Семантические модели управления. FrameNet

Синтаксис и семантика

- Синтаксис и синтаксический анализ
- Лексическая семантика (парадигматика)
 - Лексические отношения
 - WordNet
- Лексическая синтагматика
 - Как слова связаны между собой в предложении
 - Синтаксические отношения
 - Семантические отношения

Синтаксическая и семантическая модели управления

Семантические валентности

- У некоторого слова W есть <u>семантическая валентность</u> X, если слово W описывает ситуацию, в которой есть обязательный участник, выполняющий роль X
- Значения слов, связанных с W и обозначающих обязательных участников описываемой словом W ситуации, называются семантическими актантами этого слова
- Семантические актанты слова заполняют семантические валентности слова W
- Совокупность семантических валентностей семантическая модель управления

Пример: наказывать

- Лицо В совершило проступок С, а другое лицо А причиняет В некоторое зло D с целью заставить самого В (или подобных ему лиц) исправиться и впредь не совершать проступков типа С
- Семантические валентности: A, B, C, D
- Все участники необходимы для описания ситуации
 - Если нет С, то это вредительство или истязание
 - Если нет D, то это порицание

Семантические роли

- Семантические актанты слова W находятся в определенных семантических отношениях к слову W
 - А (кто наказывает) агенс, агент, субъект
 - В (кого наказывают) пациенс, адресат
 - С (проступок) мотив, причина
 - D (наказание) средство
- Предположение (60-е годы 20 века):
 - в языке существует универсальный набор семантических отношений
- Чарльс Филлмор (Fillmore):
 - Case for case (1968)
 - Семантические (глубинные) падежи

Глубинные падежи Ч. Филлмора

- Агентив падеж одушевленного инициатора действия
- Инструменталис падеж неодушевленной силы или предмета
- Датив падеж одушевленного существа, которое затрагивается состоянием или действиям, называемым глаголом
- Фактитив падеж предмета или существа, которое возникает в результате действия или состояния
- Локатив местоположение или пространственная ориентация действия или состояния
- Объектив объект действия

Однако

• С тех пор исследователи предложили много вариантов наборов семантических ролей

 Можно ли найти идеальную систему семантических отношений?

Системы семантических отношений (падежей)

- Система ЭТАП (авт. Ю.Д. Апресян) 25 семантических отношений
- Система АОТ (авт. Н.Н. Леонтьева) более 50 семантических падежей
- Филлмор (1978): не выявлено принципов, позволяющих четко отделять один падеж от другого
- Система падежей зависит от
 - Цели разработки описания
 - Предметной области и др.

Семантические модели управления vs. синтаксические модели управления

- Синтаксические актанты слова это возможные при данном слове зависимые слова
 - Синтаксическое оформление актанта зависит от самого слова (сильное управление)
 - Наказывать 4 синтаксических актанта
 - Сирконстанты слабое управление нет связи с особенностями ситуации, описанной словом: в прошлый четверг, летом, в лесу

Семантико-синтаксическая модель управления (таблица)

• Наказывать

1=A	2=B	3=C	4=D
(агенс)	(адресат)	(причина)	(средство)
Ѕим	Ѕвин	за Ѕвин	Stb

Модель управления: расщепление валентностей

• Ударить

- А ударяет Y-а X-ом (3 семантических валентности)
- Милиционер ударил грабителя дубинкой
- Милиционер ударил грабителя <u>по плечу</u> дубинкой

1=A	2=Y	3=X	4 (пораж.
(агенс)	(объект)	(инструмент)	Часть)
Ѕим	Ѕвин	Sтвор	по Ѕдат,
			в Ѕвин

Факультативные семантические валентности

• А идет из В в С (маршрутом D с целью Е)

1=A	2=B	3=C	(4=D)	(5=E)
(агенс)	(нач. точка)	(кон. точка)	(марш-рут)	(цель)
Ѕим	2. с Ѕрод	2. на Ѕвин	 Sтв по Sдат Через 	 За

Сила семантических связей

- В словарной статье слова описываются
 - Обязательные семантические актанты
 - Факультативные семантические актанты
- Сирконстанты (время, место)
- - Необязательные участники ситуации
 - Могут встретиться в разных ситуациях
 - Интерпретация сирконстантов описывается правилами
 - Иван купил автомобиль в Москве за 500 тысяч рублей = Sим V Ѕвин в Ѕпред за Ѕвин
 - Границу определить не всегда просто

Семантические валентности могут не выражаться синтаксически

• *Молчать* в значении 'не писать кому-л. или не отвечать на чьи-л. письма'

• Четырехвалетный: *кто кому на что чем не отвечает*

- Чаще всего выражается субъект, и
- Молчать о чем-либо допустимо

Выражение семантических валентностей

- Одна и та же валентность может выражаться при разных словах разными средствами
- Инструмент
 - Из+род: стрелять из ружья
 - Через+тв.п: процедить через марлю
 - Тв.падеж: резать ножом
 - С+тв.: прыгать с парашютом
 - На+предл.: тереть на терке

Сочетаемость слова

- Сочетаемость слова А это информация о требованиях, которые предъявляет слово А к слову В, синтаксически связанному со словом А.
- 1. Морфосинтаксическая сочетаемость:
 - часть речи В
 - грамматическая форма
- 2. Лексическая сочетаемость устойчивые словосочетания:
 - потупить глаза
 - похоронить заживо

Семантическая сочетаемость

- 3. Семантическая сочетаемость семантические характеристики В
 - Кто наказал человек или организация или государство
 - Кого наказали человек или организация или государство
 - Куда идти место или здание
- Система семантических характеристик

Семантические характеристики (АОТ – Леонтьева Н.Н.)

- Абстр: модель, план, тенденция, обстоятельство
- Арт: артефакты
- Велич: высокий, мощный
- Вещво
- Власть
- Вместл мешок, сейф, гараж
- Вред катастрофа, война, перегрузка
- Геогр любой географический объект
- Гос государство
- Движ движение
- Долж профессия, социальный статус
- Д-устр детали устройства

Семантические характеристики – АОТ-

- Изм действия изменения
- Инф информация утверждение, новость ...
- Носинф носитель информации: *книга, заметка, газета*
- Одеят область деятельности
- Одуш семантически одушевленный
- Предмет по размерам меньше среднего человека
- Протяж протяженные географические объекты (река)
- Собир библиотека
- Устр устройство
- Фин все, что связано с финансами: деньги, банк, долг
- Эмоции
- Явления смерч, терроризм...

Модель управления, включая семантические классы

• Наказывать

1=A	2=B	3=C	4=D
(агенс)	(адресат)	(причина)	(средство)
Ѕим	Ѕвин	за Ѕвин	Ѕтв
Одуш	Одуш	Действие	Действие,

Роль моделей управления

- Служить фильтрами при синтаксическом анализе
 - Если в синтаксической модели управления у некоторого слова указана синтаксическая валентность, то это отношение с этим словом предпочитается
 - Несоответствие семантических характеристик понижает вес отношения
- Переводить синтаксическую структуру в семантическую
 - Синтаксические отношения в семантические

Модели управления конкретных слов

- Семантические актанты, семантические характеристики, синтаксические актанты,
 - Арендовать
 - Командировать
 - Приговорить

Проекта Ч. Филлмора FrameNet

Базовое предположение проекта

- Люди понимают что-либо посредством выполнения мыслительных операций над тем, что они уже знают.
- Такое знание можно описать в виде информационных «пакетов» фреймов
- Финансируется в течение многих лет
 - National Science Foundation USA
 - DARPA USA

Деятельность в рамках проекта FrameNet

- Описать фрейм
- Найти слова, которые соответствуют этому фрейму
- Извлечь предложения из корпуса, в которых встречается отсылка на этот фрейм
- Разметить предложения элементами фрейма
- Извлечь описание семантических валентностей фрейма

The Revenge frame (Месть)

The Revenge concept involves a situation in which

- a) A has done something to harm B and
- b) B takes action to harm A in turn
- c) B's action is carried out independently of any legal or other institutional setting

Словарь для Revenge

- Nouns: revenge, vengeance, reprisal, retaliation, retribution
- Verbs: avenge, revenge, retaliate
 (against),
 get back (at), get even (with), pay back
- Adjectives: vengeful, vindictive
- V+N Phrases: take revenge, exact retribution, wreak vengeance

Выбор имен для элементов фрейма

 We develop a descriptive vocabulary for the <u>components</u> of each frame, called frame elements (FEs).

 We use FE names in labeling the constituents of sentences exhibiting the frame.

FEs for Revenge

• <u>Frame Definition</u>: Because of some **injury** to something-or-someone important to an **avenger** (maybe himself), the **avenger** inflicts a **punishment** on the **offender**. The **offender** is the person responsible for the **injury**.

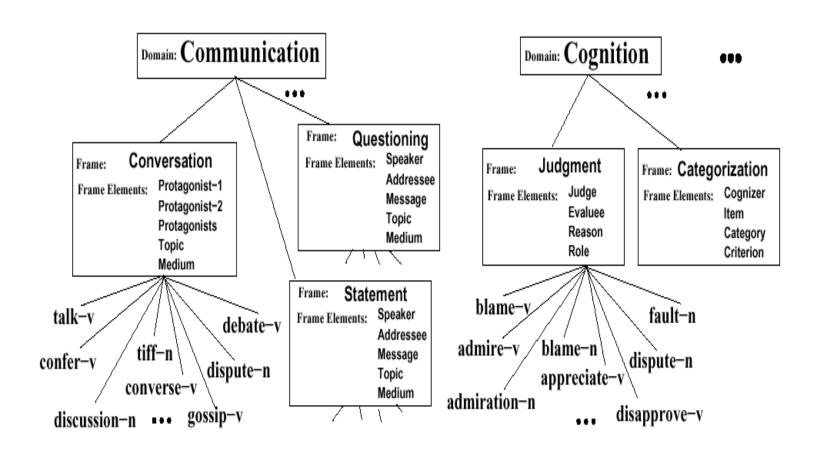
• FE List:

- avenger,
- offender,
- injury,
- injured_party,
- punishment.

Семантические роли в FrameNet

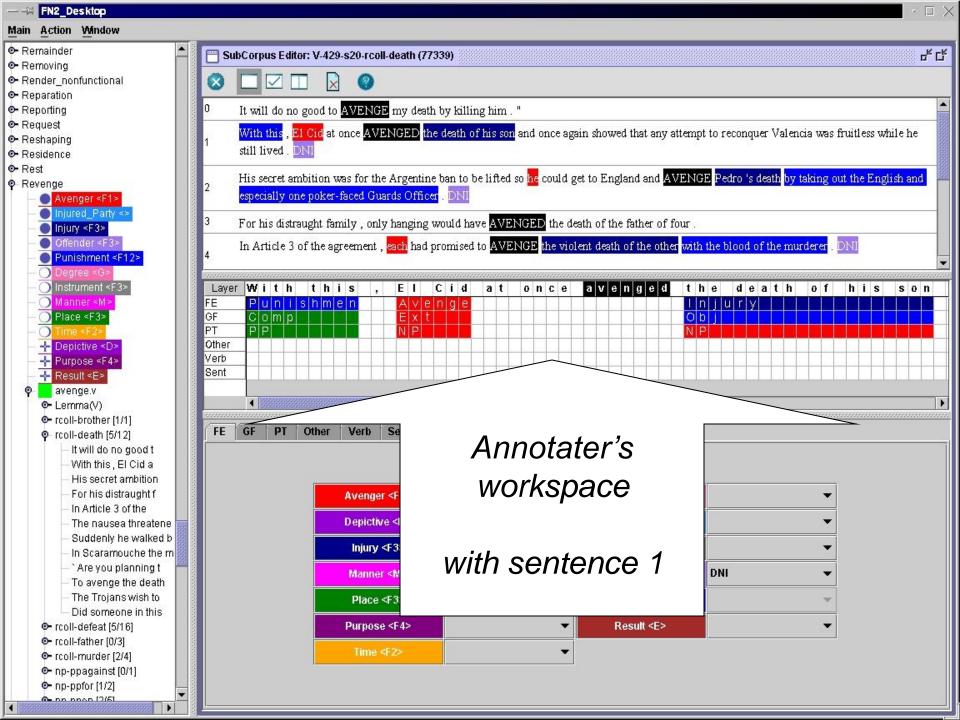
- Notice that we use such situation-specific notions as injury, offender, etc., rather than limiting ourselves to some standard list of thematic roles, like agent, patient, goal, etc.
- Причины отказа от единой системы семантических ролей
 - we would waste too much time finding criteria to do the mapping,
 - and we would have to remember what decisions we'd made.
 - !! Т.е. отказ от фиксированного набора семантических ролей

Отношения между фреймами



Сбор примеров для фрейма

- Из корпуса извлекаются примеры предложений, в которых проявляется использование каждого слова из фрейма.
- Проект делает опору на корпусные данные, а не на существующие словари или интуицию.
- Корпус:
 - Британский национальный корпус (British National Corpus);
 - Новостные тексты от Linguistic Data Consortium.
 - Всего 200М слов



Разметка предложения

 Позволяет выявить вариативность, с которой элементы фрейма могут выражаться: способы выражения валентностей

By what syntactic means is **offender** realized?

- Sometimes as direct object:
 - we'll pay you back for that
- Sometimes with the preposition <u>on</u>
 - they'll take vengeance on you
- Sometimes with *against*
 - we'll retaliate <u>against them</u>
- Sometimes with with
 - she got even with me
- Sometimes with at
 - they got back <u>at you</u>

Размеры FrameNet

- 1200 Семантических фреймов
- 13000 лексем (значений слов)
- 190 тыс. размеченных предложений

• Начало работ с 1997 года

Семантический анализ. Задача разметки семантических ролей

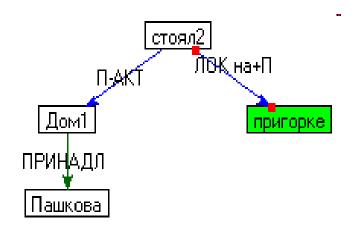
Семантический анализ текста

- Построение семантической интерпретации слов и конструкций
 - Разрешение многозначности слов

- <u>Установление семантических</u> <u>отношений между элементами текста</u>
 - Словари и правила
 - Машинное обучение по размеченным данным

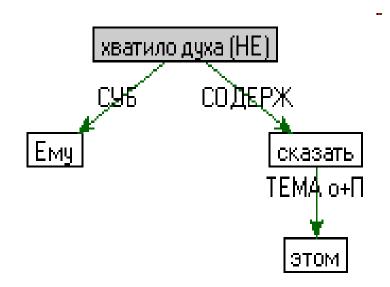
АОТ: выход семантического анализа

- Дом Пашкова стоял на пригорке
- ПРИНАДЛ (Пашков,дом) П-АКТ(дом,стоял) ЛОК (пригорке,стоял)



АОТ: выход семантического анализа -2

• Ему не хватило духа сказать об этом



Разметка семантических ролей

- Упрощенный семантический анализ
- Разметка в предложении текста заданного набора семантических ролей

Semantic Role Labeling (SRL)

• Рабочие чистят дорогу

Предикат: чистить

кто: рабочие что: дорога

• <u>Александр Зарибко</u> продал на открытом рынке принадлежащие ему <u>5 млн. акций ОАО «Дикси Групп»</u>, выручив в общей сложности <u>1,525 млрд. рублей</u>

Предикат: продать

продавец: Александр Зарибко объект продажи: 5 млн. акций

цена: 1,525 млрд. рублей

Задача SRL

• Дано:

- Синтаксический разбор предложения
- Морфологическая и лексическая информация
- Предикат
- Набор ролей
- Требуется:
 - Найти актанты этого предиката
 - Приписать актантам роли

Методы SRL

Правила

- + легче интерпретировать
- + возможна тонкая настройка
- + не нужны корпуса
- domain-dependent
- language-dependent
- высокие затраты на разработку

• Машинное обучение

- + domain-independent
- + быстрая разработка (если повезет)
- language-dependent!
- нужны корпуса
- трудно интерпретировать

Методы

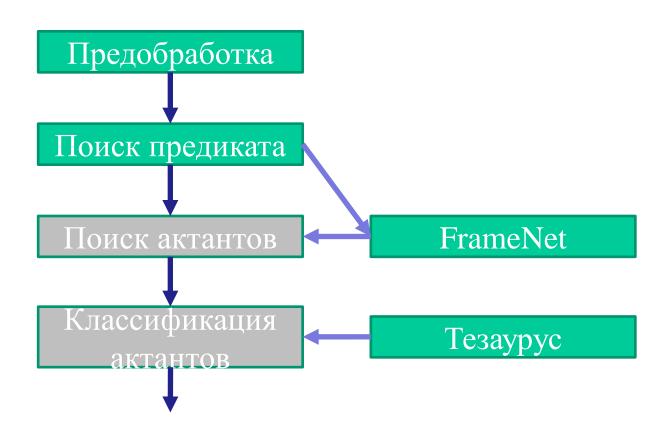
- Для английского в основном машинное обучение:
 - Есть корпуса и словари (WordNet, FrameNet)
 - Есть доступные морфологические и синтаксические анализаторы
- Для других когда как:
 - Можно делать систему на правилах
 - Можно ускорять разработку корпусных и словарных ресурсов и придумывать всякие хитрости с unsupervised learning

Классическое решение

Daniel Jurafsky & Daniel Gildea, 2002 «Automatic Labeling of Semantic Roles»

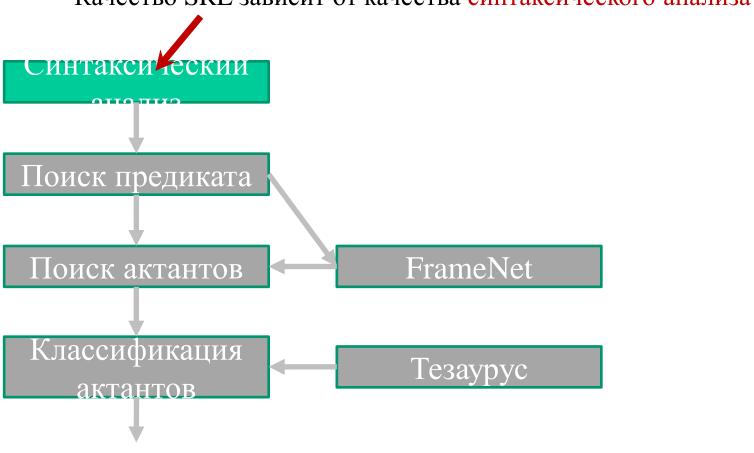
- Английский
- Синтаксис составляющих
- FrameNet

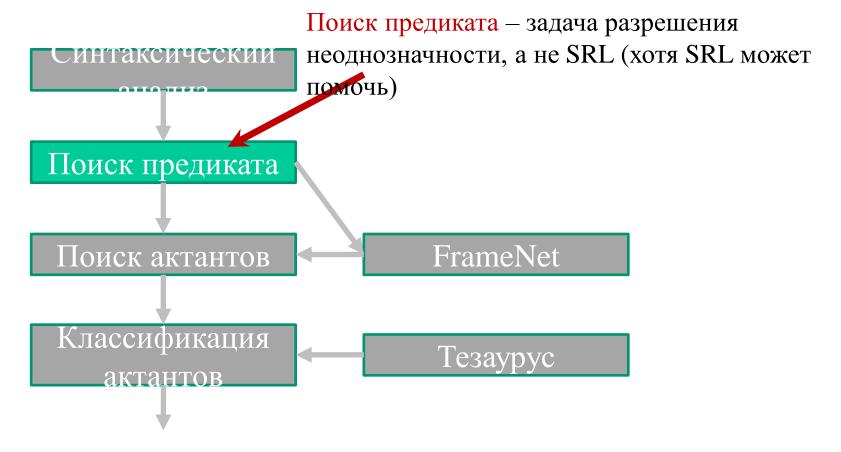
Архитектура SRL

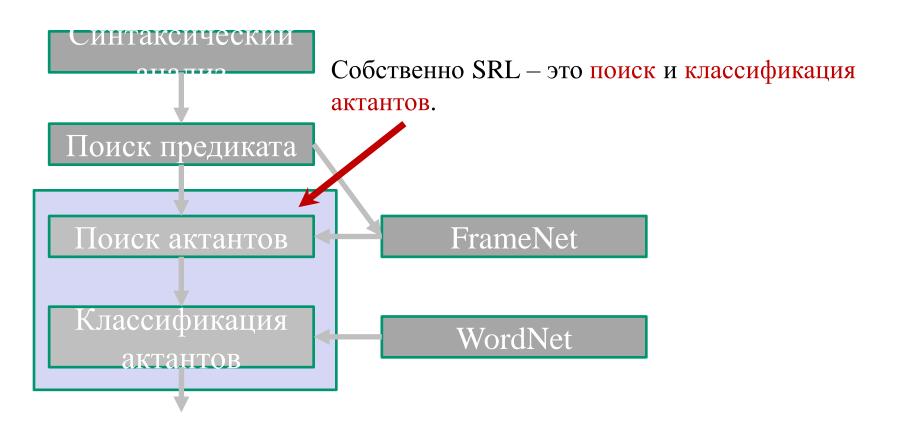


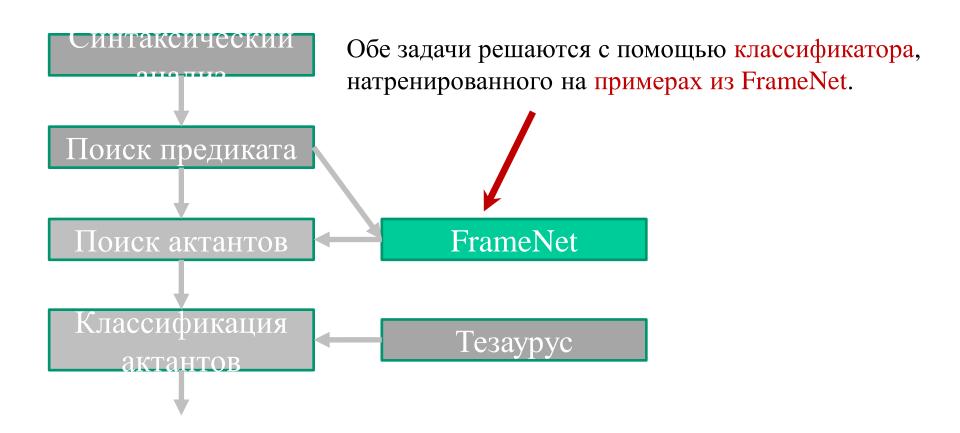
Архитектура SRL

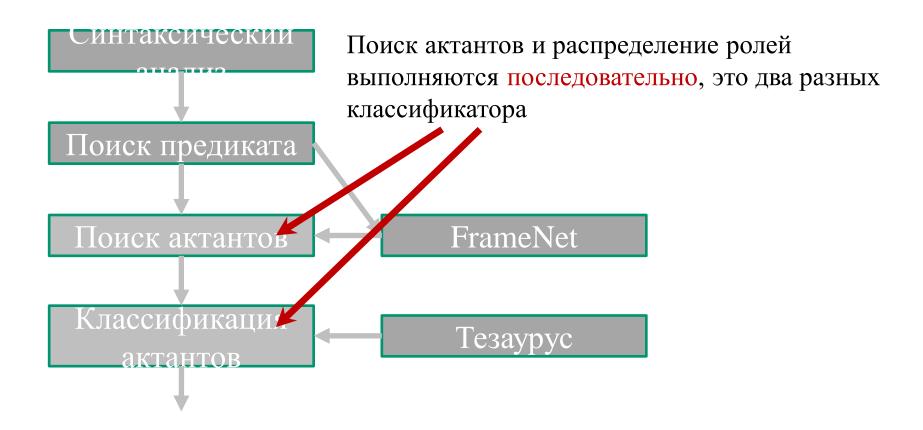
Качество SRL зависит от качества синтаксического анализа

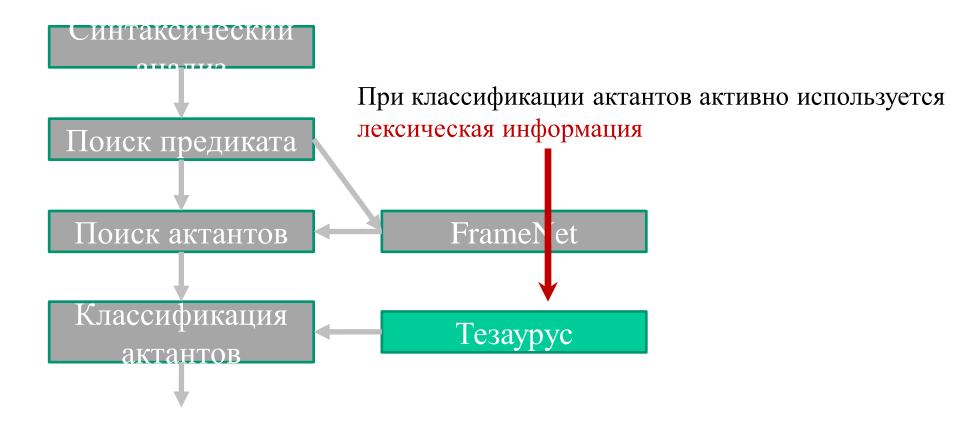












Классификатор

• Очень разреженные данные, большое признаковое пространство, мало примеров для каждого конкретного предиката

- Отдельные классификаторы для поиска актантов и для их распределения по ролям
 - Первый классифицирует все узлы синтаксического дерева
 - Второй только «актантные» узлы

Признаки

- Тип составляющей (phrase type, pt)
- Управляющая категория (gov): VP или S
- Путь от предиката (path)
- Залог глагола (voice): активный или пассивный
- Позиция относительно предиката (position): справа или слева
- Лексема в вершине актанта (head word, h)
- Лексема в вершине предиката (target word, t)

Свойства

pt: PRP

gov: S

path:

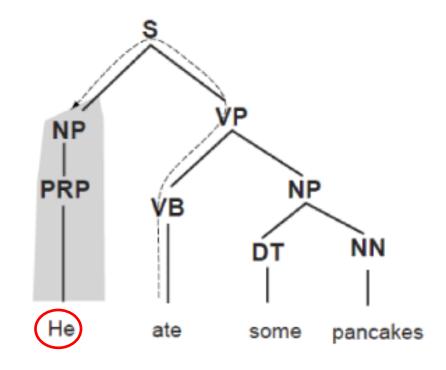
VB[↑]VP[↑]S↓NP

voice: active

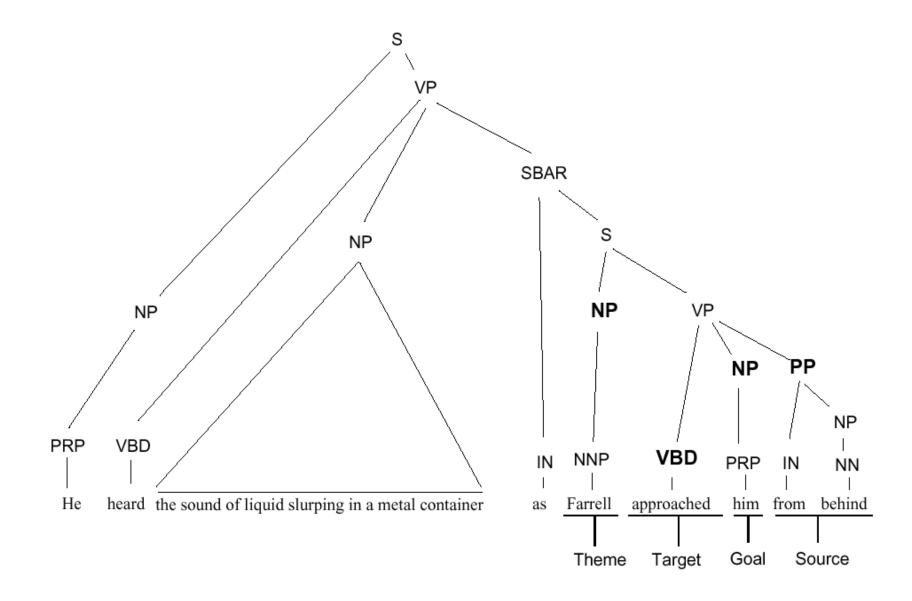
position: left

h: "he"

t: "eat"



Еще пример



Лексическая информация

 Лексическая информация обладает высокой различительной силой для классификации актантов

«Мальчик автомобиль 1000\$ купить»

- Но покрытие низкое!
- Нужно откуда-то брать информацию о близости значений лексем:
 - Готовый тезаурус (WordNet) высокая точность, низкое покрытие
 - Кластеризация лексики высокое покрытие, низкая точность
 - [Gildea, Jurafsky 2002]: кластеризация может быть использована

Полученные результаты

- Классификация актантов: точность ≈ .80
- Поиск + классификация: точность ≈ .79
- Синтаксис + поиск: точность/полнота ≈ .74/.70
- Синтаксис + поиск + классификация: ≈ .64/.61
- Полностью корректный разбор в 38% предложений

Inter-annotator agreement ≈ .70

Русский язык?

- Русский FrameNet?
- Русский WordNet?
- Синтаксический корпус СинТагРус
- Морфологический анализ доступен
- Синтаксический анализ -?