
Как помочь компьютеру понять, кто стоял на балконе

Юдина М.В. (АВВУУ)
Диалог, Москва, 30.05.2010

Проблема синтаксической неоднозначности

- Представляет проблему для синтаксических парсеров: какую стратегию выбрать?
 - Без дополнительной информации может быть принципиально неразрешима: *Маша читала и писала письма* (пример из [Юдина, Янович, Федорова 2007])
-

Разрешение неоднозначности в общении

- В большинстве случаев человек способен мгновенно разрешить неоднозначность, опираясь на контекст и модель мира
 - В подавляющем большинстве случаев (наивный) человек даже не замечает имеющуюся синтаксическую неоднозначность
-


Исследование неоднозначности

- Исследование синтаксической неоднозначности в основном заключается в перечислении омонимичных конструкций ([Иорданская 1967])
 - В компьютерных анализаторах синтаксическая неоднозначность разрешается либо случайно, либо на основании статистических данных
-

Раннее-позднее закрытие

- В этом докладе речь пойдет об одном типе синтаксической неоднозначности, называемом «ранним-поздним закрытием». Это неоднозначность сложноподчиненного предложения с относительным придаточным:

*Преступник застрелил служанку (N1) актрисы (N2),
которая стояла на балконе (RC).*



Придаточное (RC) может быть отнесено и к N1 (раннее закрытие, P3), и к N2 (позднее закрытие, ПЗ).

-
- Раннее-позднее закрытие исследовано с точки зрения влияния различных факторов и на разных языках.
 - Рассмотрим некоторые факторы, которые наиболее хорошо могут быть формализованы для использования в машинных анализаторах.
-

Длина придаточного предложения

- [Fodor 1998]: короткие придаточные чаще модифицируют N2 (позднее закрытие), в случае длинных придаточных распределение зависит от других факторов
 - Причина: короткие придаточные имеют «легкий» просодический статус, поэтому чаще присоединяются к зависимому слову ИГ
 - Проверена на русском материале ([Федорова, Янович 2004])
-

Лингвистическая настройка

- [Mitchell et al. 1995]: при разрешении неоднозначности человек склонен настраиваться на ту стратегию, которая является в языке наиболее частотной
 - Проверена на разных языках
-

Одушевленность существительных

- [Desmet et al. 2002]: при наличии в ИГ одушевленного и неодушевленного существительного одушевленное сущ. выбирается чаще в любой позиции
-

Контекст

- [Desmet, de Baecke et al. 2002], [Юдина 2006]
 - Три типа контекста: нейтральный, склоняющий к РЗ, склоняющий к ПЗ.
 - Гипотеза: после контекста, склоняющего к РЗ, будет больше РЗ, и наоборот.
-

Контекст

□ Пример экспериментального блока:

Нейтральный контекст: *Вчера на нашей улице случилось потрясшее всех ужасное происшествие. Преступник застрелил служанку актрисы, которая...*

Контекст, склоняющий к РЗ: *Эта актриса всегда давала приют нарушителям закона, пока не случилось несчастье. Преступник застрелил служанку актрисы, которая...*

Контекст, склоняющий к ПЗ: *Среди актрис этого городка было принято давать приют нарушителям закона, пока не случилось несчастье. Преступник застрелил служанку актрисы, которая...*

Контекст

- Результаты: 91% РЗ после контекста, склоняющего к РЗ, 60% РЗ после контекста, склоняющего к ПЗ
-

Синтаксический прайминг

- Говорящий склонен использовать те синтаксические конструкции, которые он недавно каким-либо образом обработал
 - Основное проявление – диалог
 - Высказывание, осуществляющее преднастройку – прайм, высказывание, на которое «нацелен» прайм – цель.
-

Синтаксический прайминг

□ [Scheepers 2003], [Юдина 2009]

□ Пример экспериментального блока:

РЗ-прайм: *На заседании утвердили бюджет организаций,
который ...*

ПЗ-прайм: *На заседании утвердили бюджет организаций,
которые ...*

Базовый прайм: *На заседании утвердили бюджет
организаций, и оказалось ...*

Цель: *Власти решили учредить конкурс газонокосилок,
которым ...*

Синтаксический прайминг

- Результаты: 57% РЗ после РЗ-прайма, 46% РЗ после ПЗ-прайма.
 - 2 гипотезы: остаточная активация и имплицитное научение
-

Общие результаты

- Подтвержденное всеми экспериментами базовое соотношение РЗ и ПЗ в русском языке: 60% - 40%
 - 296 испытуемых
 - 5400 экспериментальных предложений с разрешенной неоднозначностью
 - Структура предложений:
Subject Verb N Ngen [Relative Clause].
-

Класс глагольной вершины

- Некоторые предложения во всех экспериментах показывают схожие результаты по критерию присоединения придаточного, в независимости от экспериментального условия
 - Например, для закрытия предложения *В парке друзья встретили ассистентку профессора,...* испытуемые предпочитали присоединение придаточного предложения к первому имени, что было подтверждено результатами двух экспериментов.
-

Эффект рода?

- Мы предположили, что ИГ неравноправна с точки зрения рода, т.е. что женский род (слуга [актрисы]) способен перетягивать закрытие.
 - Дополнительный эксперимент не подтвердил этот эффект.
-

Список выделенных глаголов

- Заметить
 - Увидеть
 - Смотреть
 - Слушать
 - Встретить
 - Навестить
 - Столкнуться
 - Разругаться
 - Поссориться
 - Договориться
 - Познакомиться
 - Узнать
-

Классификация и анализ

- Классификация по тезаурусу Роже ([Roget's Thesaurus 2000])
 - Подсчет количества РЗ и ПЗ в предложениях, содержащих данный глагол.
 - Критерий оценки - количество РЗ при нейтральном контексте (или после базового прайма) и при позднем контексте (или после прайма, склоняющего к ПЗ). Предикат способствует РЗ, если число РЗ в вышеуказанных условиях составляло 100-80%, и к ПЗ – если меньше 50%.
-

предикат	Р / П	тип предиката по Роже
заметить	Р	matter → organic matter → vision
смотреть	Р	matter → organic matter → vision
слушать	Р	matter → organic matter → hearing
увидеть 1	Р	matter → organic matter → vision
встретить	Р	space → motion → arrival
навестить	Р	space → motion → arrival
столкнуться	Р	space → motion → arrival
разругаться	П	volition → individual volition → antagonism → dissention
поссориться	Р\П	volition → individual volition → antagonism → dissention
договориться	Р	volition → individual volition → antagonism → concord
познакомиться	П\Р	intellect → formation of ideas → results of reasoning → knowledge
узнать	П	intellect → formation of ideas → results of reasoning → knowledge
увидеть 2	П	intellect → formation of ideas → results of reasoning → knowledge

Увидеть 2

- Во всех экспериментальных предложениях, кроме одного, этот глагол показывал большой процент ПЗ.
 - Практически во всех предложениях глагол «увидеть» означал скорее «увидеть и узнать», например *Выходя на улицу, Катя вдруг увидела ученицу доктора, ...* (увидела и узнала, что это именно ученица доктора).
 - Лишь в предложении *В конце концов Алексей увидел посетительницу директора, ...* «увидеть» функционирует в своем прямом значении глагола зрения. Обозначим его «увидеть1», а другое значение будем маркировать как «увидеть2».
 - В словарях подобное различие не фиксируется.
-

Результаты

- Глаголы одного класса похожи в том, как разрешается синтаксическая неоднозначность в предложении, где вершиной является данный предикат.
 - Предложения с глаголами чувственного восприятия и глаголы движения имеют большой процент РЗ, глаголы, связанные с мышлением и интеллектом – ПЗ.
 - Класс глаголов отношений неоднороден: глагол «договориться» показывает практически 100% РЗ. Как нам кажется, это может быть связано со структурой ИГ: экспериментальное предложение *Я решил договориться с женой строителя, ...* почти все испытуемые заканчивали описанием того, что может сделать с (нерадивым) строителем жена (запретит пить, будет бить каждый вечер и т.п.)
-

Спасибо за внимание!

Юдина Мария (mailto:
maria_yu@abbyy.com)
