

Advanced Natural Language Processing

BM1: Advanced Natural Language Processing

University of Potsdam

Tatjana Scheffler

tatjana.scheffler@uni-potsdam.de

October 18, 2016

Today

- Introduction
- Course structure & expectations

Goal

Learning Language Games through Interaction

A Fast Unified Model

Samuel R. Bowman^{1,2,3,*}
sbowman@stanford.edu

Raghav Gupta^{2,3,6}
rgupta93@stanford.edu

¹Stanford

Automatic Labeling of Treebanks

Xiaojun Wu

Institute of Computer Science and Technology
Linguistics, Peking University
{wanxiao, jiwu}@pku.edu.cn

Danov Liang

Christopher D. Manning

Department of Linguistics

Case and Cause in Icelandic:

Reconstructing Causal Networks of Cascaded Language Changes

Fermín Moscoso del Prado Martínez

University of Cambridge

Text Summaries

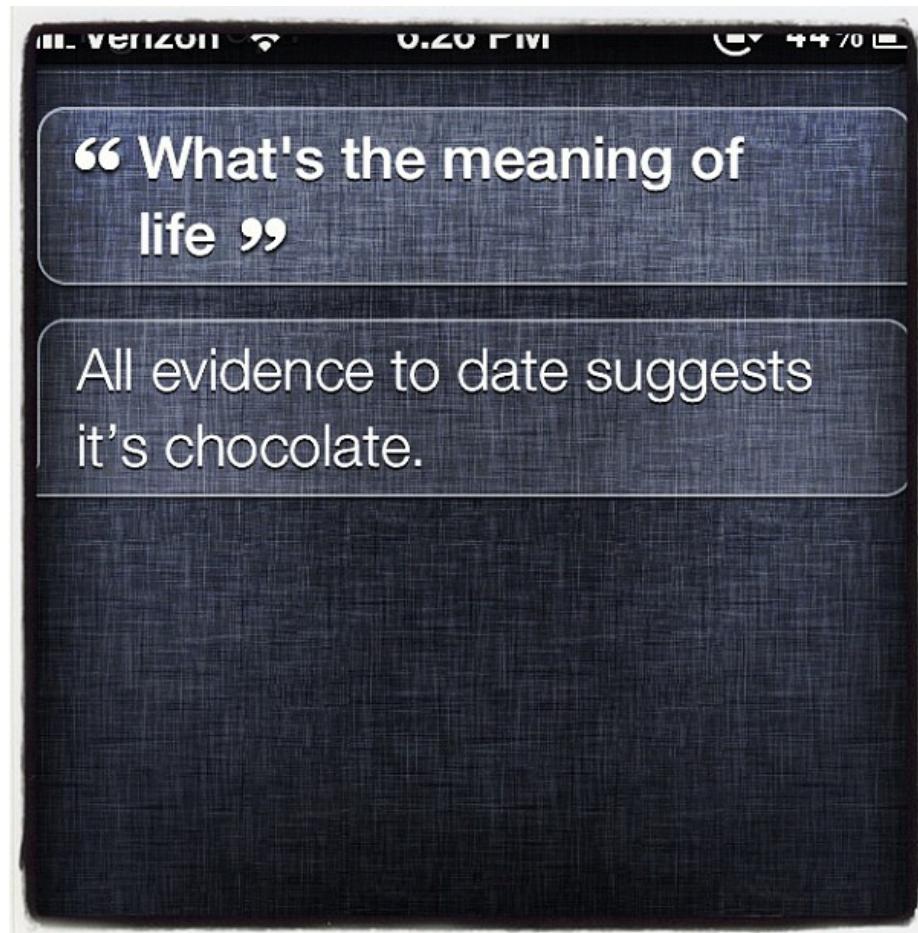
Generalized Transition-based Dependency Parsing
via Control Parameters

Bernd Bohnet, Ryan McDonald, Emily Pitler and Ji Ma

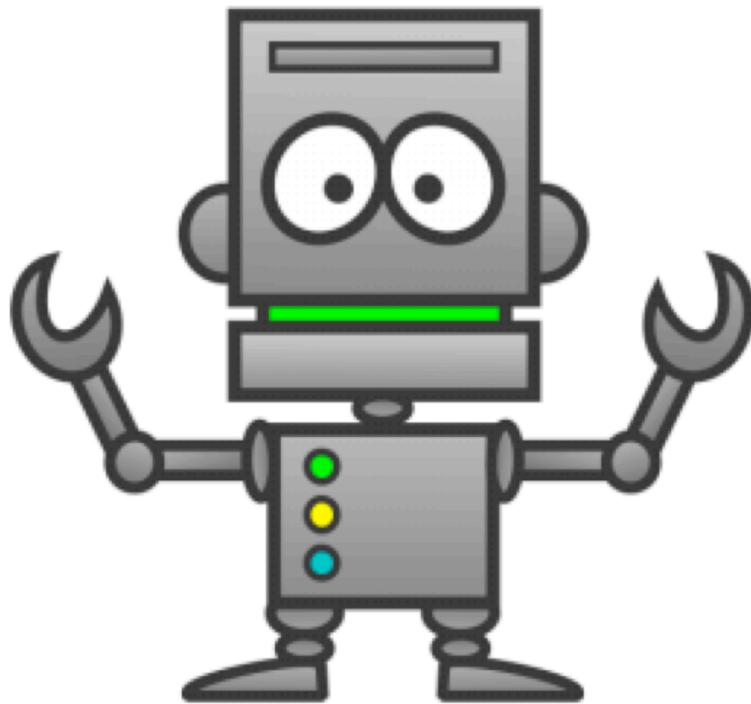
Google Inc.

{bohnetbd, ryanmcd, epitler, maji}@google.com

Natural Language Interfaces



Which tasks can my computer do?



- ❑ translate
- ❑ book my vacation
- ❑ answer my emails
- ❑ find out what people think about my product
- ❑ research castles in Potsdam

Language: the greatest interface



Christopher Manning, Stanford



SmartKom (DFKI, 2004)



Mobile scenario

Norbert Reithinger, DFKI

Mercedes-Benz Integrates Apple's Siri

By Raymond Wong | Wednesday, Feb 29, 2012 | Updated 8:00 AM PST

[View Comments \(\)](#) | Email | Print

[Tweet](#)

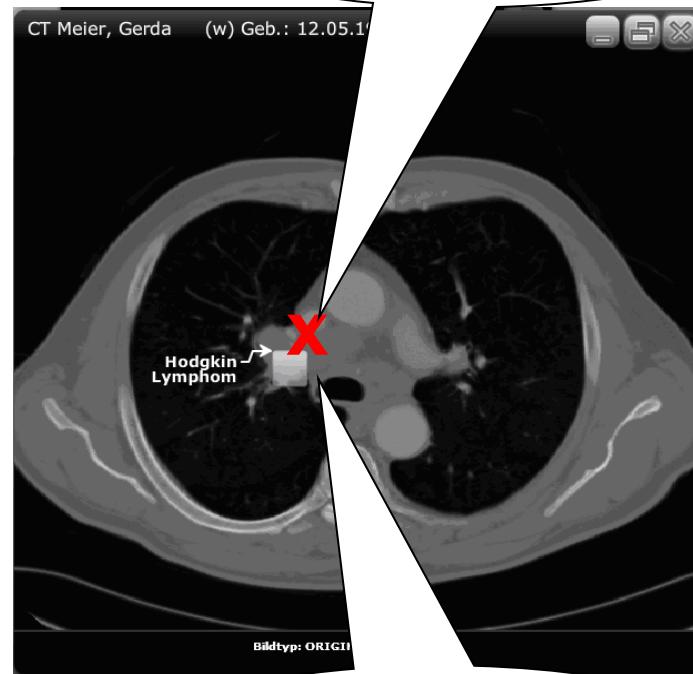
[Submit](#) 0



Mercedes-Benz has integrated the iPhone 4S and Siri into its latest dash.

Radspeech (DFKI, 2011)

**“Annotate with
Hodgkin-Lymphom.”**



**“I added the special
disease type.”**

Radiology Environment

*"This lymph node
here, annotate
Hodgkin-Lymph."*

Desktop <->

Touchscreen,
Speech
and Multimodality



*Patient Finding
Workstation*



*CT/MR Imaging
Center*

German High Tech Award



- Make dialogue-based radiology image reporting possible
- Reduce turn-over times and annotation errors
- Facilitate structured reporting



best parks in Berlin



Parks - Top10 Berlin

www.top10berlin.de/en/cat/leisure-258/parks-2638 ▾

Share this list: Rank 1: Volkspark Friedrichshain. 4.58 of 5 Top10Points. Rank 2: Tempelhofer Freiheit. 4.38 of 5 Top10Points. Rank 3: Großer Tiergarten. 4.33 of 5 Top10Points. Rank 4: Schlosspark Charlottenburg. Rank 5: Volkspark Rehberge. Rank 6: Volkspark Humboldthain. Rank 7: Viktoriapark. Rank 8: Treptower Park.

Berlin - Parks + Gardens - visitBerlin.de EN

www.visitberlin.de/en/keyword/parks-gardens ▾

Volkspark, or the 'People's Park', Prenzlauer Berg is situated in the east of Berlin's Pankow district. It spans approximately 29 hectares and is located on top of what... Berlin's best-known urban gardening project lies at the heart of Kreuzberg.

Our 10 Favorite Parks In Berlin » iHeartBerlin.de

www.iheartberlin.de/2015/05/08/our-10-favorite-parks-in-berlin/ ▾

May 8, 2015 - Berlin's getting warmer each day and the park season is in full bloom. Urban parks are more than just stretches of green amidst buildings and ...

Berlin - Parks + Gardens - visitBerlin.de EN

www.visitberlin.de › Home › See › Sightseeing ▾

Nature at its best. There are more than 2,500 public parks and gardens in Berlin, and almost a fifth of the city is covered with trees. Whether you are walking ...

My Top Berlin Parks And Glorious Green Gardens

www.berlindude.com/berlin-parks.html ▾

There is something magical about each of the 2500 public Berlin Parks and ... Known best for the Märkisches Museum, a one-stop repository of Berlin history.

Sentiviz

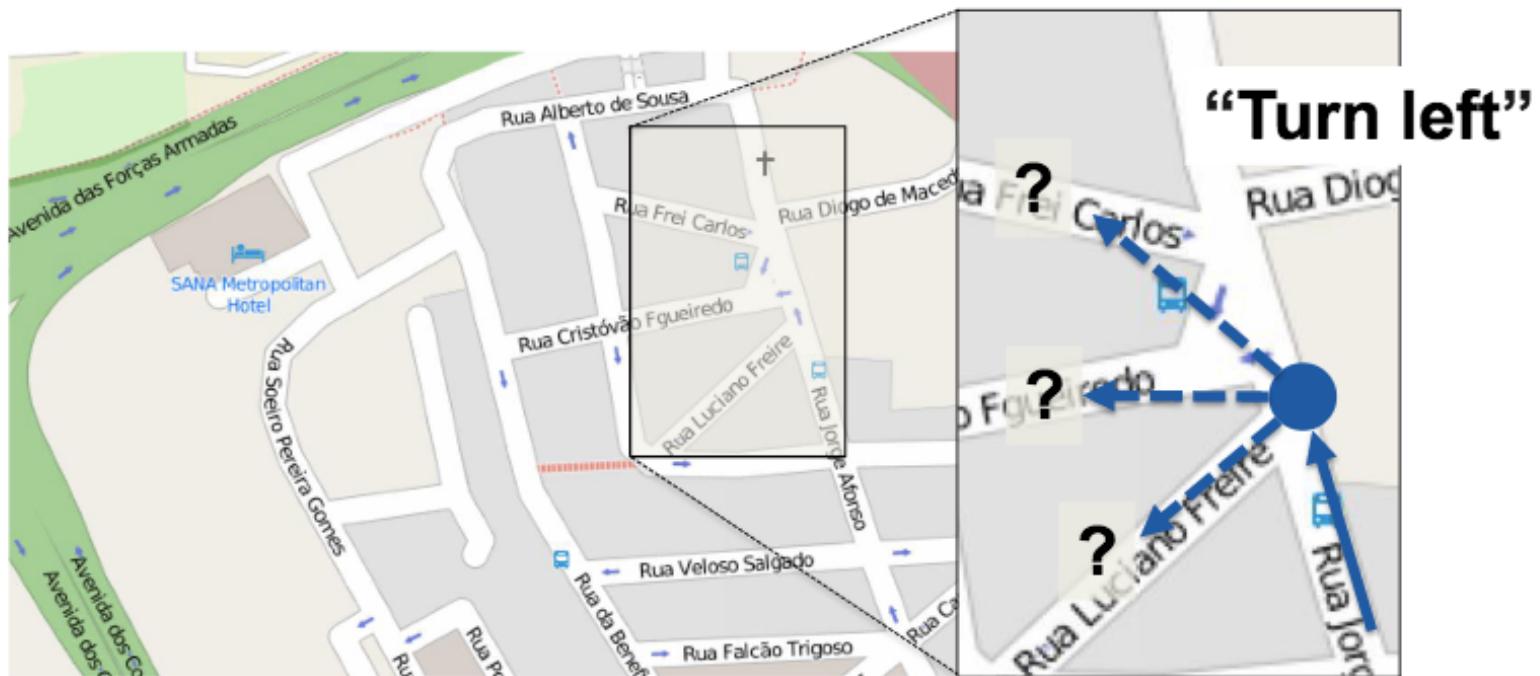


sentiment viz

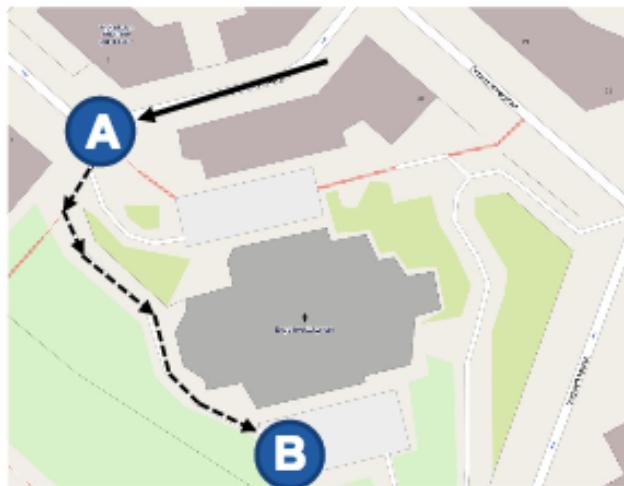
Tweet Sentiment Visualization



Pedestrian Navigation



Pedestrian Navigation



Instructions:

D: *“Turn left.”*

L: *“Walk around the church.”*

LD: *“Walk around the church to your left.”*

Google Translate

starker Kaffee = ??? coffee

starker Raucher = ??? smoker

The screenshot shows the Google Translate interface with two examples of German-to-English translation.

Example 1: The input "starker Anstieg" is translated to "strong increase".

Example 2: The input "Es gab einen starken Anstieg" is translated to "There was a sharp increase".

The interface includes a "Translate" button, language selection dropdowns ("From: German - detected" and "To: English"), and a list of other languages (English, Spanish, French, German - detected, English, Spanish, Arabic).

Translate

From: German - detected

To: English

Translate

English Spanish French German - detected

English Spanish Arabic

Es gab einen starken Anstieg

There was a sharp increase

El País (Oct 13, 2016)



15:02



Translate

Turn off instant translation



German English Russian Spanish - detected ▾



English German Russian ▾

Translate

Con un título de soterrada elegancia irónica, Los académicos y las académicas (EL PAÍS, 12 de octubre), Jesús Ruiz Mantilla da cuenta del enésimo episodio en "la más que civil batalla" (diría Juan de Mena) de quienes rechazan por sexista el uso natural y espontáneo del castellano y se empecinan en introducir especificaciones tan artificiales, tan insensatas como "nosotros y nosotras".

La cosa arranca ahora de una pieza publicada en la prensa y en la Red, en la que Arturo Pérez-Reverte embiste contra los miembros de la Academia que se negaron a hacer suya la petición que unos supuestos profesores le habían enderezado a él a título personal: se trataría de pedir amparo (?) frente a la sugerencia surgida en la Junta

With a degree of underground ironic elegance, scholars and academic (BBC, October 12), Jesus Ruiz Mantilla realizes the umpteenth episode in "the more civil battle" (say Juan de Mena) of those who reject sexist the natural and spontaneous use of Castilian and persist in introducing such artificial specifications, as foolish as "we and us".

The thing now starts a piece published in the press and on the Internet, in which Arturo Perez-Reverte onrushing against members of the Academy who refused to endorse the request that some teachers alleged had righted to him in a personal capacity : this would be to ask for protection against the suggestion that emerged in the Andalusian classroom impose the "all", "the Madrid and Madrid" and other prevarications like that (?).



Suggest an edit



Lexical Ambiguity

The image shows a screenshot of the Google Translate interface. On the left, a Spanish text block discusses a satirical title used by Jesus Ruiz Mantilla in an article for EL PAÍS. The right side shows the English translation of this text, with specific words highlighted by red boxes: 'soterrada' (underground) and 'degree' (degree). The English version also includes a link to the BBC news article.

Con un título de soterrada elegancia irónica, Los académicos y las académicas (EL PAÍS, 12 de octubre), Jesús Ruiz Mantilla da cuenta del enésimo episodio en "la más que civil batalla" (diría Juan de Mena) de quienes rechazan por sexista el uso natural y espontáneo del castellano y se especian en introducir especificaciones tan artificiales, tan inventadas como "nosotros y nosotras".

La caja de soterrada elegancia irónica de una pieza publicada en la prensa y en la Red, **EL PAÍS**, de una pieza publicada en la prensa y en la Red, Pérez-Reverte embiste contra los miembros de la Academia, que se negaron a hacer suya la petición que unos supuestos profesores le habían enderezado a él a título personal: se trataría de pedir amparo (?) frente a la sugerencia surgida en la Junta

With a degree of underground ironic elegance, scholars and academic (BBC, October 12), Jesus Ruiz Mantilla realizes the umpteenth episode in "the more civil battle" (say Juan de Mena) of those who reject sexist the natural and spontaneous use of Castilian and persist in introducing such artificial specifications, as foolish as "we and us".

The **BBC** piece published in the press and on the Internet, in which Pérez-Reverte onrushing against members of the Academy, who refused to make their own the request that some teachers alleged had rights to him in a personal capacity : this would be to ask for protection against the suggestion that emerged in the Andalusian classroom impose the "all", "the Madrid and Madrid" and other prevarications like that (?).

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the

castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle

castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle of

castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle of those

castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle of those reject

castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle of those reject as

castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist the
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist the use
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist the use natural
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist the use natural and
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist the use natural and spontaneous
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those reject as sexist the use natural and spontaneous of
castellano

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle of those reject as sexist the use natural and spontaneous of

castellano

Spanish

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del

the battle of those **who** reject as sexist the **natural and spontaneous use** of

castellano

Spanish

Translation ≈ choose words in the other language and reorder

Machine Translation

la batalla de quienes rechazan por sexista el uso natural y espontáneo del
the battle of those who reject the natural and spontaneous use of

castellano

Spanish as sexist

Translation ≈ choose words in the other language and reorder

Context



Translate

[Turn off instant translation](#)



German English Russian Spanish - detected ▾



English German Russian ▾

Translate

Con un título de soterrada elegancia irónica, Los académicos y las académicas (EL PAÍS, 12 de octubre), Jesús Ruiz Mantilla da cuenta del enésimo episodio en "la más que civil batalla" (diría Juan de Mena) de quienes rechazan por sexista el uso natural y espontáneo del castellano y se especian en introducir especificaciones tan artificiales, tan inventadas como "nosotros y nosotras".

La co...
EL PAÍS
Red, de una pieza pub...da en la prensa y en la
Mérez-Reverte...re contra los miembros de
la Academ...garon a ha...la petición que unos
supuestos profesores...trataría de pedir amparo

"nosotros y nosotras"

With a degree of underground ironic elegance, scholars and academic (BBC, October 12), Jesus Ruiz Mantilla realizes the umpteenth episode in "the more civil battle" (say Juan de Mena) of those who reject sexist the natural and spontaneous use of Castilian and persist in introducing such artificial specifications, as foolish as "we and us".

The piece published in the press and on the Internet, in Reverte onrushing again... members of the Academy wh... the request that some... others alleged had righted to him in a personal capacity : this against the suggestion that emerged in the "all", "the Madrid and Madrid" and oth...

BBC

"we and us"



Structural ambiguity

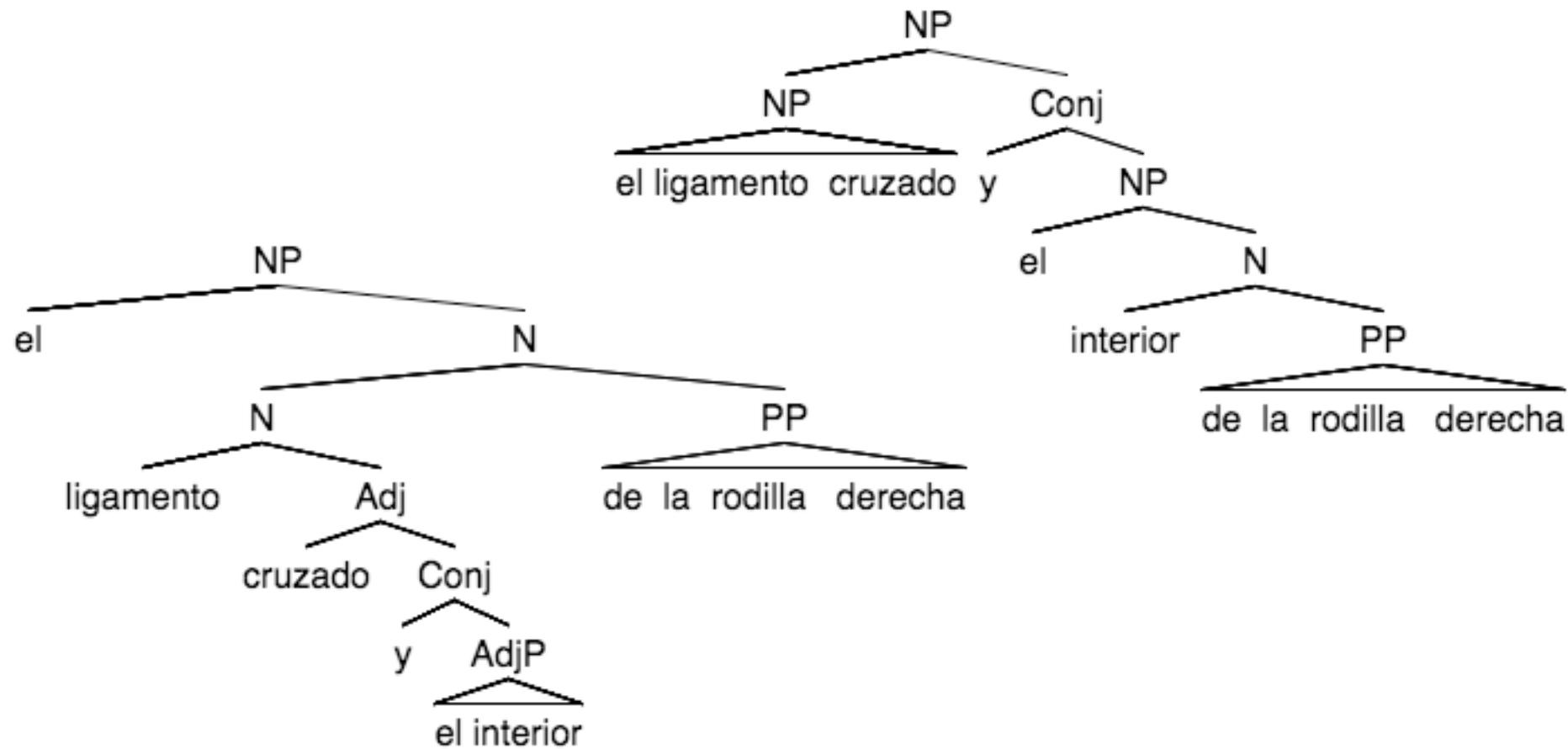
el uso **natural y espontáneo** del castellano
*the use **natural and spontaneous** of Spanish*

(se ha recuperado) el ligamento **cruzado y el interior** de la rodilla derecha
*(has self recovered) the ligament **cruciate and the interior** of the right knee*

el ligamento cruzado y el interior de la rodilla derecha
the cruciate ligament and the inside of the right knee

el ligamento cruzado y el interior de la rodilla derecha
the cruciate and the lateral ligament of the right knee

Structural ambiguity



Ambiguity: Parts of Speech

- el interior = the inside / the interior

More ambiguity...

“What's your sister doing?”

“She's painting a cow.”



Ambiguity is frequent in language

- lexical: “interest”
- syntactic: “Competent women and men hold all the jobs in this firm.” /// “I shot an elephant in my pajamas.”
- semantic: “Every student admires a professor.”
- referential: “John called Bill a Republican. He was angry.”

Ambiguity – a problem for NLP

- ❑ number of readings grows exponentially
- ❑ most naturally occurring sentences have very many possible readings
- ❑ impossible to enumerate or compute them all

=> How can we compute the correct reading efficiently?

- ❑ statistical models
- ❑ good algorithms

Uncovering hidden structure

- disambiguation = uncovering hidden structure
- This requires knowledge about language
- Classical approach: hand-written rules
 - very effective, especially in certain domains!
 - expensive/time-consuming
- Modern approach: statistical models
 - dominant paradigm since late 1990s
- Tradeoff between depth of modeling and quality of available models

Aim of this class

- Fundamental techniques of computational linguistics
- Common themes:
 - uncovering hidden linguistic structure
 - dealing with ambiguity
 - statistical methods
 - efficient algorithms

Topics

- statistical models of language
- part of speech tagging: hidden markov models
- parsing: probabilistic context-free grammars
- speech
- semantics
- machine translation
- unsupervised methods

Organization of the Course

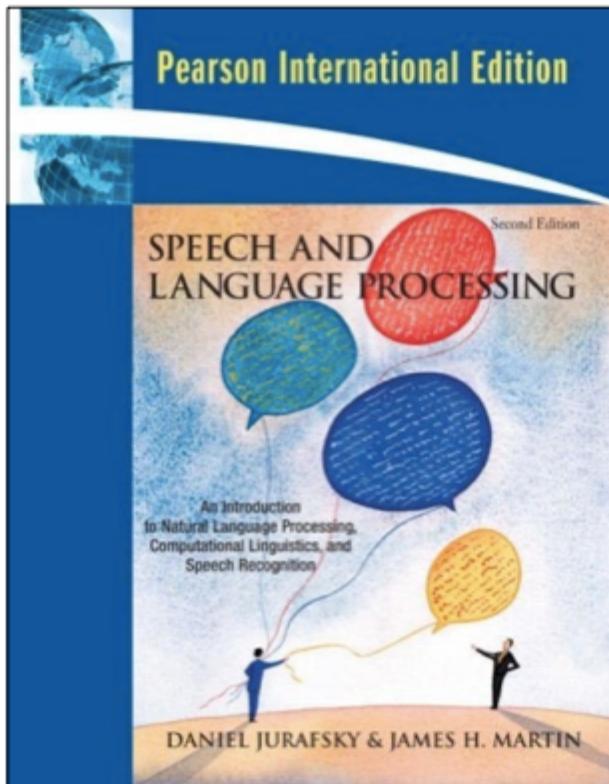
Requirements

- Serious time commitment! (270 hours of work)
- readings + discussion
- lectures
- 6 assignments
- final project

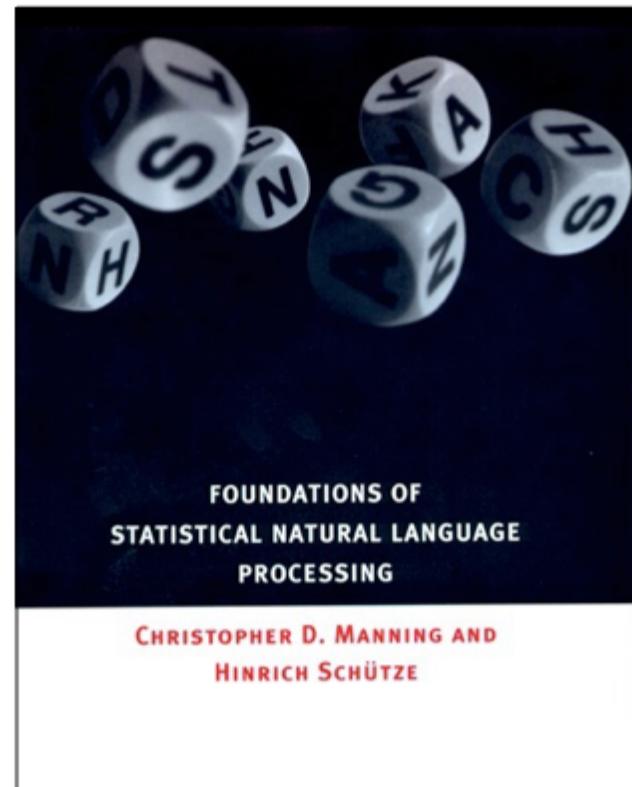
Lectures

- Tuesdays and Fridays, mixed
- We take time for discussion & assignments as needed
- Please read assigned readings before the lectures!
- Next Tuesday, Oct. 25: practical session (starting with the assignments) with the TA, Jens Johannsmeier.

Textbooks



Jurafsky/Martin: Speech and Language Processing



Manning/Schütze: Foundations of Statistical Natural Language Processing

Assignments

- Six programming assignments
- At least 5 must be turned in.
- The sum of the best 4 must be at least 250/400 to pass the course.
- Start early! 10% will be deducted for each day your assignment is late.

Final project

- Similar in scope to one assignment, to be completed during the break
- Propose and present a topic in the second half of the course
- Submission: code + documentation
- Final grade is based on:
 - proposal and its presentation
 - difficulty/creativity
 - correctness/performance
 - code readability + documentation

Class organization

- register on Moodle
=> Forum for questions/discussion!
- TA: Jens – send him a message if you have questions
- website:

<http://www.ling.uni-potsdam.de/~scheffler/teaching/2016advancednlp/index.html>

Next week

- Review of Probability Theory
- Reading:

Harald Goldstein,
A short introduction to probability and related concepts

Optional:

Manning/Schütze, chapter 2.1