

# Improving a language technological application by consulting syntactic information

**Daphne Theijssen**

Department of Linguistics

Radboud University Nijmegen

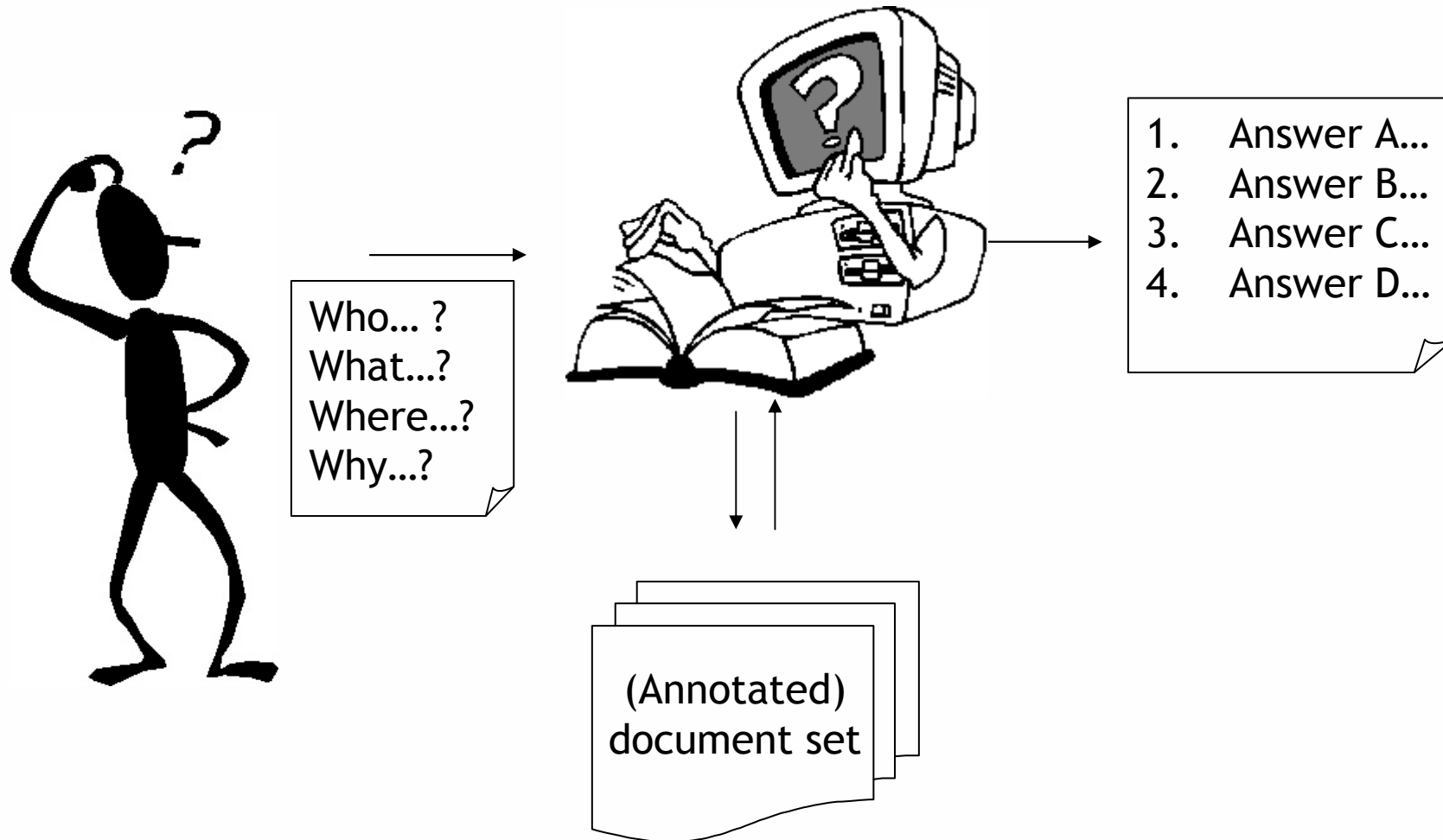
[daphnetheijssen@student.ru.nl](mailto:daphnetheijssen@student.ru.nl)

---

In search  of the WHY

- Introduction to (*why*-)Question Answering
  - Question Answering (QA)
  - A system for *why*-QA
- Exploiting syntax in question analysis
  - Question analysis
  - Research questions
  - Syntactic parsing: TOSCA (Oostdijk, 1996)
  - Data and method
  - Results
  - Answers to research questions
- Conclusion

# Question Answering (1)



# Question Answering (2)



The screenshot shows a Microsoft Internet Explorer browser window displaying the Language Computer website. The browser's address bar shows the URL: `http://www.languagecomputer.com/scripts/question.cgi?quser=&Submit22.x=238&Submit22.y=11`. The website's navigation menu includes links for COMPANY, SOLUTIONS, TECHNOLOGY, DEMOS, NEWS, PARTNERS, and CONTACT US. The main banner features a man and a woman looking at a laptop, with the text "the power to answer" and the Language Computer logo. Below the banner, a breadcrumb trail reads: [Home](#) > [Demos](#) > [Question Answering](#) > Internet Demo. The main heading is "Internet Demo". Below this, the text "Open Search. Concise Answers." is displayed. A yellow search box contains the text "Type your question:" and a text input field with the question "What is the capital of the Netherlands?". To the right of the input field is a "Power Answer" button. On the right side of the page, under the heading "Question Examples", there is a list of example questions with blue arrows pointing to the right:

- > [When did the Challenger space shuttle explode?](#)
- > [What percent of the earth air is oxygen?](#)
- > [Which is the largest volcano in Europe?](#)
- > [How many calories in a glass of wine?](#)
- > [More examples...](#)

# Question Answering (3)

Language Computer - Internet Demo - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres <http://www.languagecomputer.com/scripts/question.cgi?quser=What+is+the+capital+of+the+Netherlands%3F> Ga naar

COMPANY SOLUTIONS TECHNOLOGY DEMOS NEWS PARTNERS CONTACT US

the power LANGUAGE COMPUTER

Home > Demos > Question Answering > Int

Internet Demo

Open Search. Concise Answers.

Type your question:

The capital of the Netherlands is Amsterdam, while the seat of government is The Hague (Den Haag). Rotterdam, at the mouth of the Rhine delta, has developed...  
[www.esn-utrecht.nl/holland.php](http://www.esn-utrecht.nl/holland.php)

What is the capital of the Netherlands?

processing question... processing answers... extracting answers...

1. The capital of the Netherlands is Amsterdam, while the seat of government is The Hague (Den Haag). Rotterdam, at the mouth of the Rhine delta, has developed...  
[www.esn-utrecht.nl/holland.php](http://www.esn-utrecht.nl/holland.php)
2. Den Haag is the capital of the Netherlands , and of Zuid-Holland province. ... The national capital of the Netherlands is Amsterdam...  
[www.crwflags.com/fotw/flags/nl-zh-dt.html](http://www.crwflags.com/fotw/flags/nl-zh-dt.html)
3. Amsterdam is the capital of The Netherlands , and is also known as the ... that the , capital of the Netherlands is a different place, a special city, unique...  
[www.vueling.com/EN/cities/amsterdam.php](http://www.vueling.com/EN/cities/amsterdam.php)
4. The capital of the Netherlands is Amsterdam, even though the States-General and the government have

Gereed Internet

# Example of *why*-QA (1)

Language Computer - Internet Demo - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres <http://www.languagecomputer.com/scripts/question.cgi?quser=&Submit22.x=15&Submit22.y=17>

→ COMPANY → SOLUTIONS → TECHNOLOGY → DEMOS → NEWS → PARTNERS → CONTACT US

the power to answer

LANGUAGE COMPUTER

Home > Demos > Question Answering > Internet Demo

## Internet Demo

Open Search. Concise Answers.

Type your question:

Why is Amsterdam the capital of the Netherlands?

...and get the answer **Power Answer**

### Question Examples

- ▶ [When did the Challenger space shuttle explode?](#)
- ▶ [What percent of the earth air is oxygen?](#)
- ▶ [Which is the largest volcano in Europe?](#)
- ▶ [How many calories in a glass of wine?](#)
- ▶ [More examples...](#)

# Example of *why*-QA (2)

Language Computer - Internet Demo - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres <http://www.languagecomputer.com/scripts/question.cgi?quser=Why+is+Amsterdam+the+capital+of+the+Netherlands%3F> Ga naar

COMPANY SOLUTIONS TECHNOLOGY DEMOS NEWS PARTNERS CONTACT US

the power LANGUAGE COMPUTER

Home > Demos > Question Answering > In

Internet Demo

Open Search. Concise Answers.

Type your question:

Although there is only a small and very recent legal status for Amsterdam as . capital of the Netherlands (1983), the city has been recognized as capital ever...  
[en.wikipedia.org/wiki/Capital\\_of\\_the\\_Netherlands](http://en.wikipedia.org/wiki/Capital_of_the_Netherlands)

**Why is Amsterdam the capital of the Netherlands?**

processing question... processing answers... processing answers... extracting answers...

1. Although there is only a small and very recent legal status for Amsterdam as . capital of the Netherlands (1983), the city has been recognized as capital ever...  
[en.wikipedia.org/wiki/Capital\\_of\\_the\\_Netherlands](http://en.wikipedia.org/wiki/Capital_of_the_Netherlands)
2. Held on May 15-19, 2000 in Amsterdam , the capital of The Netherlands . Leaders from industry, academia, and government will present the latest developments...  
[www.9.org/](http://www.9.org/)
3. Amsterdam (help?info) is the capital city of the Netherlands , ... Amsterdam is the financial and business capital of the Netherlands and one of the most...  
[en.wikipedia.org/wiki/Amsterdam](http://en.wikipedia.org/wiki/Amsterdam)
4. Amsterdam , Netherlands Amsterdam is the capital of Netherlands The native name of Netherlands is

# Example of why-QA (3)

nds - Wikipedia, the free encyclopedia - Microsoft Internet Explorer

Field Favorieten Extra Help

Zoeken Favorieten

lia.org/wiki/Capital\_of\_the\_Netherlands

The only other countries at present time which have a capital between capital and seat of government are [Benin](#) and [Cuba](#).

### Constitution

Amsterdam was first referred to as capital of the Netherlands when it was sworn in, *in the city of Amsterdam*. In 1815 however, this remained the case, although the capital might be sworn in (it might be in The Hague). The status of Amsterdam as capital of the Netherlands since 1917, until the fully new constitution of 1953, was *Amsterdam*. The explicit intention of this change was to make it explicit that Amsterdam indeed is the capital.

### Perception

[edit]

Although there is only a small and very recent legal status for Amsterdam as capital of the Netherlands (1983), the city has been recognized as capital ever since 1814. Partly because it is a *Royal City* (used not only for the inauguration of kings, but also for royal weddings; note though that royal burials take place in [Delft](#)), and also because of its dominant position in Dutch history: since the end of the 16th century the city has grown to be the largest city in the Netherlands and the main centre of trade, economics, finance and culture. The Hague is never referred to as capital; Dutch school children are taught that "Amsterdam is the capital of the country".

The deeper cause of this situation is the fact that before the [Batavian Republic](#) of 1795 the Netherlands were not a *unitary state* but a *confederation*, the [United Provinces](#), that simply had no official capital at all as each province was, at least in theory, a *sovereign state*. People were thus used to being ruled from The Hague in fact while publicly acknowledging it as a mere residence only.

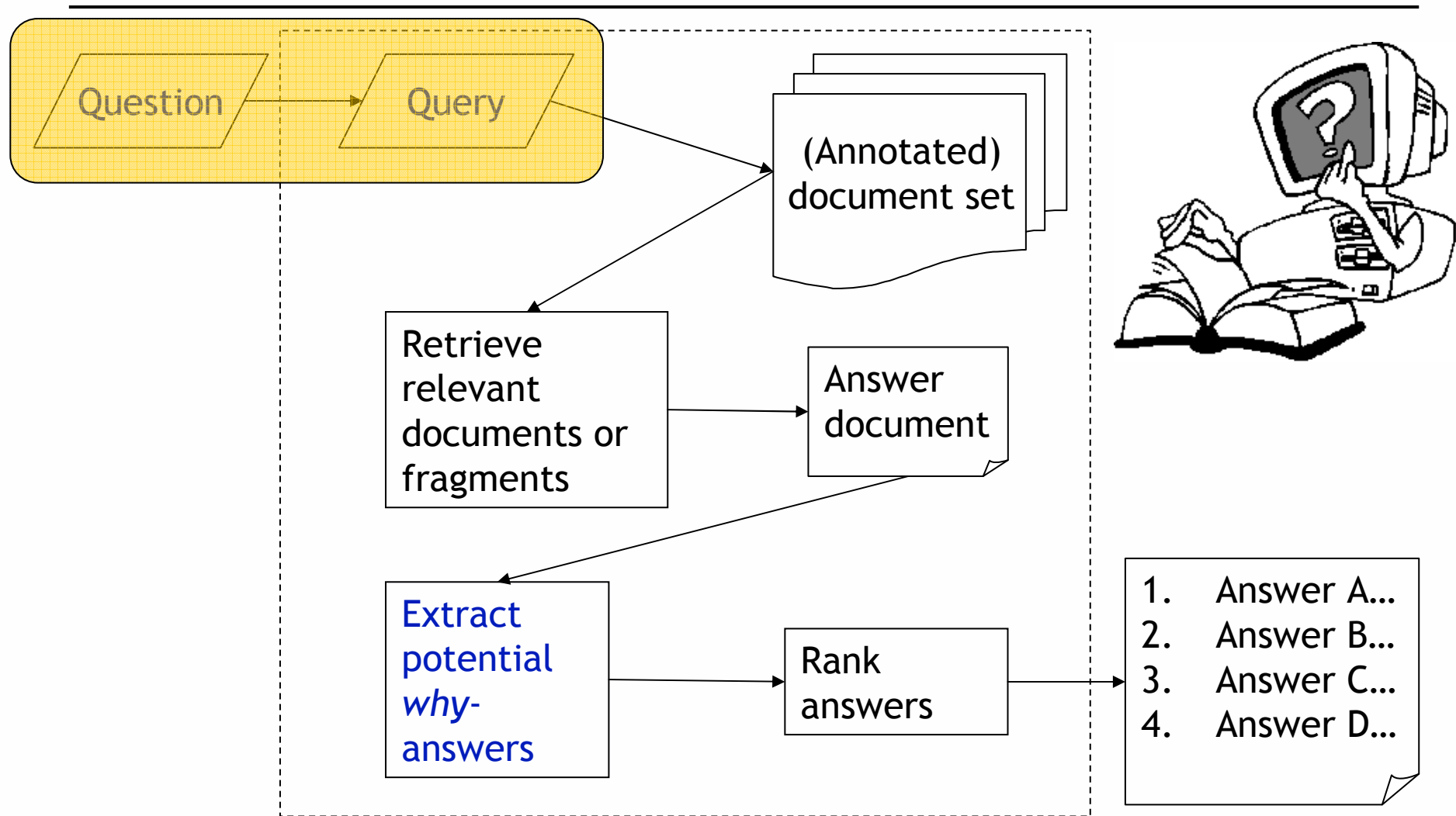
Partly because it is a Royal City and also because of its dominant position in Dutch history since the end of the 16th century...

Although there is only a small and very recent legal status for Amsterdam as . capital of the Netherlands (1983), the city has been recognized as capital ever...

[en.wikipedia.org/wiki/Capital\\_of\\_the\\_Netherlands](http://en.wikipedia.org/wiki/Capital_of_the_Netherlands)



# A system for *why*-QA



# Question analysis (1)

---

- Importance of answer type determination for quality of QA system (Hovy et al., 2001)
- Answer type for *why*-questions: ‘reason’
  - ‘cause’
  - ‘motivation’
  - ‘circumstance’
  - ‘purpose’

## Question analysis (2)

---

- Apply machine learning algorithms
- Some of the features used:
  - Subject agency
  - Verb type
  - Modality
  - Presence of intensive complementation
  - Presence of monotransitive “have”

# Research questions

---

1. Does the use of syntactic information improve the question analysis module in the *why*-QA system?
2. How do inaccuracies in syntactic trees affect the performance of the question analysis module?

## Question analysis and syntax: example (1)

---

“ After three Sagos were stolen from his home in Garden Grove, “I put a big iron stake in the ground and tied the tree to the stake with a chain,” he says proudly. ”

“ Why did Harold Smith chain his Sagos to iron stakes? ”

## Question analysis and syntax: example (2)

---

Why	Adverb
did	Auxiliary
Harold	Noun
Smith	Noun
chain	Verb
his	Pronoun
Sagos	Noun
to	Infinitive-marker
iron	Verb
stakes	Noun

- Why did Harold Smith ask his Mary to iron sheets?

# Syntactic parsing: TOSCA

---

- Deep syntactic parser
  - Categories
  - Functions
  - Attributes

# Intermezzo: example

NOFU, TXTU( )

·UTT, S(indic, inter, intr, pass, pres, unnm)

··A, AVP(inter)

···AVHD, ADV(inter) {Why}

··INTOP, AUX(indic, pass, pres) {are}

·SU, NP( )

··NPHD, N(com, plu) {films}

··V, VP(indic, intr, pass)

···MVB, LV(indic, motr, pastp) {planned}

··A, PP( )

···P, PREP( ) {for}

···PC, NP( )

····NPHD, N(com, sing) {release}

····NPPO, AVP(gen)

·····AVPR, AVP(excl)

·····AVHD, ADV(excl) {only}

·····AVHD, ADV(gen) {overseas}

·PUNC, PM(inter, qm) {?}

category

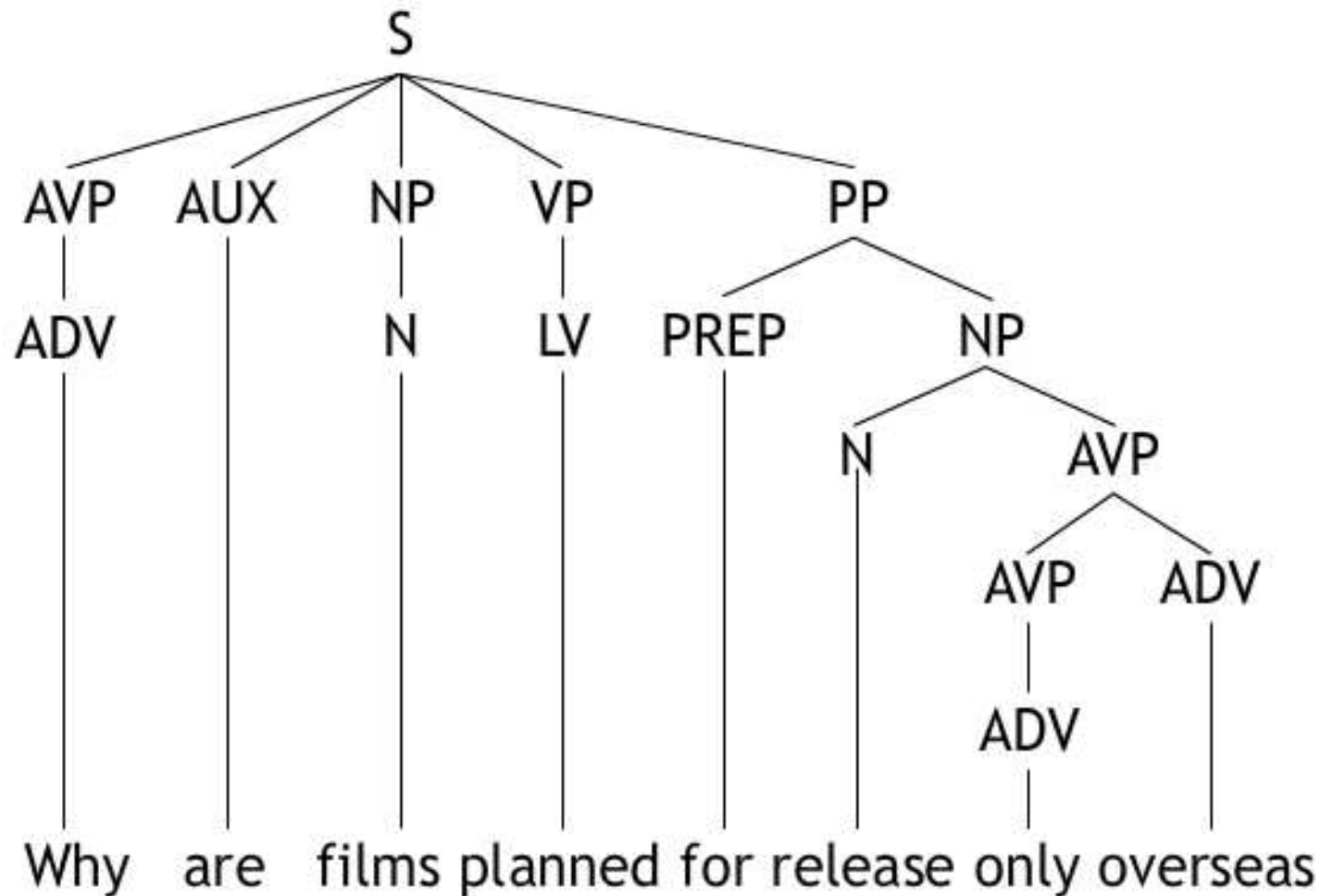
function

attribute

Why are films planned for  
release only overseas?



# Intermezzo: example



# Syntactic parsing: TOSCA

---

- Deep syntactic parser
  - Categories
  - Functions
  - Attributes
- Needs human interference at two stages:
  - After tagging → checking tags
  - After parsing → selecting desired tree
- Interactive character problematic in *why*-QA application

# Data and method (1)

---

- Focus on distinguishing ‘motivation’ and ‘cause’ (most common)
- 235 questions formulated to 13 different newspaper texts by native speakers
- Three different data sets of questions:

	coverage
1. Gold standard trees	100%
2. Semi-automatic output	98%
3. Fully automatic output	80%

## Data and method (2)

---

- Manually determine (for each question):
  - Perfect features
  - Answer type
- Automatically derive features using lexical resources and syntactic trees
- For questions that could not be parsed the part-of-speech tags were consulted
- Baseline: use words in questions as features (80.4%)

# Data and method (3)

---

## 1. Three types of feature sets:

- Words only (baseline)
- Syntactic information only
- Both words and syntactic information

## 2. Four types of sources for the features:

- Manually derived perfect features (ceiling)
- Gold standard syntactic parses
- Semi-automatic syntactic parses
- Fully automatic syntactic parses
- All offered to Timble

# Results (1)

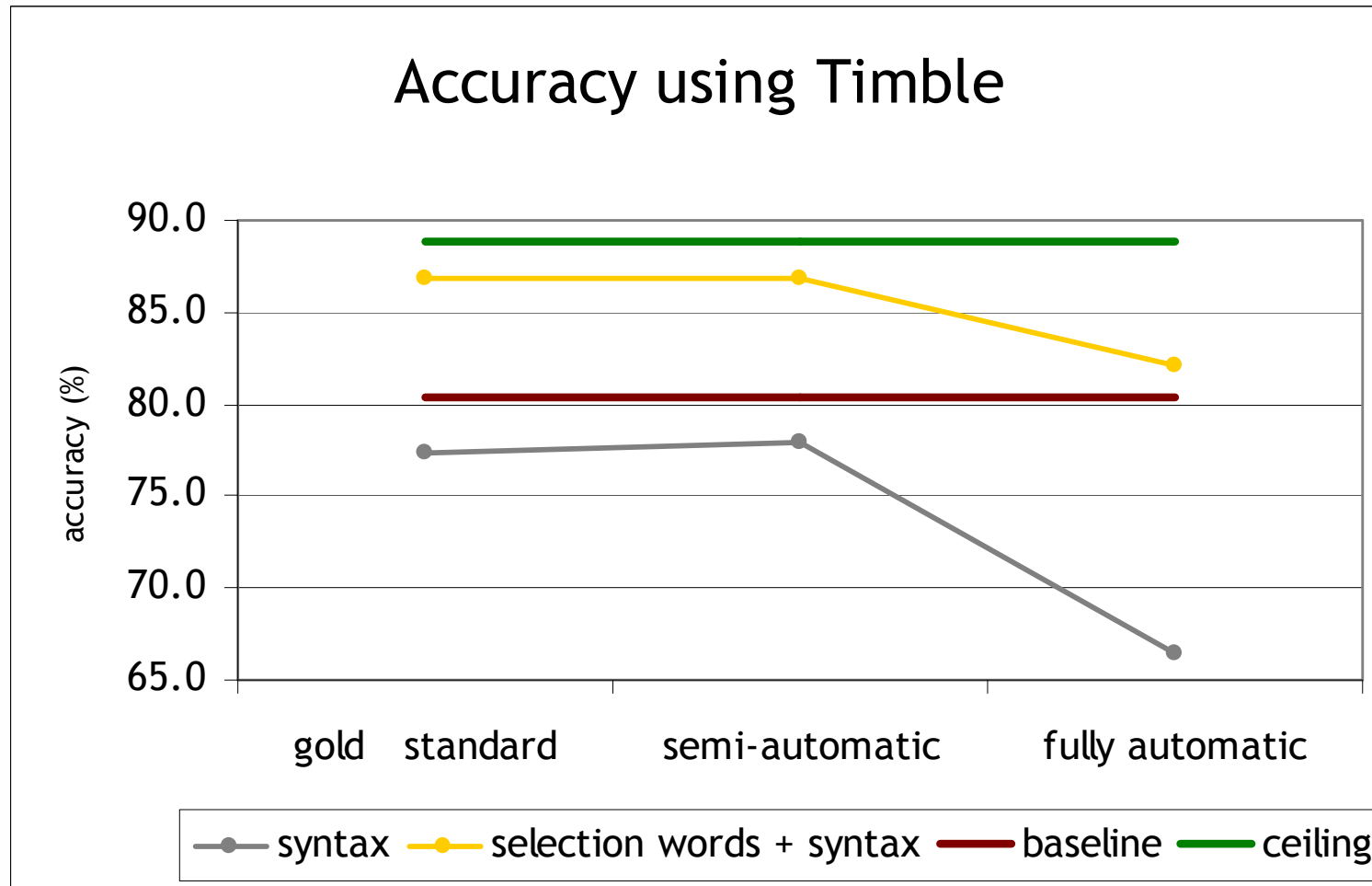
---

- Baseline: 80.4% (selection of words)
- Accuracy using Timble:

	perfect features	gold standard	semi- automatic	fully automatic
syntax	83.8	77.4	77.9	66.4
selection words + syntax	88.9	86.8	86.8	82.1

- Ceiling: 88.9% (perfect features)

## Results (2)



# Answers to research questions

---

1. Syntactic information is useful for question analysis in *why*-QA:
  - Presented method (words + syntactic information): 86.8% of 235 questions can be correctly classified as ‘motivation’ or ‘cause’, compared to 80.4% with words only
2. Too many inaccuracies decreases the benefit of using syntax considerably:
  - Automatically derived TOSCA lead to a decrease in accuracy (82.1%)



# Conclusion and further research

---

- There is a need for a question answering system for *why*-questions
- In question analysis, the use of syntactic information is beneficial as long as it is not too erroneous
- New research: automatic extraction of discourse relations from texts for the purpose of answer extraction

# Questions?

---