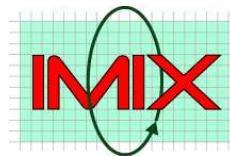


Question Answering for Dutch using Dependency Relations



Gosse Bouma Jori Mur Gertjan van Noord

Lonneke van der Plas Jörg Tiedemann

www.let.rug.nl/~gosse/Imix

Information Science

Rijksuniversiteit Groningen

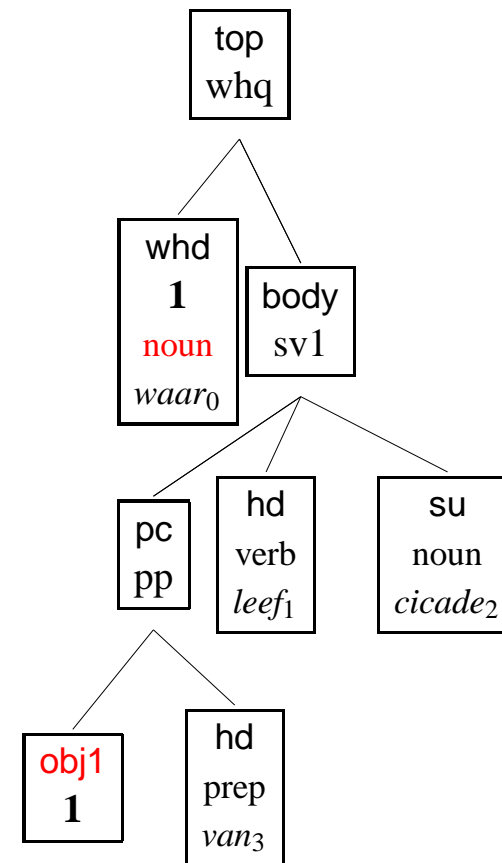
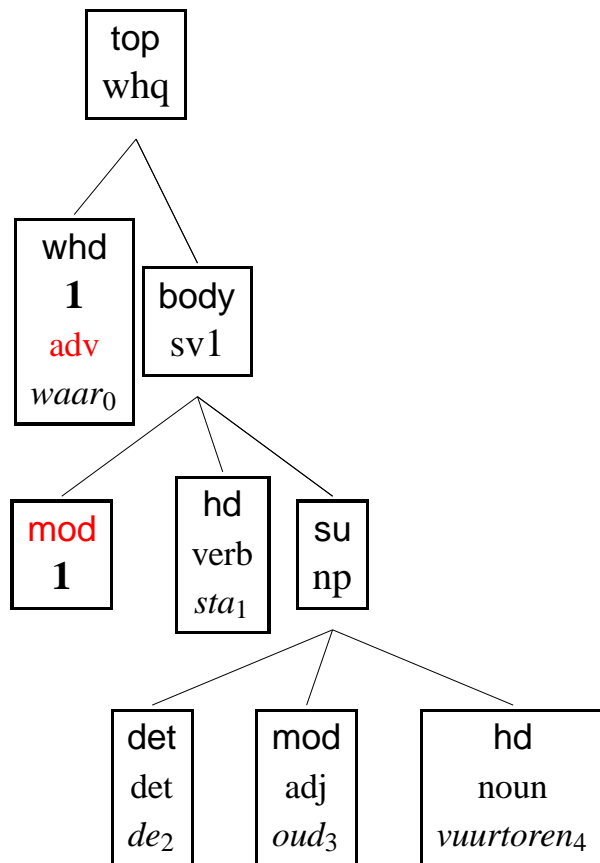
Overview

- Introduction
- Parsing for QA
- Equivalences over Dependency Relations
- Answer Extraction and Ranking
- Off-line Answer Extraction
- Answering Which questions and definition questions
- CLEF Results

Syntactic Analysis for Q Classification

- **Waar** staat de oudste vuurtoren van Nederland? (**LOC**)
 - *Where is the oldest lighthouse of the Netherlands located?*
- **Waar** staat de afkorting EMU **voor**? (**ABBR**)
 - *what does the abbreviation EMU stand for?*
- **Waar** leven cicaden **van**? (**OBJECT**)
 - *What do cicada live on?*

Classification of *Waar* questions



Using Dependency Relations for QA

- Syntactic Structure can provide useful information for
 - Question Analysis
 - Answer Extraction & Ranking
 - Off-line Answer Extraction
- **Bottlenecks**
 - Robust, wide-coverage, parsing?
 - How to use syntactic information?

But Full Parsing is impossible

Over the last five years, I don't know how often I have read a clause regretting the impossibility of full parsing and expressing the conviction that one has to make do with less desirable approximations. ... "Full parsers are slow, ..yield so many parses, ..are brittle."... When this mantra was first invented it might have expressed a truth, but by now the reality is different...

Annie Zaenen, ElsNews 13.1

Alpino Grammar and Parser

- Wide-coverage HPSG (> 500 rules)
- Large lexicon ($> 100K$ words)
- Heuristics for unknown words
- Robust left-corner parser (can return **partial parses**)
- **maximum-entropy disambiguation** model
- Output: **dependency structures**

Accuracy

Corpus	Sentences	Accuracy (%)
Eindhoven cdbl	7000	85.9
Trouw 2001	500	88.0
CLEF 03	450	95.3
CLEF 04	200	95.8
CLEF 05	200	97.6

(Malouf and van Noord 04, IJCNLP workshop
Beyond Shallow Analyses)

NE Classification in Alpino

- Named Entity **recognition** already exists
- Named Entity **classification** uses
 - **dictionary** of names (150K, mostly from Twente News Corpus (over 300 mln words))
 - **maximum entropy classifier** (trained on CONLL data)
- **88.5% accuracy** (50% dict, 50% MaxEnt)

Processing the Text Collection

- NRC 1994, 1995; AD 1994,1995
- 4.1 mln sentences, 78 mln words
- Parsed with Alpino, stored dependency trees
 - 4.1 mln XML files
 - 25 Gb
 - 500 CPU days

How to use dependency relations?

- **Wie** kreeg asiel? (*Who was given asylum?*)

- $$Q = \left\{ \begin{array}{l} \langle \text{krijg/K, obj1, asiel/A} \rangle, \\ \langle \text{krijg/K, su, Su/S} \rangle \end{array} \right\}$$

- **Mengistu** kreeg asiel in Zimbabwe (*Mengistu was given asylum by Zimbabwe*)

- $$A = \left\{ \begin{array}{l} \langle \text{krijg/2, su, mengistu/1} \rangle, \\ \langle \text{krijg/2, obj1, asiel/3} \rangle, \\ \langle \text{krijg/2, mod, in/4} \rangle, \\ \langle \text{in/4, obj1, zimbabwe/5} \rangle \end{array} \right\}$$

Syntactic Variation

- Wie verleende asiel aan Mengistu? (*Who gave asylum to Mengistu?*)
- Zimbabwe verleende asiel aan Mengistu (*Zimbabwe gave asylum to Mengistu*)
- Aan Mengistu werd asiel verleend door Zimbabwe (*Asylum was given to Mengistu by Zimbabwe*)

$$\text{equiv}(\{\langle V, \text{su}, \text{Su/S} \rangle\}, \left\{ \begin{array}{l} \langle \text{word/W}, \text{vc}, V \rangle, \\ \langle V, \text{mod}, \text{door/D} \rangle, \\ \langle \text{door/D}, \text{obj1}, \text{Su/S} \rangle \end{array} \right\}).$$

Syntactic Variation

- **Apposition vs Predicate**
 - president van Rusland, Jeltsin \Leftrightarrow
Jeltsin is president van Rusland
- **Genitive vs *van*-PP**
 - Australië's staatshoofd \Leftrightarrow
staatshoofd van Australië
- **Relatives**
 - Ulbricht liet de Berlijnse Muur bouwen \Leftrightarrow
Ulbricht, die de Berlijnse Muur liet bouwen.
- ...

Answer Ranking by Dep Match

- **Wie is de voorzitter van het Europese Parlement?** (*Who is chair of the European Parlement?*)
- **Klaus Hänsch**, voorzitter van het Europese Parlement
- **Karin Junkers** (SPD), lid van het Europese Parlement en voorzitter van de vereniging van sociaal-democratische vrouwen in Europa, (*member of the EP and chair of the organization of soc-dem women in Europe*)

$$\text{match}\left(\left\{\begin{array}{l} \langle \text{voorzitter/V, mod, van/W} \rangle, \\ \langle \text{van/W, obj1, parlement/X} \rangle \end{array}\right\}, Q\right)$$

Answer Ranking by Dep Match

- Wanneer vond de Duitse hereniging plaats? (*When did the German reunion take place?*)
- `date(hereniging)`
- Sinds de Duitse hereniging in oktober 1990 (*Since the German reunion in october 1990*)
- Al in 1962 voorspelde hij de Duitse hereniging (*As early as 1962 he predicted the German reunion*)
- Prefer answers where **date** is a dependent of `hereniging`

Answer Ranking

$$\begin{aligned} \text{a-score}(Q, A) = & \alpha.\text{type-score}(Q, A) & + \\ & \beta.\text{name-score}(Q, A) & + \\ & \gamma.\text{deprel-score}(Q, A) & + \\ & \delta.\text{frequency}(Q, A) & + \\ & \epsilon.\text{IR}(Q, A) \end{aligned}$$

Off-line Extraction

- Exhaustively search corpus for answers to specific question types
 - capital, abbreviations, number of inhabitants, year of birth, ..
- Dependency relations allow complex conditions to be stated
- Equivalences
 - reduce number of required patterns,
 - enhance recall

Founder-Organization

- < oprichten, su, **Founder** >, < oprichten, obj, **Org'n** >
- **Minderop** richtte **de Tros** op toen
- Op last van generaal De Gaulle in Londen richtte verzetsheld **Jean Moulin** in mei 1943 **de Conseil National de la Résistance (CNR)** op.
- **Het Algemeen Ouderen Verbond** is op 1 december opgericht door de nu 75-jarige **Martin Batenburg**.
- toen **de Generale Bank** bekend maakte met de Belgische Post **een "postbank"** op te richten.

Enhancing Recall

Table	-Equiv		+Equiv		Incr (%)	
	tuples	uniq	tuples	uniq	tpls	uniq
Age	15.981	13.716	22.143	18.520	38	35
Born Date	1.545	1.356	2356	1.990	54	47
Capital	1.940	406	2.146	515	10	27
Died Age	522	379	1.127	834	116	120
Founded	604	559	1.021	953	69	70
Function	54.016	28.543	77.028	46.589	43	63

Effect of Enhanced Recall

CLEF 04	off-line		on-line		total	
	# q	mrr	# q	mrr	# q	mrr
baseline	61	0.57	139	0.45	200	0.48
+equiv	71	0.73	129	0.49	200	0.58

Acquisition of ISA relations

- For Named Entities only
- Extracted 300K apposition relations
 - **veerboot** Estonia, Anna Maria Lauro, Sally Star, Herald of Free Enterprise, Prinses Beatrix, ... (112)
 - **bondscoach** Guus Hiddink, Hans Jorritsma, Joop Alberda, Terry Venables, Berti Vogts, ... (2951)

Using ISA relations

- Wh-questions
 - Welke veerboot zonk ten zuidoosten van het eiland Utö? (*Which ferry sank S-E of Utö?*)
 - *Tallinn, Estonia, Raimo Tiilikainen,...*
 - Prefer answer in **ISA** relation to **veerboot**
- Definitions
 - Wat is BMW?
 - BMW **ISA** autofabrikant, include modifying adjectives and PPs
 - **Duitse autofabrikant** (*German car manufacturer*)

Effect of ISA relations

	-isa		+isa	
	# q	CLEF	# q	CLEF
WH 03	58	0.26	58	0.39
WH 04	19	0.26	19	0.47
Definition 04	23	0.17	23	0.44

CLEF results

	# Q	# ok	% ok
Factoid Questions	114	62	54.4
Temporally Restricted Questions	26	7	26.9
Definition Questions	60	30	50.0
gron051n1n1	200	99	49.5
CLEF QA Monolingual Best	200	129	64.5
CLEF QA Monolingual Avg (42 runs)			29.5

Error Analysis

- Parsing Errors
 - oud-minister van [buitenlandse zaken, Solana]
- Echoing the question
 - Where is Bonn located? in Bonn
- Various
 - spelling variation, term variation, spelling errors, location questions, incomplete definitions, ..