ULM-1 The borders of ambiguity

Marten Postma Rubén Izquierdo Piek Vossen

VU University Amsterdam

17-10-2013

Word Sense Disambiguation

Our (1) x **project** (14) x **looks** (14) x **into** (1) x **breaking** (60) x **the** (1) x **borders** (10) x **of** (1) x **ambiguity,** (1) x **for** (1) x **which** (1) x **the** (1) x **queen** (12) x **piece** (18) x **is** (13) x **an** (1) x **example** (6) = 1.981.324.800 interpretations

Traditionally, word sense disambiguation has been defined as: "*the problem of computationally determining which 'sense' of a word is activated by the use of the word in a particular context*"

(Agirre and Edmonds, 2006, p. 1)

Discourse and background knowledge

"The winner will walk away with $1.5-million."

source: `http://www.southafrica.info/news/sport/golf-nedbank-210613.htm#.VEAWkYusVW8`
creation time: **21 June 2013**

Winners [edit]

Year	Player	Country	Score	Winner's share (US$)
2013	Thomas Bjørn	Denmark	268 (–20)	1,250,000
2012	Martin Kaymer	Germany	280 (–8)	1,250,000
2011	Lee Westwood (2)	England	273 (–15)	1,250,000
2010	Lee Westwood	England	271 (–17)	1,250,000
2009	Robert Allenby	Australia	277 (–11)^{PO}	1,200,000
2008	Henrik Stenson	Sweden	267 (–21)	1,200,000
2007	Trevor Immelman	South Africa	272 (–16)	1,200,000
2006	Jim Furyk (2)	United States	276 (–12)	1,200,000

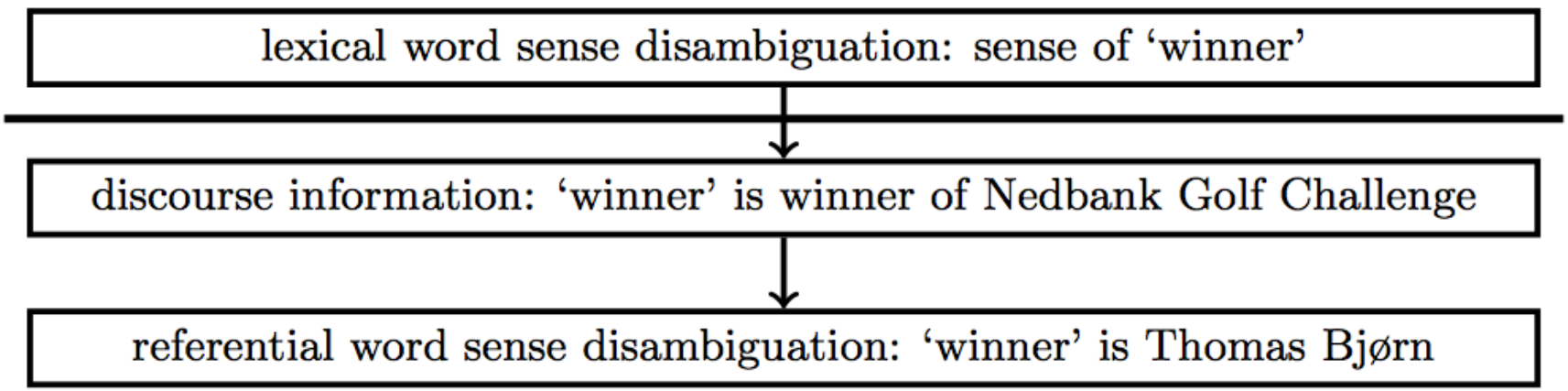
Discourse knowledge and background knowledge

Ideally, we would extract the following facts from this quote:

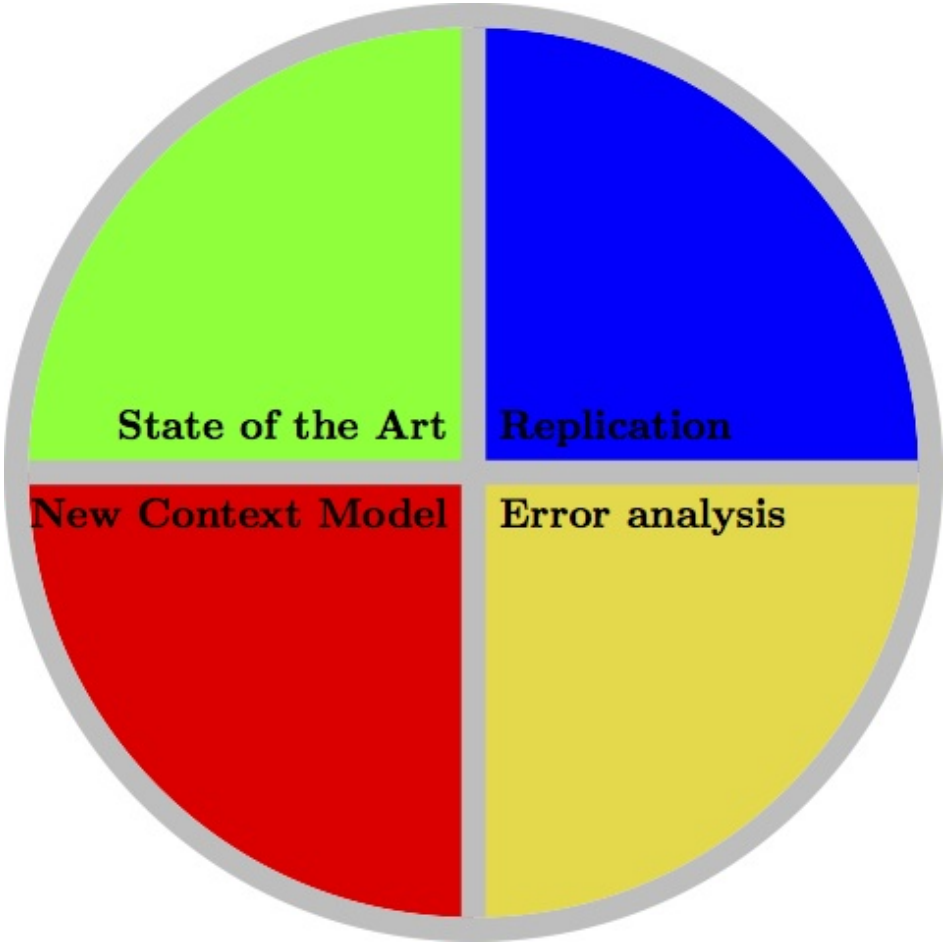
1. **winner** means "the contestant who wins the contest" (wordnet synset identifier ENG30-10782940-n)
2. the **winner** is the winner of the Nedbank Golf Challenge
3. the **winner** was Thomas Bjørn

Borders of ambiguity

Lexical word sense disambiguation



Referential word sense disambiguation


State of the Art
Replication
New Context Model
Error analysis

Classical approaches

- Supervised approaches (annotated data...)
- Knowledge based (too dependent on the resources)
- Unsupervised approaches (low performance)

General trends

- Looks at WSD as a purely classification problem
- Focus more on the low level mathematical algorithm than on the WSD problem itself
- Poor representation of the context, following the idea of"the more features, the better performance"

SensEval/SemEval

Competition	Type	Baseline	Best F1
SensEval2 (2001)	all-words	57.0	69.0 (Sup)
SensEval3 (2004)	all-words	60.9	65.1 (Sup)
SemEval1 (2007)	all-words (task17)	51.4	59.1 (Sup)
SemEval2 (2010)	all-words on specific domain (task17)	50.5	56.2 (Sup)

Replication

The next step in the circle is replication and reproduction:

1. important step
2. difficult
3. helps gain important insights into how other systems work and perform

The role of background knowledge

"One of the best moves by Gary Kasparov which includes a **queen** sacrifice..."

source: `http://www.chess.com/forum/view/chess-players/kasparov-queen-sacrifice`

The role of background knowledge

Output of the It-Makes-Sense System (Zhong and Ng, 2010)

One of the best moves by Gary Kasparov which includes a **queen** sacrifice

- 36% queen.n.1: the only fertile female in a colony of social insects such as bees and ants and termites; its function is to lay eggs
- 34% queen.n.2: a female sovereign ruler
- 30% queen.n.3: the wife or widow of a king
- ...
- **0% queen.n.6: the most powerful chess piece**

The role of background knowledge

Example of how to use background knowledge to solve wsd:

1. compare textual overlap of wiki pages Queen_(chess) and Queen_regnant to wiki page of Garry_Kasparov.
2. 170 types overlapped for Queen_(chess) and only 88 for Queen_regnant

Examples of matching words

board openings matches game press championship rules chess player king queen

Error Analysis of State of the Art

To provide evidence for our hypothesis (poor context modelling), we started doing error analysis on the single outputs of the participants of SensEval and SemEval, paying special attention to:

- Which cases/words/contexts are the most difficult (the most systems fail) and why
- Correlations between linguistic profile of a token (type, polysemy, length of the sentence...) and the performance of the systems on that token
  - Is there any correlation between the polysemy of a word and the performance of the systems?

Error Analysis on SensEval-2

- Errors on monosemous words due to pos-tagger
  - 504 occurrences of monosemous words $\rightarrow$ 33.3% systems are wrong
  - . . . through a **chance** accident in cell division . . .
  - Correct is the sense 1 as ADJECTIVE and all systems assign senses of NOUN
- Errors because of multiwords and phrasal verbs
  - 45 occurrences of multiwords or phrasal verbs, avg error rate 71%
  - Will Quinlan had not inherited a damaged retinoblastoma **supressor gene** and , therefore , faced no more risk. . .
  - Systems provide anwers for gene and not for supressor_gene

Error analysis on SensEval-2

- The more polysemous words are those with higher error rate
  - **make.v** 13 occurences, 49 senses $\rightarrow$ 96% error rate
  - **form.n** 4 ocurrences, 16 senses $\rightarrow$ 100% error rate
- Systems too skewed towards assigning always the most frequent sense (sense 1)
  - 799 tokens where the correct sense is not the most frequent one, in 84% of the cases the systems still assign the most frequent sense
  - In fact one of our goals is to detect and fix the cases where the correct meaning is not the most frequent one

Error analysis on SensEval-2

- The context is not modelled/exploited properly
  - Occurrences of the same word in different contexts have similar error rate
  - The difficulty of a word depends more on the word itself than on the context where it appears

Error analysis on SensEval-2

church.n

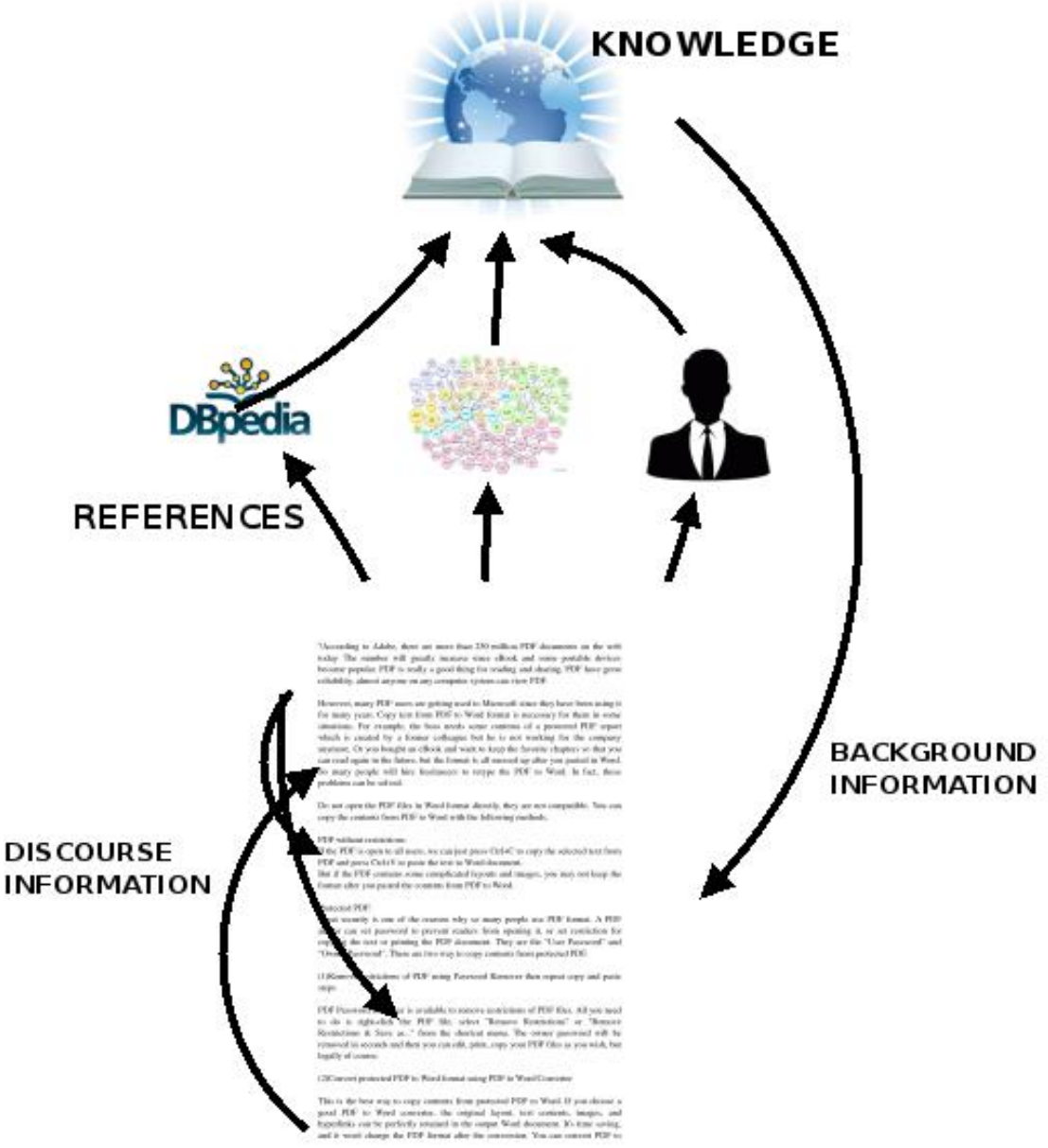
- The parishioners of *St. Michael* and *All Angels* stop to chat at the **church** door, as members here always have. (4.25 avg polysemy, 22 tokens, 81% error rate)
- A look at a Thursday night practice at *St. Mary Abbot* **church** in the Kensington district of London gives an idea of the work involved (8.6 avg polysemy, 26 tokens, 81% error rate)
- The oldest bell-ringing group in the country , the Ancient Society of College Youths , founded in 1637 , remains male-only , a fact that 's particularly galling to women because the group is the sole source of ringers for Britain 's most prestigious **churches**, *St. Paul 's Cathedral* and *Westminster Abbey*. (3.1 avg polysemy, 54 tokens, 85% error)

Error analysis on SensEval-2

new.a $\rightarrow$ 11 senses

- Sentence with avg. polysemy 5.5 and 14 tokens $\rightarrow$ 90.5%
- Sentence with avg. polysemy 5 and 39 tokens $\rightarrow$ 95.5%

Our first approach of our Next Context Model



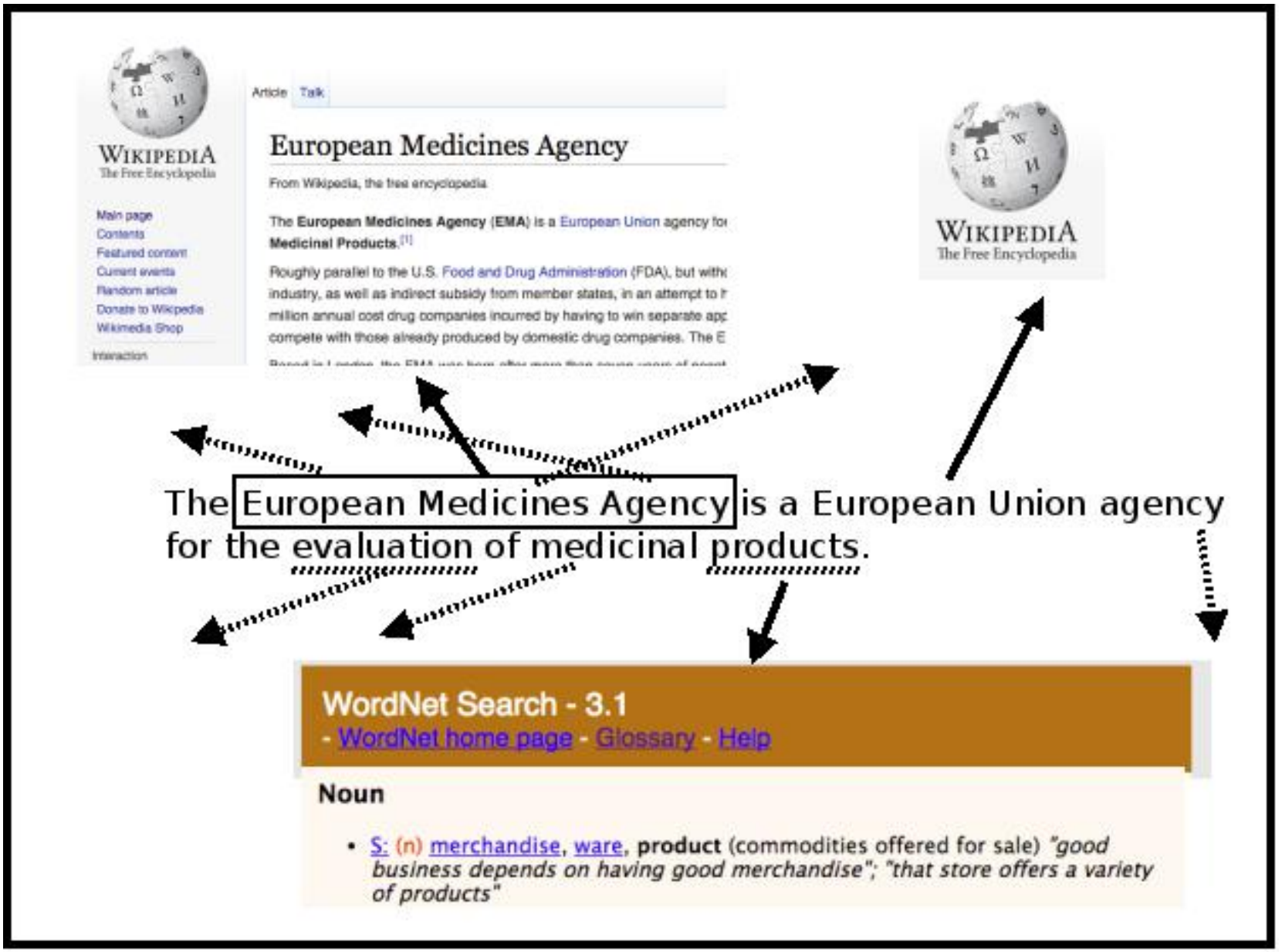
SemEval 2015

SemEval: is an ongoing series of evaluations of computational semantic analysis systems
SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking

The European Medicines Agency is a European Union agency for the evaluation of medicinal products.

SemEval 2015

SemEval 2015



SemEval 2015

- Obtain all possible entities ranked by confidence (NewsReader NERC and NED)
- Obtain all possible senses ranked by confidence (it_makes_sense system, UKB...)
- Obtain coreference sets (Newsreader coreference tool)
- Obtain discourse and background information from these 2 sources to decide on the final entities and meanings

Thank you! Questions?