

Deep Parsing in Watson



Martina Meyer

Technologien für Frage-Antwort-Systeme

WS12/13, Ch. Benzmueller

Der rote Faden

- English Slot Grammar (ESG)
- ESG Parser
- Predicate-Argument Structure (PAS) Builder
- Anpassungen
- Pattern-based relation extraction
- Deep Parsing

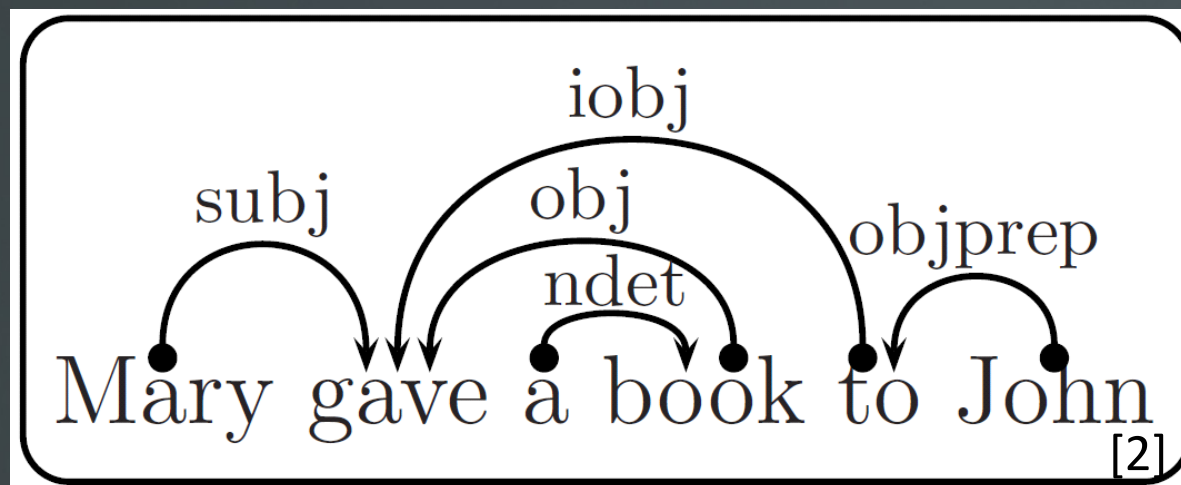
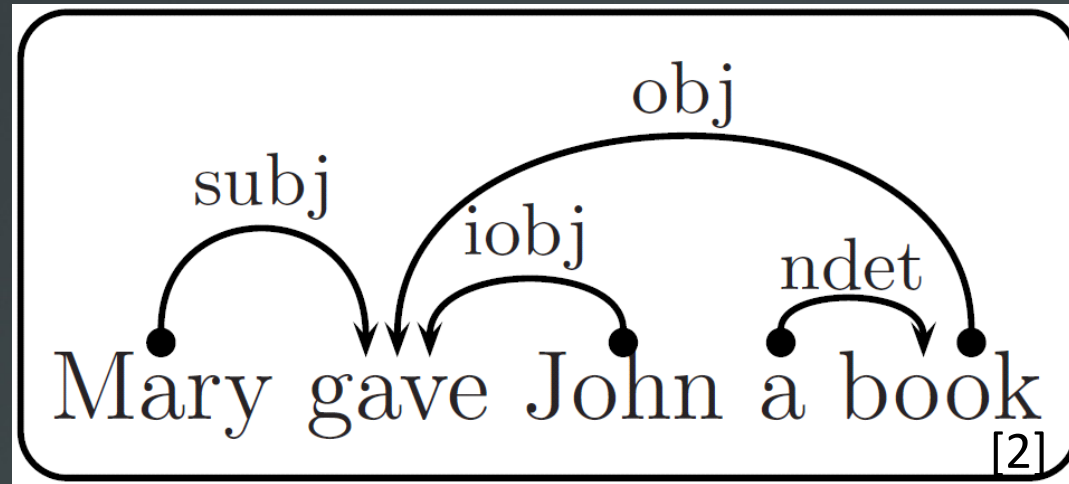


Slot Grammar (SG)

- entwickelt von M. McCord (IBM, ca. 1980)
- jedes Wort hat
 - Slots für obligatorische oder optionale Attribute
 - Filler (konkrete Inhalte, mit denen die Slots gefüllt werden)



English Slot Grammar (ESG)

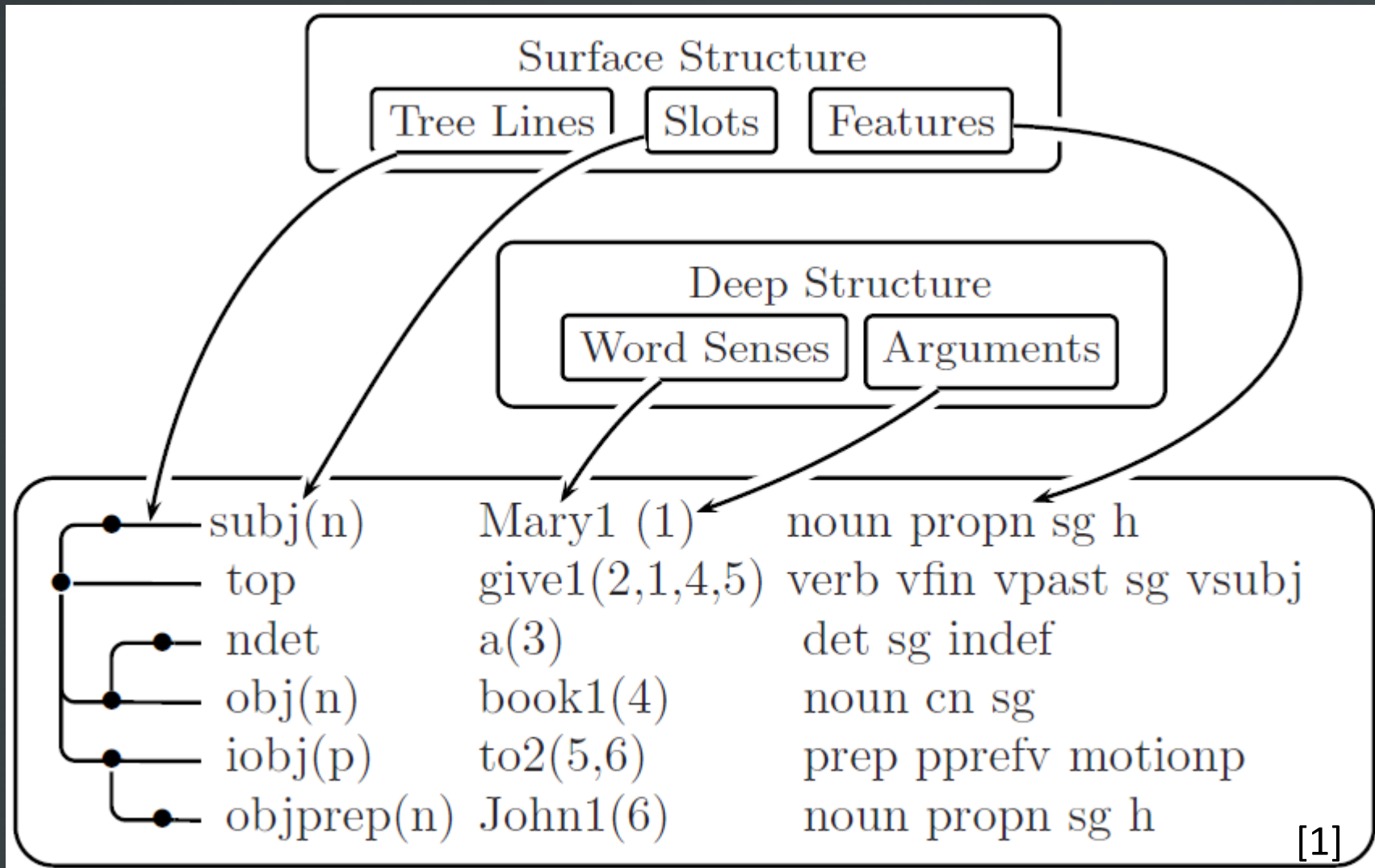


ESG Parser

- Zweiteilig: große Universal-Shell und sprachspezifische Grammatik
- ein einziger Parse Tree zeigt
 - oberflächliche syntaktische Struktur
 - tiefe logische Struktur



ESG Parser



ESG Parser

Jeopardy!™ Clue:

“Chandeliers look great but nowadays do not usually use these items from which their name is derived.”



ESG Parser

Surface Structure

Tree lines

Slots

Features

Deep Structure

Word senses

Arguments



Syntaktische Analyse

- One-word → multiword → bottom-up left-to-right parsing
- Füllen der Slots
- Bewertung findet statt



Syntaktische Analyse

Indexwort

Sense Frames

↓
talk < v (obj n (p about)) (comp (p to with))

↓
< v obj1 (comp1 (p into))

< n nsubj (nobj n (p about)) (ncomp (p to with))



Predicate-Argument Structure (PAS) Builder

- transformiert den Parse Tree
- beschrifteter gerichteter Graph mit Knoten und Kanten entsteht (in **UIMA CAS**)
- geringe Informationsverluste
- Oberflächliche und tiefe Struktur werden zusammengefügt



PAS Builder

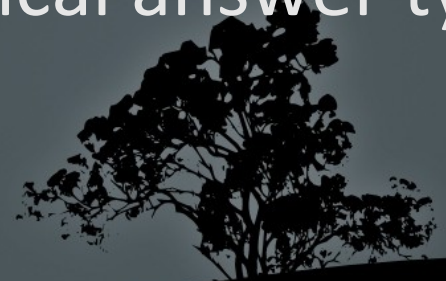
```
chandelier(1)
look(2,1,3)
great(3,1,u)
but(4)
nowadays(5,6)
do(6,1,9)
not(7,6)
usually(8,9)
use(9,1,11,u)
these(10)
item(11,u)
from(12,17,13)
which(13,11,u)
their(14)
name(15,u,u)
be(16,15,17)
derive(17,u,15,12)
```



```
chandelier (1)
look (2, subj:1, comp:3)
great (3)
but (4, lconj:2, rconj:9) [top predicate]
nowadays (5)
not (7)
usually (8)
use (9, subj:1, obj:11, vadv:5, vadv:7,
vadv:8)
item (11, nrel:17) [determiner: these]
from (12, objprep:13)
which (13)
their (14)
name (15, ndet:14) [determiner: their]
derive (17, obj:15, comp:12)
```


Anpassungen

- Syntaktische Analyse
- ESG Wörterbücher
 - einbeziehen von Princeton WordNet
 - Herleitungen von Wikipedia als Chunk Lexicon
- de-verbal nouns (celebration for celebrate)
- Nomen – Verb Ähnlichkeit (celebrant, celebrator)
- Verbesserte Identifizierung von lexical answer types (LAT)



Pattern-based relation extraction

- Erleichtert durch Analyse beim Parsen
- Identifiziert semantische (tiefe) Relationen:
authorOf, actorIn, bornOn etc.
- Problem: Komponenten des Musters im Text
erkennen



Pattern-based relation extraction

Regel:

authorOf :: [Author] [WriteVerb] [Work]

Jeopardy!™ Clue:

“The Lair of the White Worm” and “Dracula”
were written by this Dubliner.



Deep Parsing und der Nutzen für Watson

- ESG Parser und PAS Builder
 - Parsebäume für Watson-Komponenten auf jeder Stufe des DeepQA
- ➔ Deep Parsing stellt den Kern der linguistischen Analyse dar.



Fragen?



Literatur

- [1] “Deep parsing in Watson”, McCord, M. C., Murdock, J. W., Boguraev, B. K., IBM J RES. & DEV. VOL. 56(3/4), 2012
- [2] <http://domino.watson.ibm.com/UsingSlotGrammar>
- [3] <http://uima.apache.org/>

