

Semantic Role Labeling

Am Beispiel von EasySRL und SEMAFOR

Max Depenbrock

Proseminar "Tools für Computerlinguisten"

5.1.2018

Was ist Semantic Role Labeling?

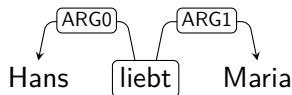
Semantic Role Labeling

Name

- Abgekürzt SRL
- Auch *Shallow Semantic Parsing* genannt

Ziel

- Ereignisse und die daran Beteiligten verstehen
- Who did what to whom?

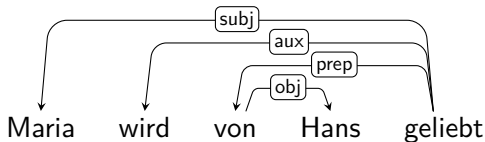
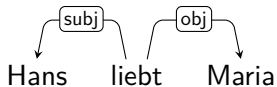


Warum braucht man SRL?

Warum nicht einfach POS-Tags / Phrasenstruktur?

- (1) **John** broke the window
 NE V DET NN
- (2) **John** broke the window with a rock
 NE V DET NN P DET NN
- (3) **The rock** broke the window
 DET NN V DET NN
- (4) **The window** broke
 DET NN V

Warum keine Dependenzbäume?



→ Grammatische Funktionen nicht eindeutig

Warum nicht gleich volles semantisches Parsing?

Satz	PL-Darstellung
Hans liebt Maria	$lieben'(hans', maria')$
Maria wird von Hans geliebt	$lieben'(hans', maria')$
Maria ist verheiratet	$\exists x : verheiratet-mit'(maria', x)$
Jeder Mensch hat einen Fehler	$\forall x : (mensch'(x) \rightarrow \exists y : (fehler'(y) \wedge hat'(x, y)))$

→ Volles semantisches Parsing zu komplex

SRL-Frameworks

SRL-Frameworks

2 verschiedene Frameworks

- Proposition Bank (PropBank)
- FrameNet

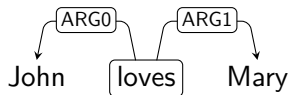
PropBank

Inhalt

- Sätze mit semantischen Rollen
- Penn TreeBank

Aufbau

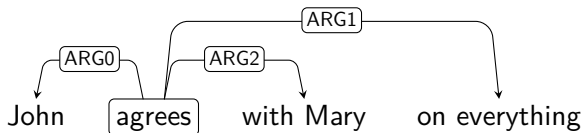
- Verben als Ereignisse
- Verben haben Rollen ARG0 - ARG5
- Genaue Semantik der Rollen in Frame-Files definiert



Frame-Files

agree.01

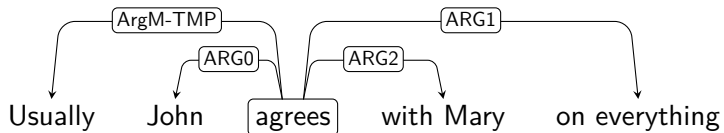
- Arg0: Agreeer
- Arg1: Proposition
- Arg2: Other entity agreeing



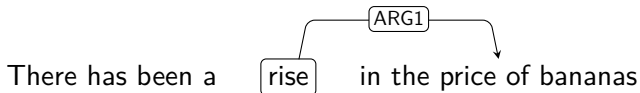
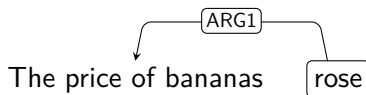
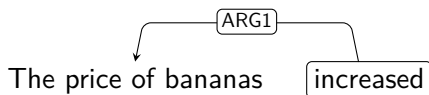
ArgMs

Nicht-nummerierte Adjunkte

- ArgM-TMP: wann?
- ArgM-LOC: wo?
- ArgM-MNR: wie?



Grenzen von PropBank



→ Gleiche Bedeutung, unterschiedliche Realisierung

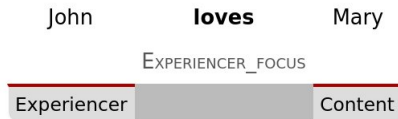
FrameNet

Idee

Wörter mit ähnlicher / gleicher Bedeutung werden zu einem Frame zusammengefasst

Frame

- Semantische Rollen (frame elements)
- Wörter, die ein Frame evozieren



Beispiel für ein Frame

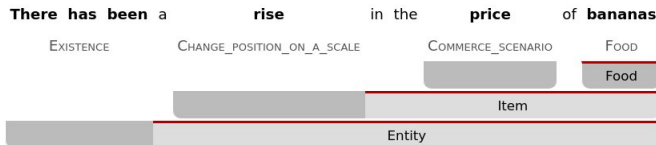
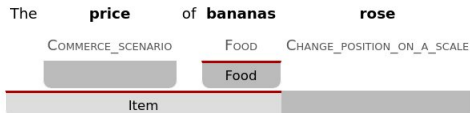
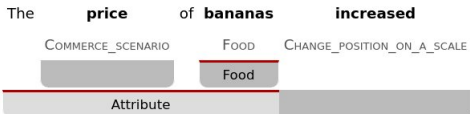
Definition: Change position on a scale

This frame consists of words that indicate the change of an Item's position on a scale (*the Attribute*) from a starting point (*Initial value*) to an end point (*Final value*).

Evoziert durch (u.a.)

- grow
- gain
- rise
- elevation
- elevated

FrameNet und Bananenpreise



PropBank vs. FrameNet

PropBank

Dank weniger Argumente leicht lernbar

FrameNet

Viele Wörter können einem Frame zugeordnet werden

Tools

Preprocessing

- 1 Tokenizer
- 2 POS-Tagger
- 3 {Dependenz, Phrasenstruktur, CCG, ...}-Parser
- 4 SRL

→ Alle Vorverarbeitungsschritte sind sprachabhängig!

Allgemeines

- Entwickelt 2015
- Basierend auf PropBank
- Trainiert mit PropBank-Daten
- Kategorialgrammatik und A*

Kategorialgrammatik (CCG)

chased Mary

VP \rightarrow V NP

NP \rightarrow N

V \rightarrow chased

N \rightarrow Mary

$$\frac{\frac{\textit{chased}}{vp/np} \quad \frac{\textit{Mary}}{np}}{vp} \rightarrow$$

A/B: Mit einem B von rechts wird daraus ein A

Kategorialgrammatik (CCG)

chased Mary

VP → V NP

NP → N

NP → Det N

V → chased

N → Mary

N → cat

Det → The

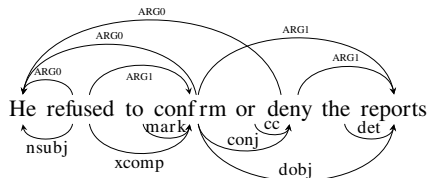
S → NP VP

$$\begin{array}{cccc}
 \textit{the} & \textit{cat} & \textit{chased} & \textit{Mary} \\
 \hline
 \textit{np/n} & \textit{n} & \textit{(s\np)/np} & \textit{np} \\
 \hline
 & & \textit{np} & \textit{s\np} \\
 & & \textit{s} &
 \end{array}
 \begin{array}{l}
 & & & \textit{>} \\
 & \textit{>} & & \\
 & & & \textit{<} \\
 & & &
 \end{array}$$

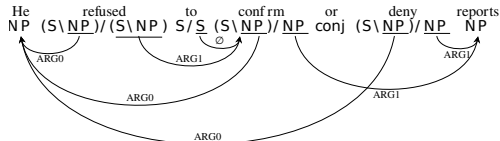
A \ B: Mit einem B von links wird daraus ein A

Motivation für CCG-basiertes SRL

Problem: Dependenzkanten \neq SRL-Kanten



Lösung: Relationen zwischen CCG-Argumenten definieren

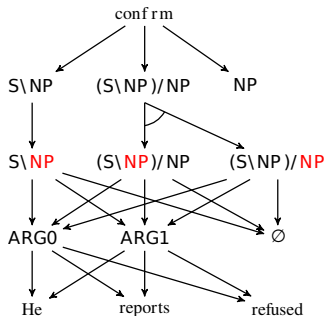


CCG-basiertes SRL – Köpfe

$$\begin{array}{c}
 \text{deny} \qquad \qquad \text{the} \qquad \text{report} \\
 \hline
 (S_{\text{deny}} \setminus NP_x) / NP_y \quad NP_z / N_z \quad N_{\text{report}} \\
 \{\text{deny} \rightarrow x, \text{deny} \rightarrow y\} \quad \{\text{the} \rightarrow z\} \quad \{\} \\
 \hline
 \qquad \qquad \qquad NP_{\text{report}} \\
 \qquad \qquad \qquad \{\text{the} \rightarrow \text{report}\} \\
 \hline
 S_{\text{deny}} \setminus NP_x \\
 \{\text{deny} \rightarrow x, \text{deny} \rightarrow \text{report}, \text{the} \rightarrow \text{report}\}
 \end{array}$$

- Jede Kategorie bekommt einen Kopf / Variable zugeordnet (Subscripte)
- In geschweiften Klammern sind Relationen definiert
- Köpfe und Variablen werden unifiziert

CCG-basiertes SRL – A*



- ① Auswahl einer Repräsentation
- ② Auswahl eines Argumentes
- ③ Auswahl einer Rolle
- ④ Auswahl eines Zieles

Allgemeines

- Entwickelt 2010
- Basierend auf FrameNet
- Trainiert mit PropBank- und SemEval-Daten
- Probabilistisches Modell basierend auf Abhängigkeitsstrukturen

Funktionsweise

1. Target Identification

- Welche Wörter evozieren Frames?
- Regelbasiert

2. Frame Identification

- Zu welchem Frame gehören die Targets?
- Potentiell ambig
- Probabilistisches Modell

3. Argument Identification

- Welche Wörter übernehmen welche Rolle?
- Weiteres probabilistisches Modell

Honorable Mentions

Shalmaneser

- Entwickelt an der UdS
- Deutsche und englische Modelle
- FrameNet-basiert

PathLSTM

- Von ehemaligem Coli-Doktorand entwickelt
- PropBank-basiert
- verwendet neuronale Netze

Übung

Docker-Image

Verzeichnisse mounten

```
docker run -it -v /local/path:/host/path srl  
/bin/bash
```

Verzeichnisstruktur

/tools

- test.txt
- EasySRL/
- semafor/

Folien

<http://depenbrock.no-ip.biz:48080/srl/slides.pdf>

EasySRL

Interaktiver Modus

```
java -jar easysrl.jar --model model --outputFormat  
srl
```

Lesen aus Datei

```
java -jar easysrl.jar --model model --outputFormat  
srl < test.txt
```

Beispielausgabe

```
===Model loaded: parsing...===  
love ARG0 John  
love ARG1 Mary
```

SEMAFOR

RAM zuweisen

```
./autoconfig_ram.sh
```

Lesen aus Datei

```
./bin/runSemafor.sh /tools/test.txt  
/tools/semafor/out.txt 1
```

Ausgabedatei darf beim Ausführen nicht existieren!

Output

```

Finished frame-semantic parsing.
*****

root@6ffe14fd9d84a:/tools/semafor# cat out.txt
{"frames":[{"target":{"name":"Experiencer_focus","spans":[{"start":1,"end":2,"text":"loves"}]},"annotationSets":[{"rank":0,"score":57.69231812749024,"frameElements":[{"name":"Content","spans":[{"start":2,"end":3,"text":"Mary"}]},{"name":"Experiencer","spans":[{"start":0,"end":1,"text":"John"}]}]}], "tokens":["John","loves","Mary"]}
{"frames":[{"target":{"name":"Experiencer_focus","spans":[{"start":1,"end":2,"text":"loves"}]},"annotationSets":[{"rank":0,"score":57.666605576772156,"frameElements":[{"name":"Content","spans":[{"start":2,"end":3,"text":"John"}]},{"name":"Experiencer","spans":[{"start":0,"end":1,"text":"Mary"}]}]}], "tokens":["Mary","loves","John"]}
root@6ffe14fd9d84a:/tools/semafor# python3 prettify_json.py out.txt
Sentence #0
John loves Mary

Experiencer_focus target: loves
  Content: Mary
  Experiencer: John
-----
Sentence #1
Mary loves John

Experiencer_focus target: loves
  Content: John
  Experiencer: Mary
-----
root@6ffe14fd9d84a:/tools/semafor# █

```

Links

- Semafor-Web-Interface:
<http://demo.ark.cs.cmu.edu/parse>
- FrameNet-Suche: https://framenet.icsi.berkeley.edu/fndrupal/framenet_search
- Frame-Übersicht: <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>
- PropBank-Definitionen: <http://verbs.colorado.edu/propbank/framesets-english-aliases/>

Zusammenfassung

Was haben wir heute gesehen?

- SRL zwischen Dependenzanalysen und semantischem Parsing
- Frameworks
 - PropBank (ARG0-ARG5)
 - FrameNet (Wörter in Frames)
- Tools
 - EasySRL (PropBank, CCG)
 - SEMAFOR (FrameNet, prob.)