



Universität Stuttgart



Institut für
Maschinelle
Sprachverarbeitung



**Janis Pagel
Ina Rösiger**

Towards Bridging Resolution in German: Data Analysis and Rule-based Experiments

**CRAC Workshop, New Orleans
June 6, 2018**

Motivation

1

Motivation

- No studies for German data (Except for Hahn, Strube, and Markert (1996))
 - Bridging resolution often only subfield of information status classification
- Close this research gap

Motivation

- No studies for German data (Except for Hahn, Strube, and Markert (1996))
 - Bridging resolution often only subfield of information status classification
- Close this research gap
-
- Available corpora: DIRNDL and GRAIN
 - Corefpro corpus (Grishina, 2016) was not available when study was conducted
- Compare performance on both corpora

Theoretical Background

2

Bridging

- Context-dependent expressions
- Reference inferable from previous discourse

*Bridging antecedents are underlined, bridging anaphors are set in **bold**.

Bridging

- Context-dependent expressions
- Reference inferable from previous discourse

(1) And now you have to be careful that you do not become the voice for the people who just want to avoid the minimum wage. **The main point of contention is the documentation requirement...***

*Bridging antecedents are underlined, bridging anaphors are set in **bold**.

Bridging

- Context-dependent expressions
- Reference inferable from previous discourse

(1) And now you have to be careful that you do not become the voice for the people who just want to avoid the minimum wage. **The main point of contention is the documentation requirement...***

- Definite
- Discourse-new
- But: not coreferent

*Bridging antecedents are underlined, bridging anaphors are set in **bold**.

Data

3

The Corpora

	DIRNDL	GRAIN
Reference	Eckart, Riester, and Schweitzer (2012) and Björkelund et al. (2014)	Eckart and Gärtner (2016) and Schweitzer et al. (2018)
Documents	618 transcribed broadcast <i>news</i>	23 transcribed broadcast <i>interviews</i>
Language	German	German
Information status guidelines	RefLex (Baumann and Riester, 2012)	RefLex (Riester and Baumann, 2017)
Audio	5 hours in total	Around 10 min per interview
Sentences	3,214	2,232
Tokens	49,515	40,430
Markables	16,799	11,578
Bridging anaphors	655	274

GRAIN – Bridging Annotations

- **Prototypical**

Aber jetzt zum Beispiel am Bürokratiewahnsinn in den Heimen, der **den Pflegekräften** die Zeit für **die Patienten** nimmt, ändert sich ja dadurch erst einmal nichts.

But for now, it changes nothing about the bureaucracy madness in the retirement homes, which takes all the time that **the caretakers** could spend on **the patients**.

GRAIN – Bridging Annotations

- **Prototypical**

Aber jetzt zum Beispiel am Bürokratiewahnsinn in den Heimen, der **den Pflegekräften** die Zeit für **die Patienten** nimmt, ändert sich ja dadurch erst einmal nichts.

But for now, it changes nothing about the bureaucracy madness in the retirement homes, which takes all the time that **the caretakers** could spend on **the patients**.

- **World-Knowledge**

[...], dass ich nicht nach Sotschi fahren konnte, obwohl ich als Sportlerin da wirklich sehr, sehr gerne jetzt auch in der neuen Rolle hingefahren wäre, um **die Sportler** zu unterstützen.

[...], that I couldn't go to Sochi, even though I really, really would have liked to go as an athlete and also in my new role, in order to support **the athletes**.

GRAIN – Bridging Annotations

- **Unspecified**

Das ist das größte Reformwerk seit Jahrzehnten in Deutschland. Und kein Wunder, dass es da **am Anfang** ruckelt.

This is the biggest reform in Germany for decades. No wonder that it is unstable **in the beginning**.

GRAIN – Bridging Annotations

- **Unspecified**

Das ist das größte Reformwerk seit Jahrzehnten in Deutschland. Und kein Wunder, dass es da **am Anfang** ruckelt.

This is the biggest reform in Germany for decades. No wonder that it is unstable **in the beginning**.

Type	Proportion
Prototypical	41%
World-Knowledge	10%
Unspecified	45%
Comparative	4%

Table 1

System

4

Rule-based system I

- Based on Hou, Markert, and Strube (2014) and Hou (2016)
- **Lexical rules:**
 - 1: Building parts
The house ... **The basement**
 - 2: Relative persons
She ... **The husband**
 - 3: Geo-political entities
Japan ... **The prime minister**
 - 4: Professional roles
Google ... **The chairman**
 - 5: Percentage NPs
22% of the firms ... **Seventeen percent**
 - 6: Numbers / Indefinite pronouns
Several problems ... **One**

Rule-based system II

- **Argument-taking ratio**

$$\frac{C_{\text{Head-mod}}}{C_{\text{Head}}} \quad (1)$$

- Example with *high* ATR: *husband*
- Example with *low* ATR: *stone*

- **Semantic Connectivity**

$$\text{log-likelihood}(C_{\text{Head-Ana+Head-Ante}}) \quad (2)$$

- Example with *high* SemCon: *door of the house*
- Example with *low* SemCon: *clock of the economy*

Rule-based system III

- Based on Hou, Markert, and Strube (2014) and Hou (2016)
- **Semantic relatedness rules:**
 - 7: Argument-taking NPs I
 - Anaphor: ATR above threshold
 - Antecedent: head occurred as modifier to the anaphor in the same document
 - 8: Argument-taking NPs II
 - A house ... **The door**
 - Anaphor: occurs in subject position, no modifications, ATR above threshold
 - Antecedent: highest SemCon

Rule-based system IV

- **New rules:**
 - 9: Country-related
 - Australia ... **The government**
 - 10: Argument-taking NPs III
 - Like *argument-taking NPs 2* rule but without requirement of being subject
 - 11: Politics NPs
 - The Green Party ... **The party leaders**
 - Extract politics nouns from GermaNet and choose as anaphor, otherwise like rule 10
 - 12: Detect familiar referents
 - Many false positives are generally known referents, such as *The vatican*
 - Only take markables as anaphors which occur once in a document

Experiments

5

Pre-processing

- Extraction of manually annotated gold markables
- **For potential anaphors:**
 - Exclusion of gold coreferent markables (except for first mention)
 - Exclusion of pronouns, proper names, indefinite expressions, NPs with embedded NPs→ Are never labeled as bridging
- **For potential antecedents:**
 - Markables stay unaltered

Hyper-Parameter

- Optimize different hyper-parameters:
 - Maximum sentences distance
 - ATR threshold
 - SemCon threshold
- Dependent on rule

Hyper-Parameter

- Optimize different hyper-parameters:
 - Maximum sentences distance
 - ATR threshold
 - SemCon threshold
- Dependent on rule

- Optimize hyper-parameters on development set
- Development set: combination of official train and devel set
- Evaluate on test set

Baseline

- Anaphor: not modified
 - By adjective
 - By PP or NP
- Antecedent: Subject of previous sentence

Baseline

- Anaphor: not modified
 - By adjective
 - By PP or NP
- Antecedent: Subject of previous sentence

	Precision	Recall	F1
Anaphor Rec.	12.6%	65.1%	21.1%
Bridging Res.	0.5%	2.3%	0.8%

(a) DIRNDL

	Precision	Recall	F1
Anaphor Rec.	15.8%	69.8%	25.9%
Bridging Res.	0.4%	1.6%	0.6%

(b) GRAIN

Table 2

Baseline

- Anaphor: not modified
 - By adjective
 - By PP or NP
- Antecedent: Subject of previous sentence

	Precision	Recall	F1
Anaphor Rec.	12.6%	65.1%	21.1%
Bridging Res.	0.5%	2.3%	0.8%

(a) DIRNDL

	Precision	Recall	F1
Anaphor Rec.	15.8%	69.8%	25.9%
Bridging Res.	0.4%	1.6%	0.6%

(b) GRAIN

Table 2

Baseline

- Anaphor: not modified
 - By adjective
 - By PP or NP
- Antecedent: Subject of previous sentence

	Precision	Recall	F1
Anaphor Rec.	12.6%	65.1%	21.1%
Bridging Res.	0.5%	2.3%	0.8%

(a) DIRNDL

	Precision	Recall	F1
Anaphor Rec.	15.8%	69.8%	25.9%
Bridging Res.	0.4%	1.6%	0.6%

(b) GRAIN

Table 2

Baseline

- Anaphor: not modified
 - By adjective
 - By PP or NP
- Antecedent: Subject of previous sentence

	Precision	Recall	F1
Anaphor Rec.	12.6%	65.1%	21.1%
Bridging Res.	0.5%	2.3%	0.8%

(a) DIRNDL

	Precision	Recall	F1
Anaphor Rec.	15.8%	69.8%	25.9%
Bridging Res.	0.4%	1.6%	0.6%

(b) GRAIN

Table 2

Results for Single Rules

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 4:	4	100.00%	0.0%
Rule 8:	47	48.9%	17.0%
Rule 9:	34	79.4%	64.7%
Rule 10:	113	44.2%	17.7%
Rule 11:	20	50.0%	20.0%
Rule 12:	113	44.2%	17.7%

(a) DIRNDL

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 1:	6	16.6%	16.6%
Rule 4:	2	0.0%	0.0%
Rule 8:	26	38.5%	11.5%
Rule 9:	32	46.9%	40.6%
Rule 10:	37	18.9%	8.1%
Rule 11:	14	7.1%	0.0%
Rule 12:	34	17.6%	8.8%

(b) GRAIN

Table 3

Results for Single Rules

Rule	Fireerate	Anaphor detection Precision	Full bridging resolution Precision
Rule 4:	4	100.00%	0.0%
Rule 8:	47	48.9%	17.0%
Rule 9:	34	79.4%	64.7%
Rule 10:	113	44.2%	17.7%
Rule 11:	20	50.0%	20.0%
Rule 12:	113	44.2%	17.7%

(a) DIRNDL

Rule	Fireerate	Anaphor detection Precision	Full bridging resolution Precision
Rule 1:	6	16.6%	16.6%
Rule 4:	2	0.0%	0.0%
Rule 8:	26	38.5%	11.5%
Rule 9:	32	46.9%	40.6%
Rule 10:	37	18.9%	8.1%
Rule 11:	14	7.1%	0.0%
Rule 12:	34	17.6%	8.8%

(b) GRAIN

Table 3

- Rule 1: Building parts
- Rule 4: Professional roles
- Rule 8: Argument-taking NPs II

Results for Single Rules

Rule	Fireerate	Anaphor detection Precision	Full bridging resolution Precision
Rule 4:	4	100.00%	0.0%
Rule 8:	47	48.9%	17.0%
Rule 9:	34	79.4%	64.7%
Rule 10:	113	44.2%	17.7%
Rule 11:	20	50.0%	20.0%
Rule 12:	113	44.2%	17.7%

(a) DIRNDL

Rule	Fireerate	Anaphor detection Precision	Full bridging resolution Precision
Rule 1:	6	16.6%	16.6%
Rule 4:	2	0.0%	0.0%
Rule 8:	26	38.5%	11.5%
Rule 9:	32	46.9%	40.6%
Rule 10:	37	18.9%	8.1%
Rule 11:	14	7.1%	0.0%
Rule 12:	34	17.6%	8.8%

(b) GRAIN

Table 3

- Rule 9: Country-related

Results for Single Rules

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 4:	4	100.00%	0.0%
Rule 8:	47	48.9%	17.0%
Rule 9:	34	79.4%	64.7%
Rule 10:	113	44.2%	17.7%
Rule 11:	20	50.0%	20.0%
Rule 12:	113	44.2%	17.7%

(a) DIRNDL

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 1:	6	16.6%	16.6%
Rule 4:	2	0.0%	0.0%
Rule 8:	26	38.5%	11.5%
Rule 9:	32	46.9%	40.6%
Rule 10:	37	18.9%	8.1%
Rule 11:	14	7.1%	0.0%
Rule 12:	34	17.6%	8.8%

(b) GRAIN

Table 3

- Rule 10: Argument-taking NPs III
- Rule 12: Detect familiar referents

Results for Single Rules

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 4:	4	100.00%	0.0%
Rule 8:	47	48.9%	17.0%
Rule 9:	34	79.4%	64.7%
Rule 10:	113	44.2%	17.7%
Rule 11:	20	50.0%	20.0%
Rule 12:	113	44.2%	17.7%

(a) DIRNDL

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 1:	6	16.6%	16.6%
Rule 4:	2	0.0%	0.0%
Rule 8:	26	38.5%	11.5%
Rule 9:	32	46.9%	40.6%
Rule 10:	37	18.9%	8.1%
Rule 11:	14	7.1%	0.0%
Rule 12:	34	17.6%	8.8%

(b) GRAIN

Table 3

- Rule 10: Argument-taking NPs III
- Rule 12: Detect familiar referents

Results for Single Rules

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 4:	4	100.00%	0.0%
Rule 8:	47	48.9%	17.0%
Rule 9:	34	79.4%	64.7%
Rule 10:	113	44.2%	17.7%
Rule 11:	20	50.0%	20.0%
Rule 12:	113	44.2%	17.7%

(a) DIRNDL

Rule	Fire rate	Anaphor detection Precision	Full bridging resolution Precision
Rule 1:	6	16.6%	16.6%
Rule 4:	2	0.0%	0.0%
Rule 8:	26	38.5%	11.5%
Rule 9:	32	46.9%	40.6%
Rule 10:	37	18.9%	8.1%
Rule 11:	14	7.1%	0.0%
Rule 12:	34	17.6%	8.8%

(b) GRAIN

Table 3

- Rule 11: Politics NPs

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Results for the Whole System

	Precision	Recall	F1		Precision	Recall	F1
Test set				Test set			
Anaphor Rec.	26.0%	18.9%	21.9%	Anaphor Rec.	45.5%	15.9%	23.5%
Bridging Res.	16.3%	11.6%	13.6%	Bridging Res.	22.7%	7.9%	11.8%
Dev set				Dev set			
Anaphor Rec.	47.6%	19.0%	27.2%	Anaphor Rec.	29.4%	15.2%	20.0%
Bridging Res.	26.7%	10.5%	15.1%	Bridging Res.	17.4%	9.0%	11.9%
Whole set				Whole set			
Anaphor Rec.	39.1%	19.1%	25.6%	Anaphor Rec.	32.1%	15.3%	20.7%
Bridging Res.	22.2%	10.7%	14.4%	Bridging Res.	18.3%	8.8%	11.9%

(a) DIRNDL

(b) GRAIN

Table 4

Oracle Lists

- Investigate if antecedents are in the general scope of the rules
- Change rules to output ranking of antecedents
- Evaluation based on length of oracle list

Oracle Lists

- Investigate if antecedents are in the general scope of the rules
- Change rules to output ranking of antecedents
- Evaluation based on length of oracle list

Without oracle:

Anaphor: Government

Antecedent: Germany

Oracle Lists

- Investigate if antecedents are in the general scope of the rules
- Change rules to output ranking of antecedents
- Evaluation based on length of oracle list

Without oracle:

Anaphor: Government
Antecedent: Germany

With oracle:

Anaphor: Government
Oracle list: [Germany, Poland, Canada]

Oracle Lists

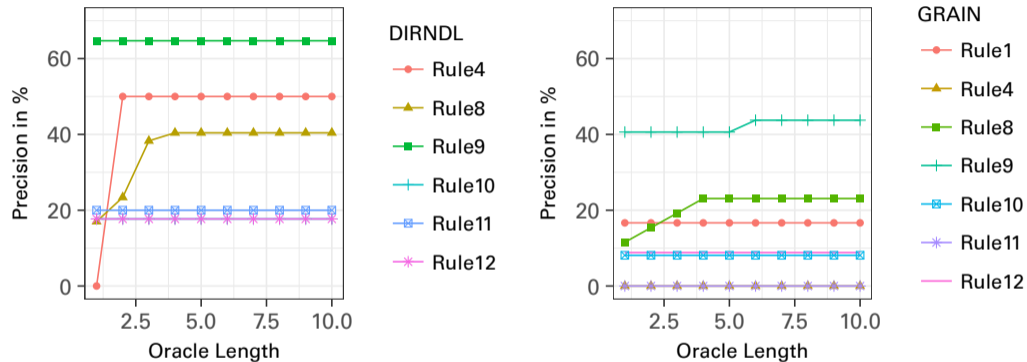


Figure 1

Relative Variable Importance

- Machine-learning did not yield usable results, but:
- Random forest allows for variable importance test
- Investigate which features are actually helpful

Relative Variable Importance

- Machine-learning did not yield usable results, but:
 - Random forest allows for variable importance test
 - Investigate which features are actually helpful
-
- Tested features:
 - Semantic Connectivity
 - Argument-Taking Ratio
 - Character count of anaphor and antecedent
 - Word count of anaphor and antecedent
 - Sentence distance between anaphor and antecedent
 - POS of head of anaphor and antecedent
 - NE class of head of antecedent

Relative Variable Importance

Feature	Variable Importance
SemanticConnectivity	32.2
AnaCharLength	31.6
AnteCharLength	30.5
AnaArgTakingRatio	29.3
AnteWordCount	25.9
AnaWordCount	22.5
SentDist	14.9
AnteHeadPOS	5.9
AnteHeadNE	5.8
AnaHeadPOS	3.3

Table 5

Relative Variable Importance

Feature	Variable Importance
SemanticConnectivity	32.2
AnaCharLength	31.6
AnteCharLength	30.5
AnaArgTakingRatio	29.3
AnteWordCount	25.9
AnaWordCount	22.5
SentDist	14.9
AnteHeadPOS	5.9
AnteHeadNE	5.8
AnaHeadPOS	3.3

Table 5

Relative Variable Importance

Feature	Variable Importance
SemanticConnectivity	32.2
AnaCharLength	31.6
AnteCharLength	30.5
AnaArgTakingRatio	29.3
AnteWordCount	25.9
AnaWordCount	22.5
SentDist	14.9
AnteHeadPOS	5.9
AnteHeadNE	5.8
AnaHeadPOS	3.3

Table 5

Relative Variable Importance

Feature	Variable Importance
SemanticConnectivity	32.2
AnaCharLength	31.6
AnteCharLength	30.5
AnaArgTakingRatio	29.3
AnteWordCount	25.9
AnaWordCount	22.5
SentDist	14.9
AnteHeadPOS	5.9
AnteHeadNE	5.8
AnaHeadPOS	3.3

Table 5

Conclusion

6

Conclusion

- Two corpora for bridging resolution in German: DIRNDL and GRAIN
- Development of new rules
- Rule-based system performs reasonably on DIRNDL and GRAIN
- Oracle list analysis shows that a lot of antecedents are not in the scope of the rules
- Variable importance analysis shows that features like length of the markable, ATR and SemCon are most helpful



Universität Stuttgart



Janis Pagel and Ina Rösiger
Institute for Natural Language Processing, University of Stuttgart

eMail `{pageljs,roesigia}@ims.uni-stuttgart.de`

Supp. `http://www.ims.uni-stuttgart.de/
institut/mitarbeiter/roesigia/
bridging-resolution-german-supplementary.
pdf`

Code `https://github.com/pagelj/publication-code/
tree/master/2018-bridging-resolution-german`

References

7

References I

- Baumann, Stefan and Arndt Rieger (2012). "Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects". In: *Prosody and Meaning*. Ed. by Gorka Elordieta and Pilar Prieto. Vol. 25. Interface of Explorations. Berlin: Mouton de Gruyter, pp. 119–162.
- Björkelund, Anders et al. (2014). "The Extended DIRNDL Corpus as a Resource for Automatic Coreference and Bridging Resolution". In: *Proceedings of the 9th International Conference on Language Resources and Evaluation*. LREC 2014. Reykjavik, pp. 3222–3228.
- Bahill, Aoife and Arndt Rieger (2012). "Automatically Acquiring Fine-Grained Information Status Distinctions in German". In: *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. SIGDIAL 2012, pp. 232–236.
- Eckart, Kerstin and Markus Gärtner (2016). "Creating Silver Standard Annotations for a Corpus of Non-Standard Data". In: *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*. Ed. by Stefanie Dipper, Friedrich Neubarth, and Heike Zinsmeister. Vol. 16. BLA: Bochumer Linguistische Arbeitsberichte. Bochum, Germany, pp. 90–96. URL: https://www.linguistics.rub.de/konvens16/pub/12_konvensproc.pdf.
- Eckart, Kerstin, Arndt Rieger, and Katrin Schweitzer (2012). "A Discourse Information Radio News Database for Linguistic Analysis". In: *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*. Ed. by Christian Chiarcos, Sebastian Nordhoff, and Sebastian Hellmann. Heidelberg: Springer, pp. 65–75.
- Grishina, Yulia (2016). "Experiments on bridging across languages and genres". In: *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016)*. NAACL 2016, pp. 7–15.
- Hahn, Udo, Michael Strube, and Katja Markert (1996). "Bridging Textual Ellipses". In: *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*. COLING '96. Copenhagen, Denmark: Association for Computational Linguistics, pp. 496–501. DOI: 10.3115/992628.992714. URL: <https://doi.org/10.3115/992628.992714>.
- Hou, Yufang (2016). "Unrestricted Bridging Resolution". PhD thesis. Heidelberg University.
- Hou, Yufang, Katja Markert, and Michael Strube (2013a). "Cascading collective classification for bridging anaphora recognition using a rich linguistic feature set". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2013, pp. 814–820.
- (2013b). "Global inference for bridging anaphora resolution". In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. NAACL 2013, pp. 907–917.
- (2014). "A rule-based system for unrestricted bridging resolution: Recognizing bridging anaphora and finding links to antecedents". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2014, pp. 2082–2093.
- Markert, Katja, Yufang Hou, and Michael Strube (2012). "Collective Classification for Fine-grained Information Status". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*. ACL 2012. Jeju Island, Korea, pp. 795–804.
- Markert, Katja, Malvina Nissim, and Natalia N. Modjeska (2003). "Using the Web for Nominal Anaphora Resolution". In: *Proceedings of the EACL Workshop on the Computational Treatment of Anaphora*. EACL 2003, pp. 39–46.
- Markert, Katja, Michael Strube, and Udo Hahn (1996). "Inferential Realization Constraints on Functional Anaphora in the Centering Model". In: *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. CogSci 1996, pp. 609–614.
- Nissim, Malvina (2006). "Learning Information Status of Discourse Entities". In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2006, pp. 94–102.
- Poesio, Massimo, Rahul Mehta, et al. (2004). "Learning to resolve bridging references". In: *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*. ACL 2004, pp. 143–150.

References II

- Poesio, Massimo, Renata Vieira, and Simone Teufel (1997). "Resolving Bridging References in Unrestricted Text". In: *Proceedings of the ACL Workshop on Operational Factors in Robust Anaphora Resolution*, pp. 1–6.
- Rahman, Altaf and Vincent Ng (2011). "Learning the Information Status of Noun Phrases in Spoken Dialogues". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2011, pp. 1069–1080.
- (2012). "Learning the Fine-Grained Information Status of Discourse Entities". In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. EACL 2012, pp. 798–807.
- Riester, Arndt and Stefan Baumann (2017). *The RefLex Scheme – Annotation Guidelines*. SinSpeC. Working papers of the SFB 732 Vol. 14. University of Stuttgart. URL: <https://elib.uni-stuttgart.de/bitstream/11682/9028/1/RefLex-SinSpec14.pdf>.
- Rösiger, Ina and Simone Teufel (2014). "Resolving Coreferent and Associative Noun Phrases in Scientific Text". In: *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics*. EACL 2014. Gothenburg, Sweden: Association for Computational Linguistics, pp. 45–55. URL: <http://aclanthology.coli.uni-saarland.de/pdf/E/E14/E14-3006.pdf>.
- Schweitzer, Katrin et al. (2018). "German Radio Interviews: The GRAIN Release of the SFB732 Silver Standard Collection". In: *Proceedings of the 11th International Conference on Language Resources and Evaluation*. LREC 2018.

Appendix

9

Types of Bridging in GRAIN

Type	Sub-type	Count	Proportion
Prototypical	Building-part	3	1%
	Professional role	1	<1%
	Country-related	19	8%
	Other prototypical	69	31%
World-Knowledge		23	10%
Unspecified		101	45%
Comparative		8	4%

Table 6

Results for the Rules

Rule	Anaphor detection			Full bridging resolution		
	Correct	Wrong	Precision	Correct	Wrong	Precision
Rule 4:	4	0	100.00%	0	4	0.0%
Rule 8:	23	24	48.9%	8	39	17.0%
Rule 9:	27	7	79.4%	22	12	64.7%
Rule 10:	50	63	44.2%	20	93	17.7%
Rule 11:	10	10	50.0%	4	16	20.0%
Rule 12:	50	63	44.2%	20	93	17.7%

(a) DIRNDL

Rule	Anaphor detection			Full bridging resolution		
	Correct	Wrong	Precision	Correct	Wrong	Precision
Rule 1:	1	5	16.6%	1	5	16.6%
Rule 4:	0	2	0.0%	0	2	0.0%
Rule 8:	10	16	38.5%	3	23	11.5%
Rule 9:	15	17	46.9%	13	19	40.6%
Rule 10:	7	30	18.9%	3	34	8.1%
Rule 11:	1	13	7.1%	0	14	0.0%
Rule 12:	6	28	17.6%	3	31	8.8%

(b) GRAIN

Table 7

Results – Oracle Lists – Overall

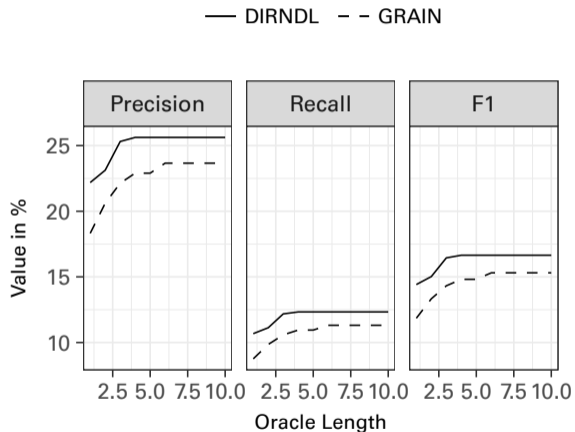


Figure 2

Related Work

- **Early bridging resolution:**
 - Markert, Strube, and Hahn (1996), Poesio, Vieira, and Teufel (1997), Markert, Nissim, and Modjeska (2003), and Poesio, Mehta, et al. (2004)
- **Bridging anaphor detection within information status classification:**
 - Nissim (2006), Rahman and Ng (2011), Rahman and Ng (2012), Cahill and Riester (2012), Markert, Hou, and Strube (2012), and Hou, Markert, and Strube (2013a)
- **Bridging resolution within coreference resolution:**
 - Rösiger and Teufel, 2014
- **Unrestricted bridging resolution:**
 - Hou, Markert, and Strube, 2013b; Hou, Markert, and Strube, 2014