Information Status in Generation Ranking

Aoife Cahill
joint work with Arndt Riester

Heidelberg Computational Linguistics Colloquium
December 9, 2010

Outline

Outline

Outlining the problem

German is considered a relatively “free word order” language (with a rich case system)

Notion dates from a time when discourse information did not play much of a role in linguistics

Our task: generating German strings from LFG F-structures

The problem: how to choose the “best” string from the many grammatical strings output by the system?

Surface Realisation System

Lexical Functional Grammar F-Structure – Basic predicate argument structure

```
"Die Nato werde nicht von der EU geführt."

    PRED       'führen<[249:von], [21:Nato]>'
               PRED  'Nato'
               CHECK   _SPEC-TYPE [_COUNT +, _DEF +, _DET attr]
                       _INFL      strong-det
    SUBJ       NTYPE  [NSYN proper]
               SPEC   DET [PRED     'die']
                          [DET-TYPE def  ]
             21 CASE nom, GEND fem, NUM sg, PERS 3
               PRED   'von<[283:EU]>'
                      PRED  'EU'
                      CHECK   _SPEC-TYPE [_COUNT +, _DEF +, _DET attr]
                              _INFL      strong-det
    OBL-AG     OBJ    NTYPE  [NSYN proper]
                      SPEC   DET [PRED     'die']
                                 [DET-TYPE def  ]
                  283 CASE dat, GEND fem, NUM sg, PERS 3
            249 PSEM dir, PTYPE sem
    ADJUNCT    { [PRED          'nicht'] }
               { 215[ADJUNCT-TYPE neg   ] }
               _AUX-FORM <werden-pass_>
    CHECK      _VLEX      [_AUX-SELECT sein]
               _VMORPH    [_PARTICIPLE perfect]
    TNS-ASP    [MOOD subjunctive, PASS-SEM dynamic_, TENSE pres]
    TOPIC      [21:Nato]
128 CLAUSE-TYPE decl, PASSIVE +, STMT-TYPE decl, VTYPE main
```

Surface Realisation System

Hand Crafted Large-Scale Grammar (Rohrer and Forst, 2006) generates all possible (grammatical) strings.

’NATO is not led by the EU.’

Die Nato werde von der EU nicht geführt.	Die Nato werde nicht von der EU geführt.
Nicht von der EU geführt werde die Nato.	Nicht werde von der EU die Nato geführt.
Nicht werde die Nato von der EU geführt.	Nicht geführt werde von der EU die Nato.
Nicht geführt werde die Nato von der EU.	Von der EU nicht geführt werde die Nato.
Von der EU werde die Nato nicht geführt.	Von der EU werde nicht die Nato geführt.
Von der EU geführt werde nicht die Nato.	Von der EU geführt werde die Nato nicht.
Geführt werde die Nato nicht von der EU.	Geführt werde die Nato von der EU nicht.
Geführt werde nicht von der EU die Nato.	Geführt werde nicht die Nato von der EU.
Geführt werde von der EU nicht die Nato.	Geführt werde von der EU die Nato nicht.

Surface Realisation System (Cahill et al., 2007)

Log-linear ranking model chooses most likely string

Linguistically Motivated Feature Types

1. C-structure	number of NPs, number of children of PP
2. C- & F-Structure	SUBJ precedes OBJ
3. Language Model	tri-gram score

Outperforms a basic tri-gram language model, but can be further improved

Surface Realisation System (Cahill et al., 2007)

Log-linear ranking model chooses most likely string

Linguistically Motivated Feature Types

1. C-structure	number of NPs, number of children of PP
2. C- & F-Structure	SUBJ precedes OBJ
3. Language Model	tri-gram score

Outperforms a basic tri-gram language model, but can be further improved

Idea: Capturing the influence of discourse information can help choose the best string

Outline

Information Status (IS) (Prince 1981,1992)

- Means of discourse analysis
- Classifying (NP/PP/DP) constituents according to their givenness
- IS is marked in prosody (Baumann, 2006; Schweitzer et al., 2009) as well as in syntax
- Corpus of German news texts manually annotated for IS
- Advantages with regard to earlier IS work:
  - proper treatment of embedded phrases
  - higher inter-annotator agreement on difficult texts
  - closer to insights from semantic theory (e.g. semantic presuppositions)

IS Labels: Riester, Lorenz, Seemann (2010)

Full	Collapsed
BRIDGING BRIDGING-CONTAINED	BRIDGING
CATAPHOR	CATAPHOR
EXPLETIVE	EXPLETIVE
GIVEN-EPITHET GIVEN-PRONOUN GIVEN-REFLEXIVE GIVEN-REPEATED GIVEN-SHORT	GIVEN
INDEF-GENERIC INDEF-NEW INDEF-PARTITIVE INDEF-PARTITIVE-CONTAINED INDEF-RESUMPTIVE	INDEF
NULL	NULL
RELATIVE	RELATIVE
SITUATIVE	SITUATIVE
UNUSED-KNOWN UNUSED-TYPE UNUSED-UNKNOWN	UNUSED

Most Important Classes

GIVEN	coreferential anaphor	Merkel ... $\boxed{\text{sie}}$
BRIDGING	non-coreferential but context dependent expression	Stuttgart ... $\boxed{\text{der Bahnhof}}$
UNUSED-KNOWN	discourse new, familiar definite	$\boxed{\text{der Mond}}$
UNUSED-UNKNOWN	discourse new, unfamiliar definite	$\boxed{\text{das neue Gesetz}}$ $\boxed{\text{zur Gesundheitsreform}}$
SITUATIVE	deictic expression	$\boxed{\text{am Dienstag}}$
INDEF	indefinite	$\boxed{\text{einige hundert}}$ $\boxed{\text{Menschen}}$

Grammaticality and markedness

Two grammatical sentences

'The army has even been able to recapture smaller territories.'

(1) Die Armee habe sogar kleinere Gebiete zurückerobern können. (ok)

(2) Kleinere Gebiete habe die Armee sogar zurückerobern können. (strongly marked)

A sentence is marked precisely if there are only few or very special contexts in which it is appropriate

Capturing context

Information status reflects context to a certain degree

IS labels taken from corpus

'The army has even been able to recapture smaller territories.'

(3) $\boxed{\textit{Die Armee}}$ GIVEN-EPITHET *habe sogar* $\boxed{\textit{kleinere Gebiete}}$ INDEF-NEW *zurückerobern können.*

The givenness/novelty of an expression characterise the class of contexts in which the expression can occur

Compute the preferred order for each pair of IS labels

Precedence of label pairs within a clause

X before Y (e.g. BRIDGING before UNUSED-UNKNOWN)

Die Gespräche $_{\text{BRIDGING}}$ sollen heute

in Jerusalem $_{\text{UNUSED-KNOWN}}$ fortgesetzt werden.

'The talks shall be continued in Jerusalem today.'

Occurrences in corpus: 49

Precedence of label pairs within a clause

X before Y (e.g. BRIDGING before UNUSED-UNKNOWN)

Die Gespräche $_{\text{BRIDGING}}$ sollen heute in Jerusalem $_{\text{UNUSED-KNOWN}}$ fortgesetzt werden.

'The talks shall be continued in Jerusalem today.'

Occurrences in corpus: 49

Y before X (e.g. UNUSED-UNKNOWN before BRIDGING)

So müsse dies die britische Regierung $_{\text{UNUSED-KNOWN}}$ den Bürgern $_{\text{BRIDGING}}$ klarmachen.

'Thus, the British Government should make this clear to the citizens.'

Occurrences in corpus: 81

Precedence of label pairs within a clause

X before Y (e.g. BRIDGING before UNUSED-UNKNOWN)

[Die Gespräche] BRIDGING sollen heute [in Jerusalem] UNUSED-KNOWN fortgesetzt werden.
'The talks shall be continued in Jerusalem today.'

Occurrences in corpus: 49   less prominent order *B*

Y before X (e.g. UNUSED-UNKNOWN before BRIDGING)

So müsse dies [die britische Regierung] UNUSED-KNOWN [den Bürgern] BRIDGING klarmachen.
'Thus, the British Government should make this clear to the citizens.'

Occurrences in corpus: 81   dominant order *A*

Defining a measure

Asymmetry ratio

A (Dominant order)	B	Asym. ratio B/A	Total
81	49	0.604	130

Compute asymmetry ratio for each pair of IS labels.

Asymmetry tables (top)

Dominant order	Asym. ratio	Freq
UNUSED-KNOWN before CATAPHOR	0.05	22
GIVEN-REPEATED before UNUSED-TYPE	0.1	11
GIVEN-PRONOUN before SITUATIVE	0.13	26
GIVEN-REFLEXIVE before INDEF-NEW	0.14	56
GIVEN-PRONOUN before CATAPHOR	0.15	23
GIVEN-PRONOUN before INDEF-NEW	0.19	142
BRIDGING before INDEF-GENERIC	0.2	12
GIVEN-SHORT before GIVEN-REPEATED	0.2	12
GIVEN-PRONOUN before UNUSED-TYPE	0.21	35
GIVEN-REFLEXIVE before UNUSED-TYPE	0.22	11
GIVEN-EPITHET before UNUSED-TYPE	0.23	27
UNUSED-KNOWN before UNUSED-TYPE	0.24	78
EXPLETIVE before INDEF-NEW	0.25	50

The crucial problem

IS is an indicator for constituent order, but ...

there is no reliable automatic annotation system for IS

The crucial problem

IS is an indicator for constituent order, but . . .

there is no reliable automatic annotation system for IS

First Attempt (Cahill and Riester, 2009): use morphosyntactic features correlated with IS

Outline

Syntactic Features

We define an inventory of syntactic features that can appear under all IS labels and automatically mark up the corpus with them. The features include:

- is simple definite
- is simple definite description with a possessive modifier
- is definite description with adjectival modifier
- is definite description with a genitive argument
- is definite description with an (obligatory/referentially restricting) PP adjunct
- is definite description including a relative clause
- is definite description including an embedded proper name and (perhaps) a title or job description
- is a combination of position/title and proper name (without article)
- is a bare proper name
- ...

Morphosyntactic correlates of IS

Some IS categories directly derive from syntactic classes (1:1 correspondence)

GIVEN-REFLEXIVE

Is a reflexive pronoun (all items)

EXPLETIVE

Is an expletive, e.g. 'es' (all items)

Morphosyntactic correlates of IS

Some IS categories are represented by various features

UNUSED-KNOWN

feature	items	example
Is a simple definite	145	the moon
Is a name with a title	55	President Obama
Is a bare noun	54	Africa
Is definite with apposition	36	the German Chancellor, Angela Merkel

...

Syntactic Features and IS phrases

Extracting information from the corpus

We have a corpus that is:

- annotated with IS labels
- marked up with syntactic features

For each phrase annotated with an IS label, look at what syntactic features are present

Collect statistics for each IS label type

Syntactic Features associated with IS labels

GIVEN-PRONOUN

Syn. Feat	Count
IS_PERS_PRON	88
IS_DA_PRON	56
IS_DEMON_PRON	41
IS_GENERIC_PRON	16

Syntactic Features associated with IS labels

GIVEN-PRONOUN

Syn. Feat	Count
IS_PERS_PRON	88
IS_DA_PRON	56
IS_DEMON_PRON	41
IS_GENERIC_PRON	16

INDEF-NEW

Syn. Feat	Count
IS_SIMPLE_INDEF	203
IS_INDEF_ATTR	95
IS_INDEF_NUM	85
IS_INDEF_GENARG	20
IS_INDEF_PPADJUNCT	19
...	

IS asymmetries with syntactic features

Label 1		Label 2		Ratio	Freq.
UNUSED-KNOWN		**CATAPHOR**		0.05	22
IS_BAREPROPER	166	IS_SIMPLE_DEF	14		
IS_SIMPLE_DEF	102	IS_DA_PRON	13		
IS_PROPER	85				
GIVEN-REPEATED		**UNUSED-TYPE**		0.1	11
IS_SIMPLE_DEF	28	IS_SIMPLE_DEF	37		
IS_BAREPROPER	23	IS_SIMPLE_INDEF	36		
GIVEN-PRONOUN		**SITUATIVE**		0.13	26
IS_PERS_PRON	88	IS_TEMP_ADV	62		
IS_DA_PRON	56	IS_SIMPLE_DEF	44		
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23		
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19		
		...			

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

IS_PERS_PRON precedes IS_TEMP_ADV

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

IS_PERS_PRON precedes IS_TEMP_ADV
IS_PERS_PRON precedes IS_SIMPLE_DEF

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

IS_PERS_PRON precedes IS_TEMP_ADV
IS_PERS_PRON precedes IS_SIMPLE_DEF
IS_PERS_PRON precedes IS_DEF_ATTR_ADJUNCT

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

IS_PERS_PRON precedes IS_TEMP_ADV
IS_PERS_PRON precedes IS_SIMPLE_DEF
IS_PERS_PRON precedes IS_DEF_ATTR_ADJUNCT
IS_PERS_PRON precedes IS_SIMPLE_INDEF

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

IS_PERS_PRON precedes IS_TEMP_ADV
IS_PERS_PRON precedes IS_SIMPLE_DEF
IS_PERS_PRON precedes IS_DEF_ATTR_ADJUNCT
IS_PERS_PRON precedes IS_SIMPLE_INDEF
IS_DA_PRON precedes IS_TEMP_ADV

New Features

From each IS asymmetry extract precedence patterns of corresponding syntactic features

GIVEN-PRONOUN		SITUATIVE	
IS_PERS_PRON	88	IS_TEMP_ADV	62
IS_DA_PRON	56	IS_SIMPLE_DEF	44
IS_DEMON_PRON	41	IS_DEF_ATTR_ADJUNCT	23
IS_GENERIC_PRON	16	IS_SIMPLE_INDEF	19

IS_PERS_PRON precedes IS_TEMP_ADV
IS_PERS_PRON precedes IS_SIMPLE_DEF
IS_PERS_PRON precedes IS_DEF_ATTR_ADJUNCT
IS_PERS_PRON precedes IS_SIMPLE_INDEF
IS_DA_PRON precedes IS_TEMP_ADV
IS_DA_PRON precedes IS_SIMPLE_DEF
IS_DA_PRON precedes IS_DEF_ATTR_ADJUNCT
IS_DA_PRON precedes IS_SIMPLE_INDEF
IS_DEMON_PRON precedes IS_TEMP_ADV

...

Improved Generation Ranking Model

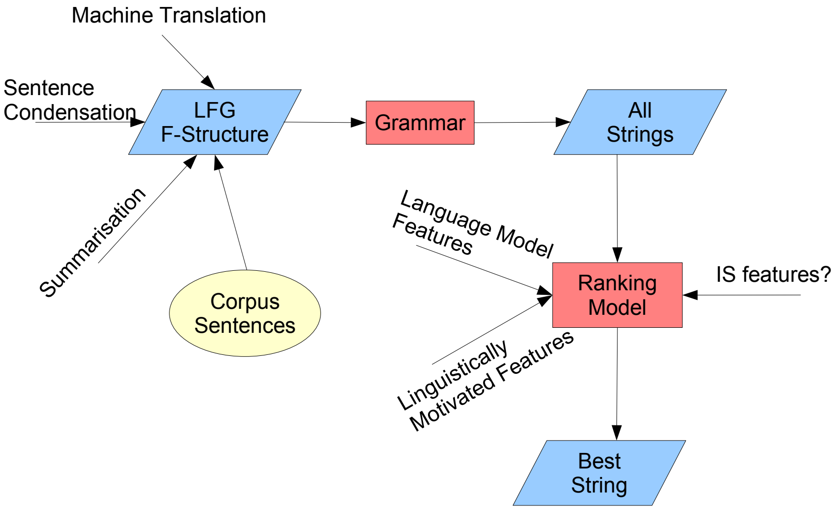
We include these new features in our svm model for generation ranking

Feature Types

1. C-structure	number of NPs, number of children of PP
2. C- & F-Structure	SUBJ precedes OBJ
3. Language Model	tri-gram score
4. IS asymmetric syntactic patterns	IS_PERS_PRON precedes IS_TEMP_ADV

Outline

System Overview



Experimental Setup

Experiment

Train svm ranking model on 7161 syntactically annotated sentences from TIGER

Tune model parameters on development set of 55 sentences

Carry out final evaluation on test set of 260 sentences

Results

Evaluation on 260 sentences
BLEU measures string similarity using ngrams

Slightly different to Cahill and Riester (2009):

- Uses SVM rank instead of log-linear model
- asymmetries calculated from more data
- ...but same features

Results

Evaluation on 260 sentences
BLEU measures string similarity using ngrams

Slightly different to Cahill and Riester (2009):

- Uses SVM rank instead of log-linear model
- asymmetries calculated from more data
- . . . but same features

	BLEU	Exact Match (%)
Baseline	0.7691	50.00
IS Approx	0.7797	51.66

Results

Evaluation on 260 sentences
BLEU measures string similarity using ngrams

Slightly different to Cahill and Riester (2009):

- Uses SVM rank instead of log-linear model
- asymmetries calculated from more data
- ...but same features

	BLEU	Exact Match (%)
Baseline	0.7691	50.00
IS Approx	0.7797	51.66

Statistically significant improvement with model including new IS-inspired syntactic features

Example Sentences

'We have learnt from the scandal'

Gold

Man hat aus der Affäre gelernt.
One has from the scandal learnt.

Example Sentences

'We have learnt from the scandal'

Gold

Man hat aus der Affäre gelernt.
One has from the scandal learnt.

Baseline

Aus der Affäre hat man gelernt.
From the scandal has one learnt.

Example Sentences

'We have learnt from the scandal'

Gold

Man hat aus der Affäre gelernt.
One has from the scandal learnt.

Baseline

Aus der Affäre hat man gelernt.
From the scandal has one learnt.

New

Man hat aus der Affäre gelernt.
One has from the scandal learnt.

Outline

Predicting Information Status?

We showed that for realisation ranking, the approximation of the morpho-syntactic features of the information status labels helped

But what if we could automatically label raw text with information status labels?

Supervised Learning Task

Given a corpus of manually annotated radio news

- 3454 sentences
- remove duplicates
- divide into ∼10% development (129 sentences), ∼90% training/test (1169 sentences)
- parse with XLE German grammar

Task: sequence labelling

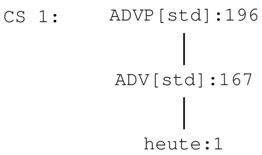
Model: Conditional Random Field

Designed Features to capture the “basic geometry of the expressions”

Capturing the Geometry of Expressions

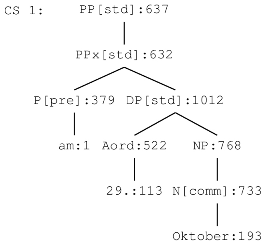
SITUATIVE

1. heute



SITUATIVE

2. am 29. oktober



Capturing the Geometry of Expressions

GIVEN-SHORT

10. meisner

```
CS 1:   NP:316
          |
        NAMEP:297
          |
        NAME:295
          |
        Meisner:1
```

GIVEN-PRONOUN

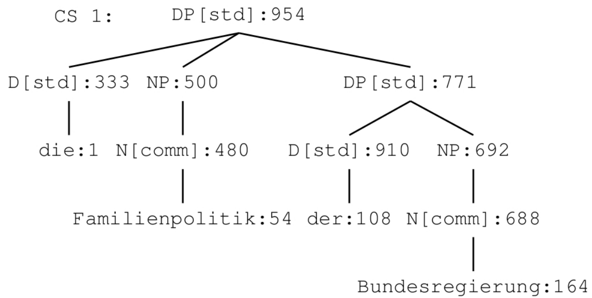
4. sie

```
CS 1:   DP[std]:249
          |
        PRON[std]:247
          |
        sie:4
```

Capturing the Geometry of Expressions

BRIDGING-CONTAINED

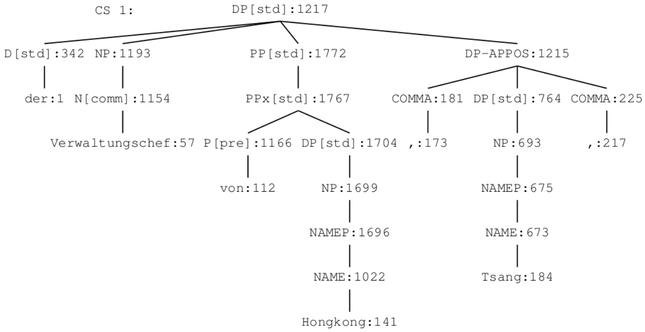
2. die familienpolitik der bundesregierung



Capturing the Geometry of Expressions

UNUSED-UNKNOWN

1. der verwaltungschef von hongkong , tsang ,



Model Features I

Starting Point

Morpho-syntactic features from previous work

Things we count

Words

Specific syntactic categories: DP, NP, DP-APPOSS, LABELP, NAMEP, YEAR, A-CARD

Children of the top category

Maximum path length from top node to POS tags

N-ary branching nodes ($n > 1$)

Model Features II

Binary Features

Coordination
Coreferent
More than 1 DP and NP
Pronoun
First/Last label in the sentences

Other Features

Determiner type (definite, indefinite, unknown)
Syntactic category of the top-most node dominating the string
Syntactic function of the substring
POS tag at left/right edge of the substring

Evaluation

Carry out 10-fold cross validation on our test/train data (1169 sentence, 3705 labels)

Evaluate on both sets of labels: full (20) and collapsed (9)

Three Baselines:

1. Randomly assign a label to each phrase
2. Always assign the most frequent label to each phrase
3. Informed: assign the most frequent label, given the morpho-syntactic features from previous experiments

Evaluation

Carry out 10-fold cross validation on our test/train data (1169 sentence, 3705 labels)

Evaluate on both sets of labels: full (20) and collapsed (9)

Three Baselines:

1. Randomly assign a label to each phrase
2. Always assign the most frequent label to each phrase
3. Informed: assign the most frequent label, given the morpho-syntactic features from previous experiments

Accuracy (%)	Full	Collapsed
Random	5.45	11.10
Most Frequent	17.65	32.31
Informed	47.98	65.26

CRF Model Prediction Results

Accuracy (%)	Full	Collapsed
Random	5.45	11.10
Most Frequent	17.65	32.31
Informed	47.98	65.26
CRF	64.87	81.65

16.89% increase in full label set accuracy,
16.39% increase on collapsed set accuracy

Detailed CRF Prediction Results

Label	Total	Precision	Recall	F-Score
BRIDGING	511	0.591	0.507	0.546
CATAPHOR	43	0.667	0.233	0.345
EXPLETIVE	73	1.000	1.000	1.000
GIVEN	768	0.959	0.993	0.976
INDEF	941	0.854	0.960	0.904
NULL	1	0.000	0.000	0.000
RELATIVE	7	1.000	0.857	0.923
SITUATIVE	164	0.759	0.518	0.616
UNUSED	1197	0.767	0.774	0.770

High level prediction could be used to suggest possible labels to annotators and possibly speed up the manual annotation process

Detailed CRF Prediction Results

Label	Total	Precision	Recall	F-Score
BRIDGING	511	0.591	0.507	0.546
CATAPHOR	43	0.667	0.233	0.345
EXPLETIVE	73	1.000	1.000	1.000
GIVEN	768	0.959	0.993	0.976
INDEF	941	0.854	0.960	0.904
NULL	1	0.000	0.000	0.000
RELATIVE	7	1.000	0.857	0.923
SITUATIVE	164	0.759	0.518	0.616
UNUSED	1197	0.767	0.774	0.770

High level prediction could be used to suggest possible labels to annotators and possibly speed up the manual annotation process

Detailed CRF Prediction Results

Label	Total	Precision	Recall	F-Score
BRIDGING	262	0.530	0.607	0.566
BRIDGING-CONTAINED	249	0.559	0.534	0.546
CATAPHOR	43	0.684	0.302	0.419
EXPLETIVE	73	1.000	1.000	1.000
GIVEN-EPITHET	230	0.647	0.870	0.742
GIVEN-PRONOUN	229	0.941	0.974	0.957
GIVEN-REFLEXIVE	97	0.990	0.979	0.984
GIVEN-REPEATED	71	0.462	0.254	0.327
GIVEN-SHORT	141	0.658	0.518	0.579
INDEF-GENERIC	102	0.385	0.196	0.260
INDEF-NEW	654	0.640	0.893	0.746
INDEF-PARTITIVE	91	0.000	0.000	0.000
INDEF-PARTITIVE-CONTAINED	72	0.443	0.375	0.406
INDEF-RESUMPTIVE	22	0.000	0.000	0.000
NULL	1	0.000	0.000	0.000
RELATIVE	7	1.000	1.000	1.000
SITUATIVE	164	0.643	0.671	0.657
UNUSED-KNOWN	627	0.739	0.710	0.724
UNUSED-TYPE	117	0.387	0.103	0.162
UNUSED-UNKNOWN	453	0.494	0.468	0.481

Detailed CRF Prediction Results

Label	Total	Precision	Recall	F-Score
BRIDGING	262	0.530	0.607	0.566
BRIDGING-CONTAINED	249	0.559	0.534	0.546
CATAPHOR	43	0.684	0.302	0.419
EXPLETIVE	73	1.000	1.000	1.000
GIVEN-EPITHET	230	0.647	0.870	0.742
GIVEN-PRONOUN	229	0.941	0.974	0.957
GIVEN-REFLEXIVE	97	0.990	0.979	0.984
GIVEN-REPEATED	71	0.462	0.254	0.327
GIVEN-SHORT	141	0.658	0.518	0.579
INDEF-GENERIC	102	0.385	0.196	0.260
INDEF-NEW	654	0.640	0.893	0.746
INDEF-PARTITIVE	91	0.000	0.000	0.000
INDEF-PARTITIVE-CONTAINED	72	0.443	0.375	0.406
INDEF-RESUMPTIVE	22	0.000	0.000	0.000
NULL	1	0.000	0.000	0.000
RELATIVE	7	1.000	1.000	1.000
SITUATIVE	164	0.643	0.671	0.657
UNUSED-KNOWN	627	0.739	0.710	0.724
UNUSED-TYPE	117	0.387	0.103	0.162
UNUSED-UNKNOWN	453	0.494	0.468	0.481

Detailed CRF Prediction Results

Label	Total	Precision	Recall	F-Score
BRIDGING	262	0.530	0.607	0.566
BRIDGING-CONTAINED	249	0.559	0.534	0.546
CATAPHOR	43	0.684	0.302	0.419
EXPLETIVE	73	1.000	1.000	1.000
GIVEN-EPITHET	230	0.647	0.870	0.742
GIVEN-PRONOUN	229	0.941	0.974	0.957
GIVEN-REFLEXIVE	97	0.990	0.979	0.984
GIVEN-REPEATED	71	0.462	0.254	0.327
GIVEN-SHORT	141	0.658	0.518	0.579
INDEF-GENERIC	102	0.385	0.196	0.260
INDEF-NEW	654	0.640	0.893	0.746
INDEF-PARTITIVE	91	0.000	0.000	0.000
INDEF-PARTITIVE-CONTAINED	72	0.443	0.375	0.406
INDEF-RESUMPTIVE	22	0.000	0.000	0.000
NULL	1	0.000	0.000	0.000
RELATIVE	7	1.000	1.000	1.000
SITUATIVE	164	0.643	0.671	0.657
UNUSED-KNOWN	627	0.739	0.710	0.724
UNUSED-TYPE	117	0.387	0.103	0.162
UNUSED-UNKNOWN	453	0.494	0.468	0.481

Detailed CRF Prediction Results

Label	Total	Precision	Recall	F-Score
BRIDGING	262	0.530	0.607	0.566
BRIDGING-CONTAINED	249	0.559	0.534	0.546
CATAPHOR	43	0.684	0.302	0.419
EXPLETIVE	73	1.000	1.000	1.000
GIVEN-EPITHET	230	0.647	0.870	0.742
GIVEN-PRONOUN	229	0.941	0.974	0.957
GIVEN-REFLEXIVE	97	0.990	0.979	0.984
GIVEN-REPEATED	71	0.462	0.254	0.327
GIVEN-SHORT	141	0.658	0.518	0.579
INDEF-GENERIC	102	0.385	0.196	0.260
INDEF-NEW	654	0.640	0.893	0.746
INDEF-PARTITIVE	91	0.000	0.000	0.000
INDEF-PARTITIVE-CONTAINED	72	0.443	0.375	0.406
INDEF-RESUMPTIVE	22	0.000	0.000	0.000
NULL	1	0.000	0.000	0.000
RELATIVE	7	1.000	1.000	1.000
SITUATIVE	164	0.643	0.671	0.657
UNUSED-KNOWN	627	0.739	0.710	0.724
UNUSED-TYPE	117	0.387	0.103	0.162
UNUSED-UNKNOWN	453	0.494	0.468	0.481

Confusion Matrix (Human Annotators)

Riester, Lorenz, Seemann (2010)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	122	25	7	3	20	2						1								
B	7	125	2								3									
C	1	3	32	1		5														1
D	35	5	8	21	5	8														
E	22	5	1		51	1														
F	3	4	2	4	3	5														
G							65										1			
H								14												
I	1	6			3				28											1
J		1	2						2	23										
K			6								38	34	2							
L	5	1	4		7						20	98	1	7						1
M			3								4	12	9							
N					1							1		6						
O											1	3			1					
P												1				12	3			
Q																	11			
R																		4		
S												1							5	
T											1									45

Confusion Matrix (Automatic System)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	159	8								1	18	2					6	26	5	37
B	4	133									12		10					7		83
C			13		17	10		1	2											
D				73																
E			2		200	2		10	16											
F			3		2	223	1													
G						2	95													
H			1		32			18	20											
I					58			10	73											
J	1									20	76	4							1	
K	1	6								15	584	8	19				5	9	3	4
L	3									2	78		1				1	3	3	
M		5									35		27					1		4
N										3	19									
O																	1			
P																7				
Q	12									1	22						110	12	2	5
R	53	5									22	1	1				28	445	1	71
S	30	1								8	28	3	2				11	9	12	13
T	37	80								2	18		1				9	90	4	212

Confusion Matrix (Automatic System)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	159	8								1	18	2					6	26	5	37
B	4	133									12		10					7		83
C			13		17	10		1	2											
D				73																
E			2		200	2		10	16											
F			3		2	223	1													
G						2	95													
H			1		32			18	20											
I					58			10	73											
J	1									20	76	4							1	
K	1	6								15	584	8	19				5	9	3	4
L	3									2	78		1				1	3	3	
M		5									35		27					1		4
N										3	19									
O																	1			
P																7				
Q	12									1	22						110	12	2	5
R	53	5									22	1	1				28	445	1	71
S	30	1								8	28	3	2				11	9	12	13
T	37	80								2	18		1				9	90	4	212

Confusion Matrix

	BRIDGING	K	R
A	159	18	26
BRIDGING-CONTAINED	4	12	7
C			
D			
E			
F			
G			
H			
I			
INDEF-GENERIC	1	76	
INDEF-NEW	1	584	9
INDEF-PARTITIVE	3	78	3
M		35	1
N		19	
O			
P			
SITUATIVE	12	22	12
UNUSED-KNOWN	53	22	445
UNUSED-TYPE	30	28	9
UNUSED-UNKNOWN	37	18	90

Confusion Matrix

	BRIDGING	K	R
A	159	18	26
BRIDGING-CONTAINED	4	12	7
O			
P			
SITUATIVE	12	22	12
UNUSED-KNOWN	53	22	445
UNUSED-TYPE	30	28	9
UNUSED-UNKNOWN	37	18	90

Confusing BRIDGING with UNUSED-KNOWN

Human annotators have the same confusion 5/89 times

(4) Die Behörden gaben eine Tsunami-Warnung für die
The authorities gave a Tsunami-warning for the
Westküste heraus.
west coast out.
'The authorities gave a Tsunami-warning for the west coast'

Confusion Matrix

	A	INDEF-NEW	R
BRIDGING	159	18	26
BRIDGING-CONTAINED	4	12	7
C			
D			
E			
F			
G			
H			
I			
INDEF-GENERIC	1	76	
K	1	584	9
INDEF-PARTITIVE	3	78	3
INDEF-PARTITIVE-CONTAINED		35	1
INDEF-RESUMPTIVE		19	
O			
P			
SITUATIVE	12	22	12
UNUSED-KNOWN	53	22	445
UNUSED-TYPE	30	28	9
UNUSED-UNKNOWN	37	18	90

Confusion Matrix

	A	INDEF-NEW	R
BRIDGING	159	18	26
BRIDGING-CONTAINED	4	12	7
O			
P			
SITUATIVE	12	22	12
UNUSED-KNOWN	53	22	445
UNUSED-TYPE	30	28	9
UNUSED-UNKNOWN	37	18	90

Confusing INDEF-NEW with INDEF-GENERIC

Human annotators have the same confusion 20/144 times

(5) Nach Angaben japanischer Medien kam ein Mensch
According to reports Japanese media came a person

ums Leben, viele Einwohner wurden verletzt.
for life, many inhabitants were injured.

'According to Japanese media reports, one person died, many inhabitants were injured'

Confusion Matrix

	A	K	UNUSED-KNOWN
BRIDGING	159	18	26
BRIDGING-CONTAINED	4	12	7
C			
D			
E			
F			
G			
H			
I			
J	1	76	
INDEF-NEW	1	584	9
INDEF-PARTITIVE	3	78	3
INDEF-PARTITIVE-CONTAINED		35	1
N		19	
O			
P			
SITUATIVE	12	22	12
UNUSED-KNOWN	53	22	445
UNUSED-TYPE	30	28	9
UNUSED-UNKNOWN	37	18	90

Confusion Matrix

	A	K	UNUSED-KNOWN
BRIDGING	159	18	26
BRIDGING-CONTAINED	4	12	7
O			
P			
SITUATIVE	12	22	12
UNUSED-KNOWN	53	22	445
UNUSED-TYPE	30	28	9
UNUSED-UNKNOWN	37	18	90

Confusing UNUSED-KNOWN with UNUSED-UNKNOWN

Human annotators have the same confusion 7 / 134 times

(6) Der Kölner Erzbischof Meisner kritisiert die
The Cologne Archbishop Meisner criticised the
Familienpolitik der Bundesregierung.
family politics of the federal government.
'The Archbishop of Cologne, Meisner, criticised the family policies of the federal government'

Addressing our underlying assumptions

1. Gold-standard co-reference information (D-GIVEN)
2. Gold-standard markables

Addressing our underlying assumptions

1. Gold-standard co-reference information (D-GIVEN)
2. Gold-standard markables

Real-world applications will not have access to this information

Test two automatic co-reference systems on the data

Accuracy (%)	Full	Collapsed
Gold	64.87	81.65
None	57.02	71.32
Simple	57.29	71.85
Unsupervised	58.15	72.90

Summary of Automatic IS Label Prediction

Trained a CRF on manually annotated text

Results are high for collapsed label set (81.65%) and well above baseline for full label set (64.87%)

Often the mistakes made by the automatic system are similar to the disagreements that human annotators have

Summary of Automatic IS Label Prediction

Trained a CRF on manually annotated text

Results are high for collapsed label set (81.65%) and well above baseline for full label set (64.87%)

Often the mistakes made by the automatic system are similar to the disagreements that human annotators have

Q: How useful is it in practice?

Outline

An application for IS Label Prediction

Revisit our earlier realisation ranking experiments

No need to use approximations of IS Labels any more

Train CRF on 1169 sentences of manually annotated corpus (test/train)

Automatically assign an IS label to every DP/NP in our TIGER training data (21,341 phrases)

Extract IS Label order patterns directly

Even Newer Generation Ranking Model

We include the IS Label asymmetric patterns directly into the svm ranking model now

Feature Types

1. C-structure	number of NPs, number of children of PP
2. C- & F-Structure	SUBJ precedes OBJ
3. Language Model	tri-gram score
~~4. IS asymmetric syntactic patterns~~	~~IS_PERS_PRON precedes IS_TEMP_ADV~~
4. IS label asymmetric patterns	D-GIVEN-SHORT precedes INDEF-NEW

Evaluation

Evaluate on 260 sentences

	BLEU	Exact Match (%)
Baseline	0.7691	50.00
IS Approx	0.7797	51.66
IS Label (full)	0.8001	54.16
IS Label (collapsed)	0.7784	51.66

Difference between the IS Label (full) model and all other models is statistically significant

Evaluation

Evaluate on 260 sentences

	BLEU	Exact Match (%)
Baseline	0.7691	50.00
IS Approx	0.7797	51.66
IS Label (full)	0.8001	54.16
IS Label (collapsed)	0.7784	51.66

Difference between the IS Label (full) model and all other models is statistically significant

Evaluation

Evaluate on 260 sentences

	BLEU	Exact Match (%)
Baseline	0.7691	50.00
IS Approx	0.7797	51.66
IS Label (full)	0.8001	54.16
IS Label (collapsed)	0.7784	51.66

Difference between the IS Label (full) model and all other models is statistically significant

Sample Improvement

(7) Im September forderten 85000 Demonstranten den Abzug
in September demanded 85,000 demonstrators the withdrawal
der 29000 auf der Insel stationierten US-Soldaten.
of the 29,000 on the island stationed US soldiers.
'85,000 demonstrators demanded the withdrawal of the 29,000 US soldiers that were stationed on the island'

IS Approximations

85000 Demonstranten forderten den Abzug der 29000 auf der Insel stationierten US-Soldaten im September .

IS Labels

Im September forderten 85000 Demonstranten den Abzug der 29000 auf der Insel stationierten US-Soldaten .

Outline

Conclusions

We have shown that a realisation ranking system can benefit from information status

Approximating the information status markup using morpho-syntactic features works well

Using automatically assigned information status labels works better

We trained a CRF model to automatically predict an IS label for a phrase, given its parse

Prediction quality on a subset of more general labels is high (81.65%) and for the full label set is well above the informed baseline (64.87%)

Outstanding Issues and Future Directions

Investigate the integration of lexical (and other) resources to improve the classification of certain phrases

Currently we still only consider single sentences. Future work will also look at preceding context

Look into carrying out an experiment with human annotators, automatically suggesting labels for them

Continue working with colleagues to improve the automatic co-reference detection for our purposes and also apply it to the TIGER training corpuse

Investigate other parsers during feature extraction for IS label prediction model

Thank you!

This work was funded by the Collaborative Research Centre (SFB 732) at the University of Stuttgart.