Evaluating Answers to Reading Comprehension Questions in Context: Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott & Janina Kopp

Universität Tübingen
SFB 833, Projekt A4

EMNLP TextInfer-Workshop 2011
July 30, 2011

Evaluating Answers to RC Questions in Context: Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp

Introduction

Our Corpus
- Data sets used

CoMiC Approach
- Annotation
- Alignment
- Classification Features

Experiment
- Overall results
- Detailed evaluation

Information Structure
- Givenness filter
- Alternative question problem
- From Givenness to Focus
- Towards annotating focus

Conclusion

References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Overview

Introduction

Our Corpus

CoMiC Approach

Experiment

Information Structure

Conclusion

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Long-term research questions

- What **linguistic representations** can be used robustly and efficiently in automatic meaning comparison?
- What is the role of **context** and how can we utilize knowledge about it in comparing meaning automatically?
  - Context here means questions and reading texts in reading comprehension tasks.

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Aims of this talk

- Present first content assessment approach for German
- Explore impact of
  - question types and
  - ways of encoding information in the text
- Discuss the importance of explicit language-based context
  - here: information structure of answers given questions and text

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Connection to RTE and Textual Inference

- What is Content Assessment?
  - The task of determining whether a response actually answers a given question about a specific text.
- Two possible perspectives in connection with RTE:
  1. Decide whether reading text *T* supports student answer *SA*, i.e., whether *SA* is entailed by *T*.
  2. Decide whether student answer *SA* is paraphrase of target answer *TA*. ⇒ bi-directional entailment

In this talk, we focus on the second perspective.

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Example from our corpus

T: *(Reading comprehension text)*

Q: **Was sind die Kritikpunkte, die Leute über Hamburg äußern?**

'What are the objections people have about Hamburg?'

TA:	**Der**	**Gestank**	**von**	**Fisch**	**und**	**Schiffsdiesel**	**an den Kais**	**.**
	The	stink	of	fish	and	fuel	on the quays	.

SA:	**Der**	**Geruch**	**zon**	**Fish**	**und**	**Schiffsdiesel**	**beim**	**Hafen** **.**
	The	smell	of_{err}	$fish_{err}$	and	fuel	at the port	.

Evaluating Answers to RC Questions in Context: Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Data source: CREG

Corpus of **R**eading Comprehension **E**xercises in **G**erman

- Consists of
  - reading texts,
  - reading comprehension questions,
  - target answers formulated by teachers,
  - student answers to the questions.

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Data source: CREG

Corpus of **R**eading Comprehension **E**xercises in **G**erman

- Consists of
  - reading texts,
  - reading comprehension questions,
  - target answers formulated by teachers,
  - student answers to the questions.
- Is being collected in two large German programs in US
  - The Ohio State University (Prof. Kathryn Corl)
  - Kansas University (Prof. Nina Vyatkina)

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Data source: CREG

Corpus of **R**eading Comprehension **E**xercises in **G**erman

- Consists of
  - reading texts,
  - reading comprehension questions,
  - target answers formulated by teachers,
  - student answers to the questions.
- Is being collected in two large German programs in US
  - The Ohio State University (Prof. Kathryn Corl)
  - Kansas University (Prof. Nina Vyatkina)
- Two research assistants independently rate each student answer with respect to meaning.
  - Did student provide meaningful answer to question?
  - Binary categories: adequate/inadequate
  - Annotators also identify target answer for student answer

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

Data sets used

- From the corpus in development, we took a snapshot
  - with full agreement in binary ratings,
  - and with half of the answers being rated as *inadequate* (random base line = 50%).
- Resulted in one data set for each of the two sites
  - No overlap in exercise material

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

Data sets used

- From the corpus in development, we took a snapshot
  - with full agreement in binary ratings,
  - and with half of the answers being rated as *inadequate* (random base line = 50%).
- Resulted in one data set for each of the two sites
  - No overlap in exercise material

	KU data set	OSU data set
Target Answers	136	87
Questions	117	60
Student Answers	**610**	**422**
# of Students	141	175
SAs per question	5.21	7.03
avg. Token #	9.71	15.00

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

General CoMiC Approach

(Bailey & Meurers 2008; Meurers, Ziai, Ott & Bailey 2011)

The overall approach has three phases:

1. **Annotation** uses NLP to enrich the student and target answers, as well as the question text, with linguistic information on different levels and types of abstraction.
2. **Alignment** maps elements of the learner answer to elements of the target response using annotation.
   - Global alignment solution computed by Traditional Marriage Algorithm (Gale & Shapley 1962)
3. **Classification** analyzes the possible alignments and labels the learner response with a binary content assessment and a detailed diagnosis code.

Evaluating Answers to RC Questions in Context: Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Annotation

NLP Components

Annotation Task	**NLP Component**
Sentence Detection	OpenNLP http://incubator.apache.org/opennlp
Tokenization	OpenNLP
Lemmatization	TreeTagger (Schmid 1994)
Spell Checking	Edit distance (Levenshtein 1966) igerman98 word list http://www.j3e.de/ispell/igerman98
Part-of-speech Tagging	TreeTagger (Schmid 1994)
Noun Phrase Chunking	OpenNLP
Lexical Relations	GermaNet (Hamp & Feldweg 1997)
Similarity Scores	PMI-IR (Turney 2001)
Dependency Relations	MaltParser (Nivre et al. 2007)

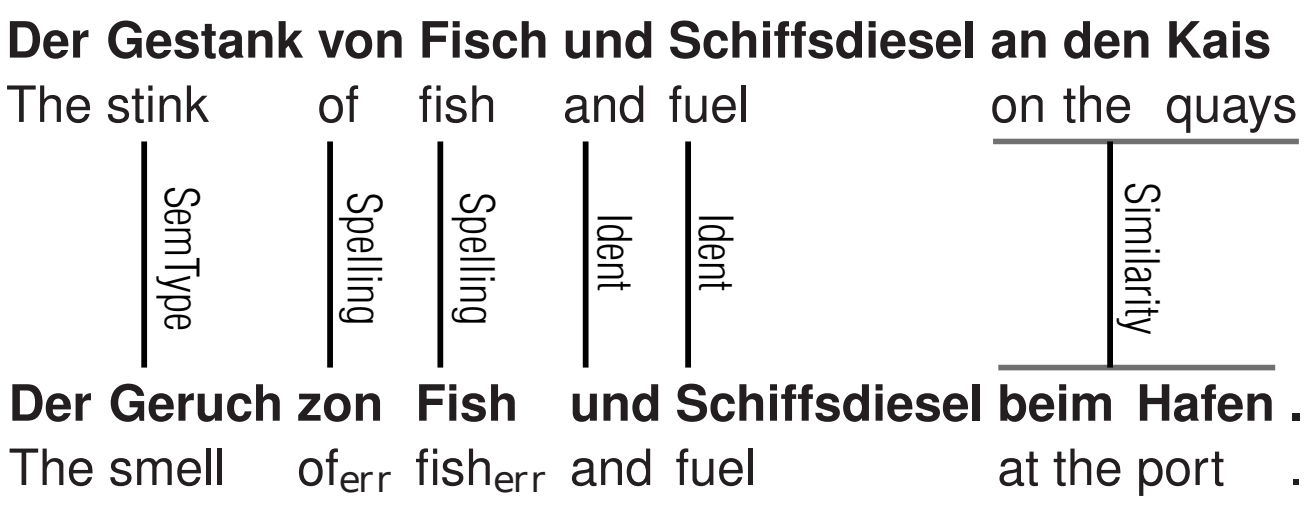
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Alignment

Example

Q: **Was sind die Kritikpunkte, die Leute über Hamburg äußern?**

'What are the objections people have about Hamburg?'



Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

Classification

Features

- Content Assessment is based on 13 features:

  % of Overlapping Matches:

  - keyword (head)
  - target/learner token
  - target/learner chunk
  - target/learner triple

  Nature of Matches:

  - % token matches
  - % lemma matches
  - % synonym matches
  - % similarity matches
  - % sem. type matches
  - match variety

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Classification

Features

- Content Assessment is based on 13 features:

  % of Overlapping Matches:

  - keyword (head)
  - target/learner token
  - target/learner chunk
  - target/learner triple

  Nature of Matches:

  - % token matches
  - % lemma matches
  - % synonym matches
  - % similarity matches
  - % sem. type matches
  - match variety

- We combined the evidence with memory-based learning (TiMBL, Daelemans et al. 2007)
  - Trained seven classifiers using different distance metrics, overall outcome obtained through majority voting.
  - Used *leave-one-out* testing: For each test item train on all answer pairs except the test item itself.

Experiment

Overall results

	KU data set	OSU data set
# of answers	610	422
Accuracy	**84.6%**	**84.6%**

- Remarkable similarity of results across completely different data sets
- Same overall results when macro-averaging over individual questions
- Competitive with results obtained for English (78%) in Bailey & Meurers (2008) and related results of C-Rater for short answer scoring (Leacock & Chodorow 2003).

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Detailed Evaluation

- Global accuracy scores do not tell us how well the system fares, e.g., in terms of question types.
- First step towards deeper analysis of results: manual annotation of reading comprehension question properties
- Annotation scheme follows Day & Park (2005) guidelines for development of reading comprehension questions
  - Comprehension Types:
    - nature & depth of comprehension required by learner to answer the question
    - in our data: “Literal”, “Reorganization” and “Inference”
  - Question Forms:
    - Surface-based question classes such as “yes/no” or “who” questions

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Detailed Evaluation

Accuracy by question form and comprehension

	Literal	Reorganization	Inference	Total
Alternative	0 (1)	– (0)	66.67 (6)	57.14 (7)
How	85.71 (126)	83.33 (12)	100 (7)	86.21 (145)
What	87.04 (247)	74.19 (31)	83.33 (6)	85.56 (284)
When	85.71 (7)	– (0)	– (0)	85.71 (7)
Where	88.89 (9)	– (0)	– (0)	88.89 (9)
Which	92.35 (183)	100 (14)	83.33 (6)	92.61 (203)
Who	73.91 (23)	94.44 (18)	– (0)	82.93 (41)
Why	80.47 (128)	57.14 (14)	84.38 (32)	79.31 (174)
Yes/No	– (0)	100 (5)	– (0)	100 (5)
Several	82.11 (95)	68.42 (38)	75 (24)	77.71 (157)
Total	85.96 (819)	78.03 (132)	81.48 (81)	84.59 (1032)

- Answer counts shown in brackets
- Error bars indicate 95% confidence intervals

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Most important results

Comprehension types

	Literal	Reorganization	Inference	Total
Alternative	0 (1)	– (0)	66.67 (6)	57.14 (7)
How	85.71 (126)	83.33 (12)	100 (7)	86.21 (145)
What	87.04 (247)	74.19 (31)	83.33 (6)	85.56 (284)
When	85.71 (7)	– (0)	– (0)	85.71 (7)
Where	88.89 (9)	– (0)	– (0)	88.89 (9)
Which	92.35 (183)	100 (14)	83.33 (6)	92.61 (203)
Who	73.91 (23)	94.44 (18)	– (0)	82.93 (41)
Why	80.47 (128)	57.14 (14)	84.38 (32)	79.31 (174)
Yes/No	– (0)	100 (5)	– (0)	100 (5)
Several	82.11 (95)	68.42 (38)	75 (24)	77.71 (157)
Total	85.96 (819)	78.03 (132)	81.48 (81)	84.59 (1032)

- “Literal” questions (86.0%) seem to be easier than “Reorganization” (78.0%) and “Inference” (81.5%).

Evaluating Answers to RC Questions in Context: Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Most important results

Question forms: easy case

	Literal	Reorganization	Inference	Total
Alternative	0 (1)	– (0)	66.67 (6)	57.14 (7)
How	85.71 (126)	83.33 (12)	100 (7)	86.21 (145)
What	87.04 (247)	74.19 (31)	83.33 (6)	85.56 (284)
When	85.71 (7)	– (0)	– (0)	85.71 (7)
Where	88.89 (9)	– (0)	– (0)	88.89 (9)
Which	92.35 (183)	100 (14)	83.33 (6)	92.61 (203)
Who	73.91 (23)	94.44 (18)	– (0)	82.93 (41)
Why	80.47 (128)	57.14 (14)	84.38 (32)	79.31 (174)
Yes/No	– (0)	100 (5)	– (0)	100 (5)
Several	82.11 (95)	68.42 (38)	75 (24)	77.71 (157)
Total	85.96 (819)	78.03 (132)	81.48 (81)	84.59 (1032)

- Accuracy for *wh*-questions based on concrete information from text is rather high, e.g., 92.6% for “which” questions.

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Most important results

Question forms: hard case

	Literal	Reorganization	Inference	Total
Alternative	0 (1)	– (0)	66.67 (6)	57.14 (7)
How	85.71 (126)	83.33 (12)	100 (7)	86.21 (145)
What	87.04 (247)	74.19 (31)	83.33 (6)	85.56 (284)
When	85.71 (7)	– (0)	– (0)	85.71 (7)
Where	88.89 (9)	– (0)	– (0)	88.89 (9)
Which	92.35 (183)	100 (14)	83.33 (6)	92.61 (203)
Who	73.91 (23)	94.44 (18)	– (0)	82.93 (41)
Why	80.47 (128)	57.14 (14)	84.38 (32)	79.31 (174)
Yes/No	– (0)	100 (5)	– (0)	100 (5)
Several	82.11 (95)	68.42 (38)	75 (24)	77.71 (157)
Total	85.96 (819)	78.03 (132)	81.48 (81)	84.59 (1032)

- "why" questions are difficult (79.3%): Asking for reasons/causes supports more answer variation.

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Most important results

Question forms: a puzzle

	Literal	Reorganization	Inference	Total
Alternative	0 (1)	– (0)	66.67 (6)	57.14 (7)
How	85.71 (126)	83.33 (12)	100 (7)	86.21 (145)
What	87.04 (247)	74.19 (31)	83.33 (6)	85.56 (284)
When	85.71 (7)	– (0)	– (0)	85.71 (7)
Where	88.89 (9)	– (0)	– (0)	88.89 (9)
Which	92.35 (183)	100 (14)	83.33 (6)	92.61 (203)
Who	73.91 (23)	94.44 (18)	– (0)	82.93 (41)
Why	80.47 (128)	57.14 (14)	84.38 (32)	79.31 (174)
Yes/No	– (0)	100 (5)	– (0)	100 (5)
Several	82.11 (95)	68.42 (38)	75 (24)	77.71 (157)
Total	85.96 (819)	78.03 (132)	81.48 (81)	84.59 (1032)

- “Alternative” questions are near random level (57.1%).
  - Why?

Evaluating Answers to RC Questions in Context: Results for German and the Role of Information Structure

Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Information Structure

- Information Structure (IS) research investigates:
  - How is the meaning of a sentence integrated into the discourse?
- One relevant notion is **Givenness**:
  - "A constituent C counts as Given if there is a salient antecedent A for C, such that A either
    - co-refers with C,
    - is a synonym of C or
    - is a hyponym of C." (Büring 2006)
- Our system as a first approximation excludes all words from alignment that appear in the question.
  - Motivation: Mentioned lexical material typically does not contain new information answering the question.
- However, in some interesting cases, the answer to a question does include given information.
  - Example: "Alternative" questions

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Alternative question example

Q: Ist die Wohnung in einem Neubau oder einem Altbau?

'Is the flat in a new building or in an old building?'

TA: Die Wohnung ist in einem Neubau .
The flat is in a new building

SA: Die Wohnung ist in einem Neubau
The flat is in a new building

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

SFB 833
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

From Givenness to Focus

- The IS notion of a **Focus** as the expression which addresses an explicit or implicit question under discussion (Krifka 2004) helps address the issue.
  - → Given information is relevant when it is part of the focus.

- Making the focus explicit can also help in cases such as:

Q: Was muss die Meerjungfrau erleiden, wenn sie Menschenbeine haben will?

'What must the mermaid suffer if she wants to have human legs?'

TA: Die Meerjungfrau muss schreckliche Qualen erleiden bei jedem Schritt .
The mermaid must horrible torment suffer with every step .

SA: Sie erleidt bei jedem Schritt.
She suffer with every step.

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Towards annotating focus

- Idea: Integrate an automatic focus identification component into CoMiC.
- Approach should be informed by manual approaches to annotating information structure aspects:
  - Those targeting focus are moderately successful (Dipper et al. 2007; Calhoun et al. 2010).
  - In the CREG corpus, the explicit linguistic context (text, question) may support more reliable focus identification.
  - Information Status (Given vs. New) of referential expressions (Riester et al. 2010) may help as “backbone”.

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

Conclusion

- We presented the first content assessment system for German, CoMiC-DE
  - accuracy of 84.6% on authentic classroom data
  - competitive with results for English
- Detailed evaluation by question form and comprehension type
  - clear differences in performance
  - identifies avenues for future research improving analysis for specific question forms and comprehension types
- To identify which parts of an answer are most relevant for content assessment, information structure distinctions should be integrated.
  - manual annotation of the focus of an answer is a first step
  - explicit language-based context of task is crucial

Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

The End

Thank you!

Evaluating Answers to RC Questions in Context:
Results for German and the Role of Information Structure
Detmar Meurers, Ramon Ziai, Niels Ott, Janina Kopp
Introduction
Our Corpus
Data sets used
CoMiC Approach
Annotation
Alignment
Classification Features
Experiment
Overall results
Detailed evaluation
Information Structure
Givenness filter
Alternative question problem
From Givenness to Focus
Towards annotating focus
Conclusion
References

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

References I

Bailey, S. & D. Meurers (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein & R. D. Felice (eds.), *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications (BEA-3) at ACL'08*. Columbus, Ohio, pp. 107–115.

Büring, D. (2006). Intonation und Informationsstruktur. In H. Blühdorn, E. Breindl & U. H. Waßner (eds.), *Text — Verstehen. Grammatik und darüber hinaus*. Berlin/New York: de Gruyter, pp. 144–163.

Calhoun, S., J. Carletta, J. Brenier, N. Mayo, D. Jurafsky, M. Steedman & D. Beaver (2010). The NXT-format Switchboard Corpus: A Rich Resource for Investigating the Syntax, Semantics, Pragmatics and Prosody of Dialogue. *Language Resources and Evaluation* 44, 387–419.

Daelemans, W., J. Zavrel, K. van der Sloot & A. van den Bosch (2007). *TiMBL: Tilburg Memory-Based Learner Reference Guide, ILK Technical Report ILK 07-03*. Induction of Linguistic Knowledge Research Group Department of Communication and Information Sciences, Tilburg University, P.O. Box 90153, NL-5000 LE, Tilburg, The Netherlands, version 6.0 ed.

Day, R. R. & J.-S. Park (2005). Developing Reading Comprehension Questions. *Reading in a Foreign Language* 17(1), 60–73.

Dipper, S., M. Götze & S. Skopeteas (eds.) (2007). *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure*, vol. 7 of *Interdisciplinary Studies on Information Structure*. Potsdam, Germany: Universitätsverlag Potsdam.

Gale, D. & L. S. Shapley (1962). College Admissions and the Stability of Marriage. *American Mathematical Monthly* 69, 9–15.

Hamp, B. & H. Feldweg (1997). GermaNet – a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid.

Krifka, M. (2004). The semantics of questions and the focusation of answers. In C. Lee, M. Gordon & D. Büring (eds.), *Topic and Focus: A Cross-Linguistic Perspective*, Dordrecht: Kluwer Academic Publishers, pp. 139–151.

Leacock, C. & M. Chodorow (2003). C-rater: Automated Scoring of Short-Answer Questions. *Computers and the Humanities* 37, 389–405.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady* 10(8), 707–710.

References II

Meurers, D., R. Ziai, N. Ott & S. Bailey (2011). Integrating Parallel Analysis Modules to Evaluate the Meaning of Answers to Reading Comprehension Questions. *IJCEELL. Special Issue on Automatic Free-text Evaluation* 21(4), 355–369.

Nivre, J., J. Nilsson, J. Hall, A. Chanev, G. Eryigit, S. Kübler, S. Marinov & E. Marsi (2007). MaltParser: A Language-Independent System for Data-Driven Dependency Parsing. *Natural Language Engineering* 13(1), 1–41.

Riester, A., D. Lorenz & N. Seemann (2010). A Recursive Annotation Scheme for Referential Information Status. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*. Valletta, Malta.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK, pp. 44–49.

Turney, P. (2001). Mining the Web for Synonyms: PMI-IR Versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001)*. Freiburg, Germany, pp. 491–502.

Introduction

Our Corpus

Data sets used

CoMiC Approach

Annotation

Alignment

Classification Features

Experiment

Overall results

Detailed evaluation

Information Structure

Givenness filter

Alternative question problem

From Givenness to Focus

Towards annotating focus

Conclusion

References