Evaluating the Meaning of Answers to Reading Comprehension Questions: A Semantics-Based Approach

Michael Hahn Detmar Meurers

Department of Linguistics, SFB 833
University of Tübingen

BEA 7 Workshop, NAACL-HLT
Montreal, 7. June 2012

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Introduction

- A range of approaches have been proposed for short answer meaning assessment (Ziai, Ott & Meurers 2012).

Introduction

- A range of approaches have been proposed for short answer meaning assessment (Ziai, Ott & Meurers 2012).
- Meaning comparison generally relies on a combination of surface-based and deeper linguistic representations,
  - but essentially no use is made of semantic formalisms created by theoretical linguists to represent meaning.

Introduction

- A range of approaches have been proposed for short answer meaning assessment (Ziai, Ott & Meurers 2012).
- Meaning comparison generally relies on a combination of surface-based and deeper linguistic representations,
  - but essentially no use is made of semantic formalisms created by theoretical linguists to represent meaning.
    - deep linguistic analysis of formal semantics often lacks coverage and robustness
    - semantic structures are complex to derive and compare

Introduction

- A range of approaches have been proposed for short answer meaning assessment (Ziai, Ott & Meurers 2012).
- Meaning comparison generally relies on a combination of surface-based and deeper linguistic representations,
  - but essentially no use is made of semantic formalisms created by theoretical linguists to represent meaning.
    - deep linguistic analysis of formal semantics often lacks coverage and robustness
    - semantic structures are complex to derive and compare
    + semantic representations abstract away from lexical and syntactic variation in the realization of the same meaning
    + they precisely expose meaning distinctions and support linking meaning to discourse

Introduction

- A range of approaches have been proposed for short answer meaning assessment (Ziai, Ott & Meurers 2012).
- Meaning comparison generally relies on a combination of surface-based and deeper linguistic representations,
  - but essentially no use is made of semantic formalisms created by theoretical linguists to represent meaning.
    - deep linguistic analysis of formal semantics often lacks coverage and robustness
    - semantic structures are complex to derive and compare
    + semantic representations abstract away from lexical and syntactic variation in the realization of the same meaning
    + they precisely expose meaning distinctions and support linking meaning to discourse
- We present a short answer assessment approach based on underspecified formal semantic representations.

Outline

Empirical challenge: CREG

- Empirical basis: Corpus of Reading Comprehension Exercises in German (CREG; Ott, Ziai & Meurers 2012)
  - CREG consists of texts, questions, target answers, and student answers written by learners of German.

Empirical challenge: CREG

- Empirical basis: Corpus of Reading Comprehension Exercises in German (CREG; Ott, Ziai & Meurers 2012)
  - CREG consists of texts, questions, target answers, and student answers written by learners of German.
- CREG data was collected and assessed in two large German programs in the US: KU and OSU
  - For each student answer, two independent annotators evaluated whether it correctly answers the question.
    - Answers were only assessed with respect to meaning, not orthography or grammaticality.

Empirical challenge: CREG

- Empirical basis: Corpus of Reading Comprehension Exercises in German (CREG; Ott, Ziai & Meurers 2012)
  - CREG consists of texts, questions, target answers, and student answers written by learners of German.
- CREG data was collected and assessed in two large German programs in the US: KU and OSU
  - For each student answer, two independent annotators evaluated whether it correctly answers the question.
    - Answers were only assessed with respect to meaning, not orthography or grammaticality.
- Data freely available, and reference results available for CoMiC-DE system (Meurers, Ziai, Ott & Kopp 2011),
  - a system not using formal semantic representations

Representations: Lexical Resource Semantics

- LRS (Richter & Sailer 2003) is an underspecified semantic formalism:
  - standard model-theoretic semantics
  - semantic representations are not completely specified but subsume a set of possible resolved expressions

Representations: Lexical Resource Semantics

- LRS (Richter & Sailer 2003) is an underspecified semantic formalism:
  - standard model-theoretic semantics
  - semantic representations are not completely specified but subsume a set of possible resolved expressions
- Advantage of an underspecified semantic formalism for content assessment:
  - provides access to fine-grained semantic distinctions
    - all parts of the semantic representation are accessible in a flat representation
  - how the parts are combined is separately encoded (variable bindings, dominance)
    - avoids costly computation of all readings
    - similar parts can be compared independent of where they appear in the overall semantics

Example for LRS representation

(1) *Alle Zimmer haben nicht eine Dusche.*
all rooms have not a shower

'Not every room has a shower.' *vs.* 'No room has a shower.'

Example for LRS representation

(1) *Alle Zimmer haben nicht eine Dusche.*
all rooms have not a shower

'Not every room has a shower.' *vs.* 'No room has a shower.'

- INTERNAL CONTENT: core semantic contribution of head
- EXTERNAL CONTENT: semantic representation of sentence
- PARTS: all subterms of the representation

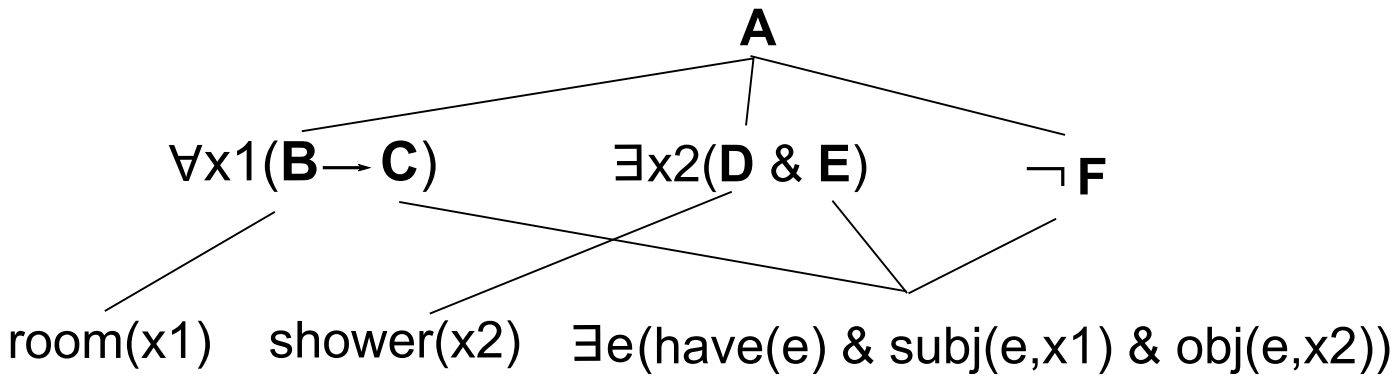
$$\begin{bmatrix} \textsc{incont} & \textit{have(e)} \\ \textsc{excont} & \mathrm{A} \\ \textsc{parts} & \left\langle \begin{array}{l} \mathrm{A},\ \mathrm{have(e)},\ \forall x_1(\mathrm{B} \rightarrow \mathrm{C}),\ \exists x_2\ (\mathrm{D} \wedge \mathrm{E}),\ \neg\ \mathrm{F}, \\ \mathrm{room}(x_1),\ \mathrm{shower}(x_2),\ \mathrm{subj(e,}x_1\mathrm{)},\ \mathrm{obj(e,}x_2\mathrm{)} \\ \exists \mathrm{e}(\mathrm{have(e)} \wedge \mathrm{subj(e,}x_1\mathrm{)} \wedge \mathrm{obj(e,}x_2\mathrm{)}) \end{array} \right\rangle \end{bmatrix}$$

Example for LRS representation (cont.)

- The readings of the sentence are obtained by identifying the meta-variables A, . . . , F with the subformulas.

Example for LRS representation (cont.)

- The readings of the sentence are obtained by identifying the meta-variables A, . . . , F with the subformulas.
- LRS representations include dominance constraints, which restrict possible identifications, e.g.:



Our general approach

1. automatically derive LRS representations for the student answer, the target answer, and the question
   - method described in Hahn & Meurers (2011)
   - based on statistical dependency parsing
   - always results in an LRS structure, also for ill-formed input

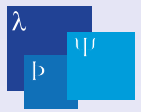
Our general approach

1. automatically derive LRS representations for the student answer, the target answer, and the question
   - method described in Hahn & Meurers (2011)
   - based on statistical dependency parsing
   - always results in an LRS structure, also for ill-formed input
2. align LRS representations of target and student answers
   - *local* measures of semantic similarity
   - *global* measures of extent to which alignment preserves semantic structure (variable bindings, dominance)

Our general approach

1. automatically derive LRS representations for the student answer, the target answer, and the question
   - method described in Hahn & Meurers (2011)
   - based on statistical dependency parsing
   - always results in an LRS structure, also for ill-formed input
2. align LRS representations of target and student answers
   - *local* measures of semantic similarity
   - *global* measures of extent to which alignment preserves semantic structure (variable bindings, dominance)

   alignments also computed between answers and question



Our general approach

1. automatically derive LRS representations for the student answer, the target answer, and the question
   - method described in Hahn & Meurers (2011)
   - based on statistical dependency parsing
   - always results in an LRS structure, also for ill-formed input
2. align LRS representations of target and student answers
   - *local* measures of semantic similarity
   - *global* measures of extent to which alignment preserves semantic structure (variable bindings, dominance)

   alignments also computed between answers and question
3. perform overall meaning comparison based on numerical scores representing quality of alignment

Aligning Meaning Representations

- An alignment between two LRS representations is a bijective partial mapping between PARTS lists p_1^n and q_1^m
  - Every element of one representation can be aligned to at most one element of the other representation.

Aligning Meaning Representations

- An alignment between two LRS representations is a bijective partial mapping between PARTS lists p_1^n and q_1^m
  - Every element of one representation can be aligned to at most one element of the other representation.
- A simple example: “John left.” vs. “Jon vanished.”

∃x A B	——	∃y E F
john(x)	——	jon(y)
∃e C D	——	∃f G H
subj(e,x)	——	subj(f,y)
leave(e)	——	vanish(f)

Aligning Meaning Representations

- An alignment between two LRS representations is a bijective partial mapping between PARTS lists p_1^n and q_1^m
  - Every element of one representation can be aligned to at most one element of the other representation.
- A simple example: “John left.” vs. “Jon vanished.”

∃x A B	∃y E F
john(x)	jon(y)
∃e C D	∃f G H
subj(e,x)	subj(f,y)
leave(e)	vanish(f)

- Best alignment is determined automatically using a maximization criterion.

Automatically Deriving Alignments

Maximization criterion

- combines three measures of alignment quality:
  - **LinkScore**: similarity of the alignment links
  - **VariableScore**: consistency of alignments with respect to the induced variable bindings θ
  - **DominanceScore**: consistency with respect to dominance constraints
- $Q(a, \theta | S, T) = \quad LinkScore(a|S,T) \cdot VariableScore(\theta) \cdot DominanceScore(a|S,T)$
- The alignment maximizing the criterion is found efficiently using the A* algorithm.

LinkScore: Similarity for Alignment Links

- Base cases:
  - Variables can be matched with any variable of same type.
  - For other semantic terms, compute the maximum score of
    - Levenshtein distance, to account for spelling errors
    - Synonyms: score 1 if in GermaNet (Hamp & Feldweg 1997)
    - Dissimilar elements of same category: constant costs, empirically determined for pairs of
      – grammatical function terms
      – special terms (affirmative or negative natural language expressions and logical negation)
    - ...

LinkScore: Similarity for Alignment Links

- Base cases:
  - Variables can be matched with any variable of same type.
  - For other semantic terms, compute the maximum score of
    - Levenshtein distance, to account for spelling errors
    - Synonyms: score 1 if in GermaNet (Hamp & Feldweg 1997)
    - Dissimilar elements of same category: constant costs, empirically determined for pairs of
      – grammatical function terms
      – special terms (affirmative or negative natural language expressions and logical negation)
    - ...
- Complex expressions are compared recursively.

LinkScore: Similarity for Alignment Links

- Base cases:
  - Variables can be matched with any variable of same type.
  - For other semantic terms, compute the maximum score of
    - Levenshtein distance, to account for spelling errors
    - Synonyms: score 1 if in GermaNet (Hamp & Feldweg 1997)
    - Dissimilar elements of same category: constant costs, empirically determined for pairs of
      – grammatical function terms
      – special terms (affirmative or negative natural language expressions and logical negation)
    - ...
- Complex expressions are compared recursively.

$\Rightarrow$ Overall **LinkScore** = sum of similarity of all alignment links

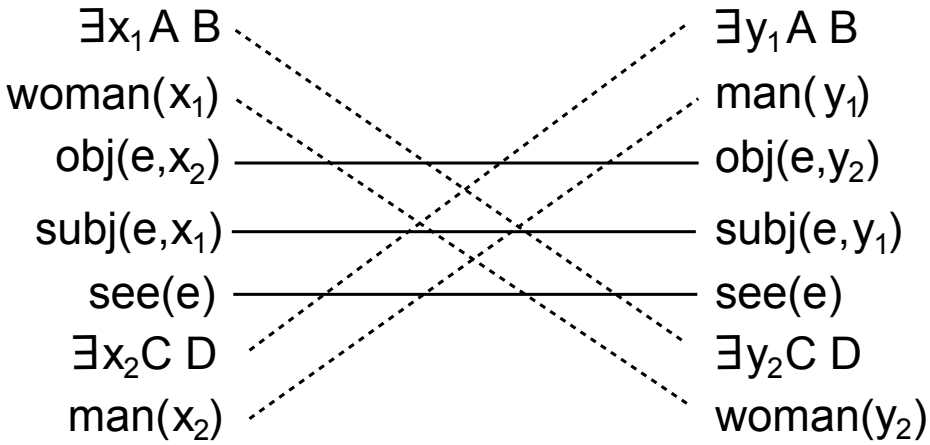
- unaligned elements: constant cost μ_{NULL} (may be smaller than costly alignment link in another overall alignment)

VariableScore

- Every alignment induces a unifier, which unifies all variables which are matched by the alignment.

VariableScore

- Every alignment induces a unifier, which unifies all variables which are matched by the alignment.
- “A woman sees a man.” vs. “A man sees a woman.”



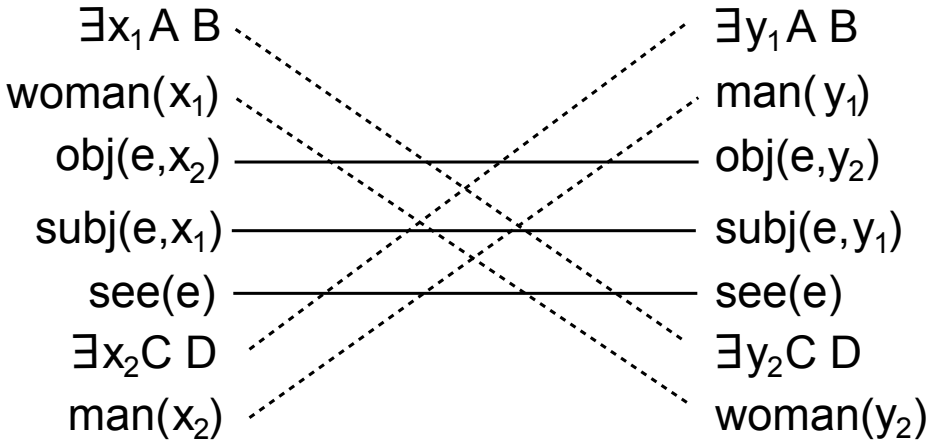
  - line links require unifying x_1 with y_1 and x_2 with y_2
  - adding the dotted links would result in unifying all variables

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

VariableScore

- Every alignment induces a unifier, which unifies all variables which are matched by the alignment.
- “A woman sees a man.” vs. “A man sees a woman.”



  - line links require unifying x_1 with y_1 and x_2 with y_2
  - adding the dotted links would result in unifying all variables

- An alignment which preserves the structure will not unify two distinct variables from the same LRS representation.

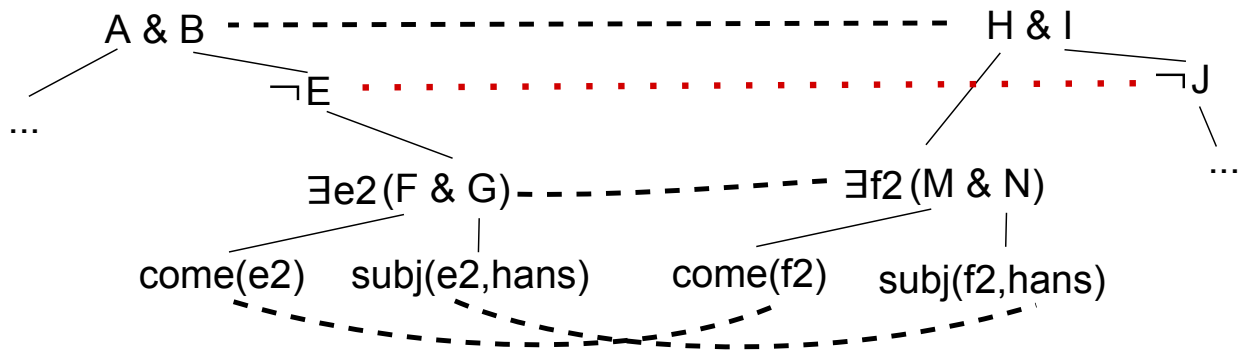
$\Rightarrow$ **VariableScore** = information loss resulting from unification

DominanceScore

- Mismatches in the structure of the linked semantic representations need to be taken into account.
- For example:

(2) a. *Peter will come but Hans will **not** come.*

b. *Peter will **not** come but Hans will come.*



⇒ **DominanceScore** = extent to which an alignment defines an isomorphism

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

From Alignment to Meaning Comparison

Basic measures

- Based on the best overall alignment identified using A*, we compute several measures for meaning comparison.
- **ALIGN** measure, based on alignment quality Q:

$$\text{ALIGN} = \frac{\text{alignment quality(student answer, target answer)}}{\text{\# of elements in shorter PARTS list}}$$

From Alignment to Meaning Comparison

Basic measures

- Based on the best overall alignment identified using A*, we compute several measures for meaning comparison.
- **ALIGN** measure, based on alignment quality Q:

$$\text{ALIGN} = \frac{\text{alignment quality(student answer, target answer)}}{\text{\# of elements in shorter PARTS list}}$$

- **EQUAL** measure, based on number of alignment links:

$$\text{Student} = \frac{\text{\# of alignment links(student answer, target answer)}}{\text{\# of elements on PARTS list of } \textit{student} \text{ answer}}$$

$$\text{Target} = \frac{\text{\# of alignment links(student answer, target answer)}}{\text{\# of elements of PARTS list of } \textit{target} \text{ answer}}$$

Average = average of Student and Target measures

From Alignment to Meaning Comparison

Studying Impact of Functional Elements

- **EQUAL** measures treat all semantic elements the same

From Alignment to Meaning Comparison

Studying Impact of Functional Elements

- **EQUAL** measures treat all semantic elements the same
- Define measures to help identify the impact of functional elements (quantifiers, lambda operator, *subj, obj*, . . .):

From Alignment to Meaning Comparison

Studying Impact of Functional Elements

- **EQUAL** measures treat all semantic elements the same
- Define measures to help identify the impact of functional elements (quantifiers, lambda operator, *subj, obj*, . . .):
- **IGNORE** measures: ignore all functional elements
- **WEIGHTED** measures: weight elements so that functional and non-functional ones differ in impact
  - weights are empirically determined using grid search on a development set

From Alignment to Meaning Comparison

Studying Impact of Information Structure

- Information Structure (Krifka 2008): structuring of the meaning of a response in relation to the discourse
  - *given* (vs. *new*): part of meaning known from question
  - *focus* (vs. *background*): part of meaning selecting between the set of alternatives that the question raises

From Alignment to Meaning Comparison

Studying Impact of Information Structure

- Information Structure (Krifka 2008): structuring of the meaning of a response in relation to the discourse
  - *given* (vs. *new*): part of meaning known from question
  - *focus* (vs. *background*): part of meaning selecting between the set of alternatives that the question raises
- Basing meaning comparison on semantic representation allows us to directly represent Information Structure.

From Alignment to Meaning Comparison

Studying Impact of Information Structure

- Information Structure (Krifka 2008): structuring of the meaning of a response in relation to the discourse
  - *given* (vs. *new*): part of meaning known from question
  - *focus* (vs. *background*): part of meaning selecting between the set of alternatives that the question raises
- Basing meaning comparison on semantic representation allows us to directly represent Information Structure.
- Some previous approaches exclude *given* material from alignment (Bailey & Meurers 2008; Mohler et al. 2011):
  - greatly improves classification accuracy

From Alignment to Meaning Comparison

Studying Impact of Information Structure

- Information Structure (Krifka 2008): structuring of the meaning of a response in relation to the discourse
  - *given* (vs. *new*): part of meaning known from question
  - *focus* (vs. *background*): part of meaning selecting between the set of alternatives that the question raises
- Basing meaning comparison on semantic representation allows us to directly represent Information Structure.
- Some previous approaches exclude *given* material from alignment (Bailey & Meurers 2008; Mohler et al. 2011):
  - greatly improves classification accuracy
- Meurers, Ziai, Ott & Kopp (2011) show that the relevant linguistic aspect here is not *givenness* but *focus*.

From Alignment to Meaning Comparison

How we integrate information structure

- Needed: A component which automatically identifies the focus of an answer in a question-answer pair.
  - First approximation: an element on the PARTS lists of an answer is marked as focused if it is not aligned to the question, except for alignment with explicit alternatives.

From Alignment to Meaning Comparison

How we integrate information structure

- Needed: A component which automatically identifies the focus of an answer in a question-answer pair.
  - First approximation: an element on the PARTS lists of an answer is marked as focused if it is not aligned to the question, except for alignment with explicit alternatives.
- **FOCUS** measures: BASIC measures counting only those elements which are recognized as focused
- **GIVEN** measures: BASIC measures counting only elements not aligned to the question

Experiments

Setup

- Corpus
  - 1032 answers from the CREG corpus, used for evaluating the CoMiC-DE system (Meurers, Ziai, Ott & Kopp 2011)
  - balanced: same number of correct and incorrect answers

Experiments

Setup

- Corpus
  - 1032 answers from the CREG corpus, used for evaluating the CoMiC-DE system (Meurers, Ziai, Ott & Kopp 2011)
  - balanced: same number of correct and incorrect answers
- Preparation
  - optimized numerical parameters using grid search on a separate development set of 379 answers from CREG

Experiments

Setup

- Corpus
  - 1032 answers from the CREG corpus, used for evaluating the CoMiC-DE system (Meurers, Ziai, Ott & Kopp 2011)
  - balanced: same number of correct and incorrect answers
- Preparation
  - optimized numerical parameters using grid search on a separate development set of 379 answers from CREG
- Experiment
  - explored all measures for meaning assessment
  - binary classification is based on a threshold
    - arithmetic mean of the average result of correct and the average result of incorrect answers
  - training and testing performed using the leave-one-out scheme Weiss & Kulikowski (1991)

Results

	BASIC	GIVEN	FOCUS
WEIGHTED Average	80.9	86.1	**86.3**
IGNORE Average	79.8	84.7	84.9
EQUAL Average	76.6	80.8	80.7
ALIGN	77.1		
CoMiC-DE		84.6	

- Best accuracy with WEIGHTED Average FOCUS measure

Results

	BASIC	GIVEN	FOCUS
WEIGHTED Average	80.9	86.1	**86.3**
IGNORE Average	79.8	84.7	84.9
EQUAL Average	76.6	80.8	80.7
ALIGN	77.1		
CoMiC-DE		84.6	

- Best accuracy with WEIGHTED Average FOCUS measure
- Including functional elements improves accuracy (+1.4%)

SFB 833

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Results

	BASIC	GIVEN	FOCUS
WEIGHTED Average	80.9	86.1	**86.3**
IGNORE Average	79.8	84.7	84.9
EQUAL Average	76.6	80.8	80.7
ALIGN	77.1		
CoMiC-DE		84.6	

- Best accuracy with WEIGHTED Average FOCUS measure
- Including functional elements improves accuracy (+1.4%)
  - weight should differ from content elements (+5.6%)

Results

	BASIC	GIVEN	FOCUS
WEIGHTED Average	80.9	86.1	**86.3**
IGNORE Average	79.8	84.7	84.9
EQUAL Average	76.6	80.8	80.7
ALIGN	77.1		
CoMiC-DE		84.6	

- Best accuracy with WEIGHTED Average FOCUS measure
- Including functional elements improves accuracy (+1.4%)
  - weight should differ from content elements (+5.6%)
- Information Structure
  - Focus helps target relevant part of answer (+5.4%)

Results

	BASIC	GIVEN	FOCUS
WEIGHTED Average	80.9	86.1	**86.3**
IGNORE Average	79.8	84.7	84.9
EQUAL Average	76.6	80.8	80.7
ALIGN	77.1		
CoMiC-DE		84.6	

- Best accuracy with WEIGHTED Average FOCUS measure
- Including functional elements improves accuracy (+1.4%)
  - weight should differ from content elements (+5.6%)
- Information Structure
  - Focus helps target relevant part of answer (+5.4%)
- Outperforms CoMiC-DE, also integrating givenness
  - supports usefulness of semantic representations

Results

Experiment testing impact of grammaticality

- We manually annotated 220 student answers for grammatical well-formedness.
  - 66% were ungrammatical
- Accuracy on this sample:
  - 83% for ungrammatical answers
  - 81% for grammatical answers

$\Rightarrow$ semantics-based approaches can be robust, not directly linked to grammaticality

Conclusion

- We presented a system for evaluating the content of answers to reading comprehension questions.
- Unlike previous content assessment systems, it is based on comparing formal semantic representations.
  - integrates a novel approach for comparing underspecified semantic representations
- Formal semantic representations readily support the integration of information structural differences.
  - connects content-assessment to information structure research in formal semantics and pragmatics
- The system presented outperforms our shallower CoMiC-DE system on the same CREG data set.
  - formal semantic representations can be competitive for content assessment in real-world contexts

References

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. URL http://aclweb.org/anthology/W08-0913.

Haghighi, A. D., A. Y. Ng & C. D. Manning (2005). Robust textual inference via graph matching. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 387–394.

Hahn, M. & D. Meurers (2011). On deriving semantic representations from dependencies: A practical approach for evaluating meaning in learner corpora. In K. Gerdes, E. Hajicová & L. Wanner (eds.), *Depling 2011 Proceedings*. Barcelona, pp. 94–103.

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*. pp. 9–15.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica* 55(3), 243–276.

MacCartney, B., M. Galley & C. D. Manning (2008). A phrase-based alignment model for natural language inference. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 802–811.

Meurers, D., R. Ziai, N. Ott & J. Kopp (2011). Evaluating Answers to Reading Comprehension Questions in Context: Results for German and the Role of Information Structure. In *Proceedings of the TextInfer 2011 Workshop on Textual Entailment*. Edinburgh, Scotland, UK: Association for Computational Linguistics, pp. 1–9. URL http://www.aclweb.org/anthology/W11-2401.

Mohler, M., R. Bunescu & R. Mihalcea (2011). Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments. In *Proceedings of the 49th Annual Meeting of the Association for Comnputational Linguistics*. pp. 752–762.

Ott, N., R. Ziai & D. Meurers (2012). Creation and Analysis of a Reading Comprehension Exercise Corpus: Towards Evaluating Meaning in Context. In T. Schmidt & K. Wörner (eds.), *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam: Benjamins, Hamburg Studies in Multilingualism (HSM). URL http://purl.org/dm/papers/ott-ziai-meurers-12.html.

Richter, F. & M. Sailer (2003). Basic Concepts of Lexical Resource Semantics. In A. Beckmann & N. Preining (eds.), *ESSLLI 2003 – Course Material I*. Wien: Kurt Gödel Society, vol. 5 of *Collegium Logicum*, pp. 87–143.

Rus, V., A. Graesser & K. Desai (2007). Lexico-syntactic subsumption for textual entailment. *Recent Advances in Natural Language Processing IV: Selected Papers frp, RANLP 2005* pp. 187–196.

Sammons, M., V. Vydiswaran et al. (2009). Relation Alignment for Textual Entailment Recognition. In *Text Analysis Conference (TAC)*.

Weiss, S. M. & C. A. Kulikowski (1991). *Computer systems that learn*. San Mateo, CA: Morgan Kaufmann.

Ziai, R., N. Ott & D. Meurers (2012). Short Answer Assessment: Establishing Links Between Research Strands. In *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-7) at NAACL-HLT 2012*. Montreal. URL http://purl.org/dm/papers/ziai-ott-meurers-12.html.

Full Results

Measure	BASIC	GIVEN	FOCUS
ALIGN	77.1		
EQUAL			
Student	69.8	75.3	75.2
Target	70.0	75.5	75.2
Average	76.6	80.8	80.7
IGNORE			
Student	75.8	80.1	80.3
Target	77.2	82.2	82.3
Average	79.8	84.7	84.9
WEIGHTED			
Student	75.0	80.6	80.7
Target	76.1	83.3	83.3
Average	80.9	86.1	**86.3**
CoMiC-DE	84.6		

From Alignment to Meaning Comparison

Information Structure: Example for Focus vs. Given

- Alternative questions: focused information determining whether answer is correct is explicitly given in question.

(3) *Ist die Wohnung in einem Altbau oder Neubau?*
is the flat in a old house or new house

(4) a. *Die Wohnung ist in einem* ***Altbau****.*
the flat is in a old house

b. *Die Wohnung ist in einem* ***Neubau****.*
the flat is in a new house

- All words in answers mentioned in the question, but some are **focused**, shown in boldface.