

Wie wurden die RTE-Evaluierungsdaten erstellt?

- IE, IR, QA, SUM
- vorhandene Anwendungssoftware
- nicht kontrolliert: Art der Entailment-Relation (welches Wissen wird benötigt?)
- 94% der T-H-Paare in RTE2-Trainingsset sind Paraphrasen
- erforderliche Inferenzschritte eher oberflächlich

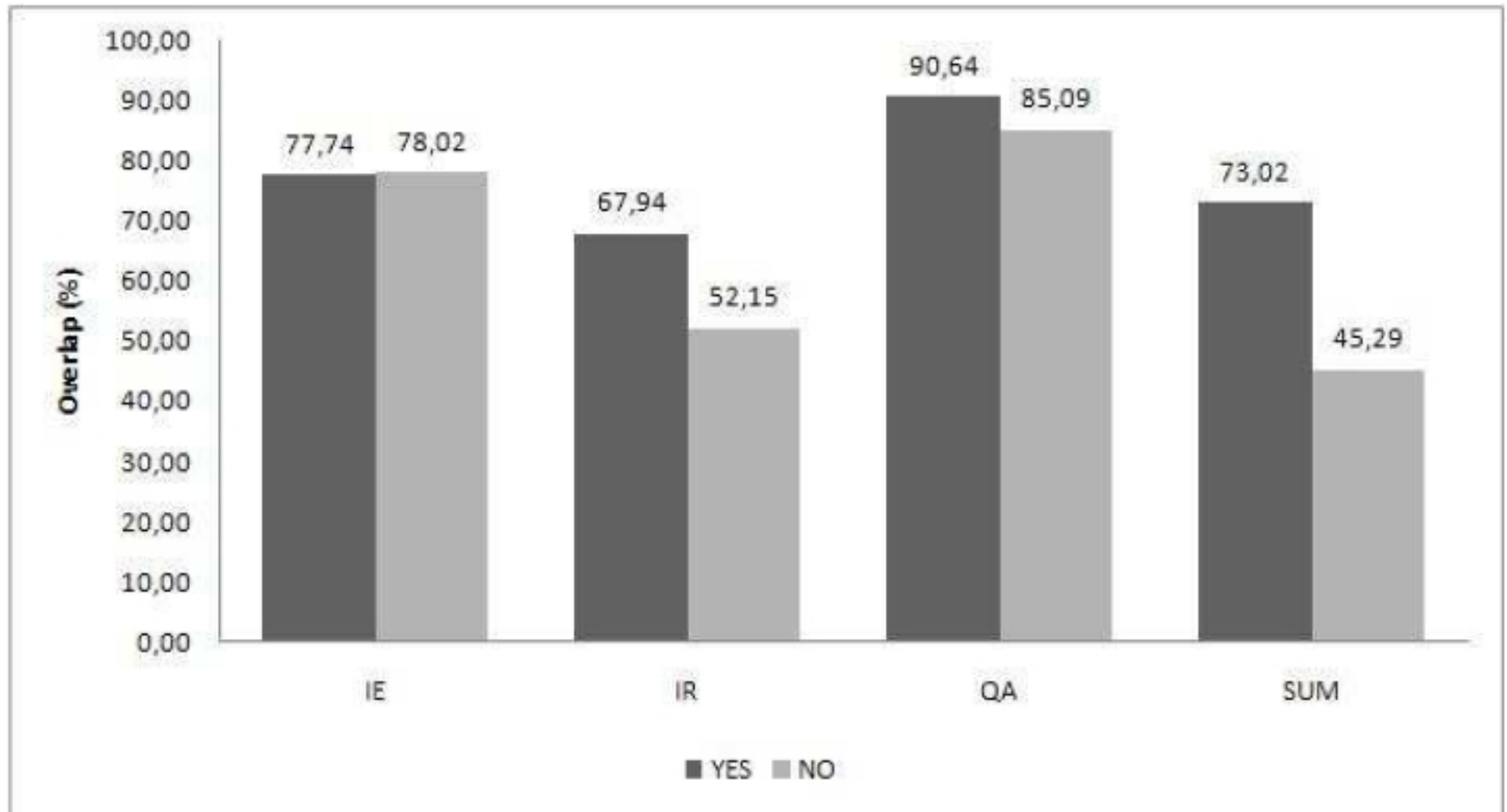
Analyse der RTE-Daten

- Annotating RTE (ARTE): Garoufi 2007
- drei Ebenen: Alignment, Context, Coreference
- Analyse von Word Overlap sowie Art des benötigten Wissens
- Word Overlap:

$$\text{Overlap} = \frac{\# \text{ common lemmas of T-H}}{\# \text{ lemmas of H}}$$

Analyse der RTE-Daten

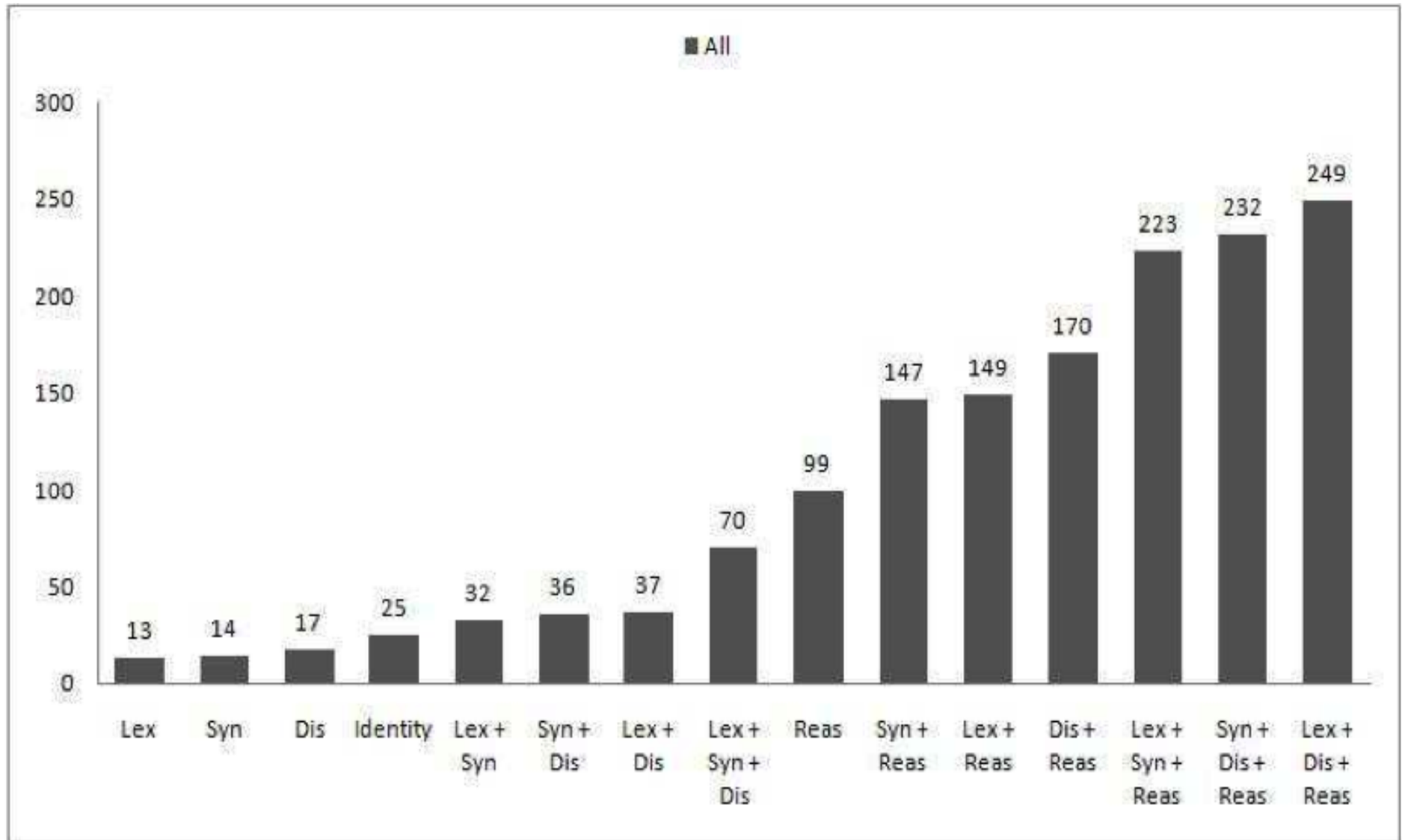
- Analyse:
Word Overlap



Analyse der RTE-Daten

- Arten des benötigten Wissens:
 - Identity
 - Lexical
 - Syntax
 - Discourse
 - Reasoning

Analyse der RTE-Daten



Analyse der RTE-Daten

	Inference	Overlap	Lexical DB	Alignment
Lex + Reas	0.5772	0.6154	0.6418	0.6412
Reas	0.5636	0.6395	0.6430	0.6515
Lex + Dis + Reas	0.5847	0.6290	0.6547	0.6565
Lex + Syn + Reas	0.5991	0.6275	0.6595	0.6587
Lex + Syn + Dis + Reas	0.6075	0.6345	0.6668	0.6673
Dis + Reas	0.5882	0.6633	0.6686	0.6767
Syn + Reas	0.5932	0.6567	0.6735	0.6822
Syn + Dis + Reas	0.6043	0.6691	0.6797	0.6858
Lex	0.8000	0.6982	0.7356	0.7170
Lex + Dis	0.7514	0.7110	0.7542	0.7568
Dis	0.7294	0.7285	0.7629	0.7710
Lex + Syn + Dis	0.7543	0.7253	0.7710	0.7791
Lex + Syn	0.7562	0.7308	0.7754	0.7757
Syn + Dis	0.7500	0.7393	0.7786	0.7907
Syn	0.7429	0.7692	0.8058	0.8265
Identity	0.9120	0.8554	0.8762	0.8971

BPI Test Suite

- 250 T-H-Paare
- Types of knowledge needed:
 - Synonyms
 - Hypernyms
 - Noun-adjective associations
 - mental - of the mind; Iraqi - of/from/in/to do with Iraq
 - Adverb-adjective associations
 - mentally - mental
 - Irregular singular/plurals
 - witness - person, witnesses - people
 - Noun-verb-adjectives relationships
 - strengthen - strong - stronger - strength

BPI Test Suite

- Types of knowledge needed (2):
 - Causative/inchoative/stative relationships
 - break the window -> window is broken
 - Role names and their relations to verb argument positions
 - buyer - buy, tennis player - plays tennis
 - Verb/Nominalization links
 - performance - perform
 - Proper name classification
 - Apple, Associated Press, Golden Retriever
 - Function/purpose of artifacts (telic roles)
 - airplanes fly, camera -> photo

BPI Test Suite

- Types of knowledge needed (3):
 - Typical co-occurrences/selectional restrictions
 - read x -> x is probably written material
 - General and specific world knowledge
 - increase -> greater/more, try to x -> want x
 - Actions: preconditions and effects
 - x buys y from z <- z has y
 - NN/PP disambiguation
 - test missile -> missile was tested, FBI secrets
 - Transfers: metonymy rules
 - Japan -> people of Japan, begin a fight -> fight
 - Other

Literatur

- K. Garoufi (2007): *Towards a better understanding of applied textual entailment*. MA thesis, Saarbrücken.
- <http://www.cs.utexas.edu/~pclark/bpi-test-suite/bpi-rte-knowledge-types.txt>