BIUTEE and the NIST*

(BIUTEE under Search)

Shachar Mirkin, Roy Bar-Haim, Jonathan Berant,
Ido Dagan, Eyal Shnarch, Asher Stern, Idan Szpektor

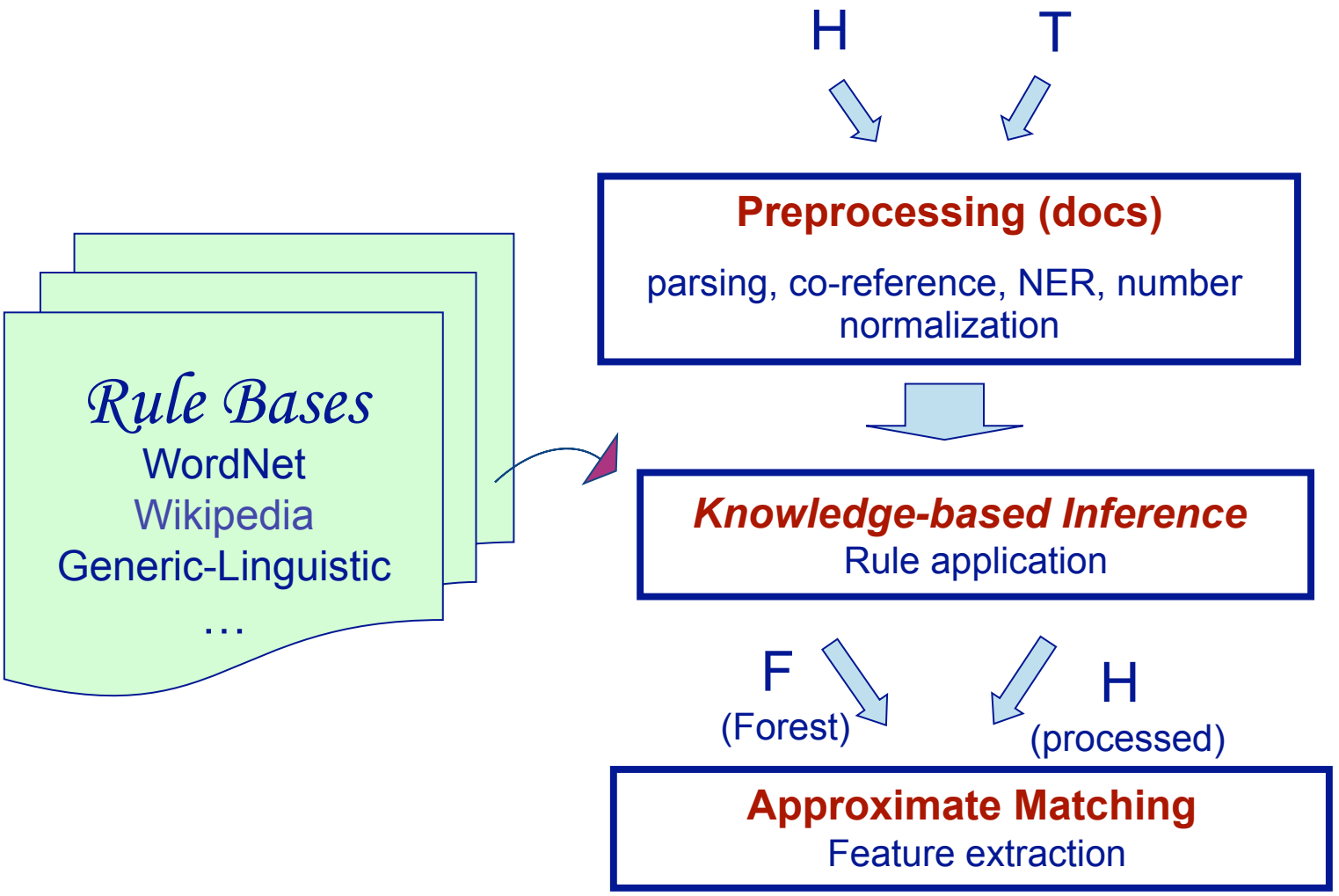
TAC 2009 / RTE Track

**** Bar-Ilan University's Textual Entailment Engine***

Outline

- *BIUTEE*
  - System architecture
  - Knowledge Resources
- Retrieval step
- *Discourse impact on inference*
  - Analysis of inference-oriented discourse phenomena
  - Our implementation to address some identified phenomena
- Submissions & Results
- Conclusions & Future Work

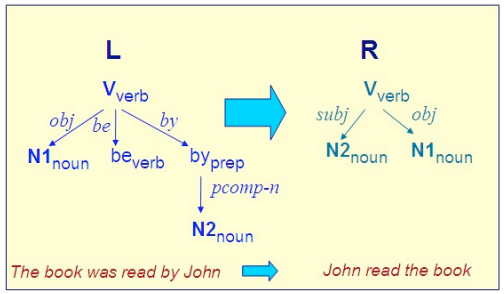
BIUTEE: System Architecture (as in RTE4)



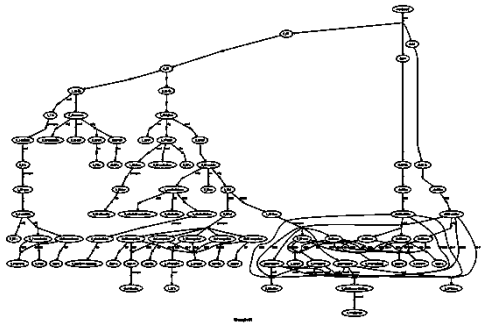
BIUTEE: Inference Rules are Tree Transformations

- Uniform representation for a vast range of semantic knowledge

- Single unified inference mechanism
  - Apply tree transformations
  - Rules can be chained (vs. alignments!)
  - Generate consequents



- Rule applications on T generate many consequent trees
  - Efficiently stored in a ***Compact Forest*** F (EMNLP-09)



BIUTEE: Approximate Matching

- Measure similarity between processed H and F
  - Compensate for knowledge gaps

Features:

- Coverage of H by F
  - Lexical coverage (words, verbs, numbers, named entities)
  - Local syntactic coverage (edges)
  - Global structural matching
    - Aim to match maximal sub-trees of H in F
- Predicate coverage in F
- Polarity mismatch (*forgot to buy* vs. *bought*)
- Argument match and coverage for corresponding predicates in F & H

Candidate Retrieval

- Dev set contains ~20K T:H pairs
- Only 810 (4%) are entailing
  - Assuming similar ratio on test set

- A naïve approach:
  - Reduce the task to T:H pairs
  - Apply main-task techniques on each pair

  - Inefficient
  - Won't be feasible in larger scale search settings (e.g. QA)

⇒ A prior step of candidate retrieval is necessary

Retrieving Candidates in RTE5

- Entailment-based query expansion
  - Using a set of entailment-rules resources – for recall increase
- Retrieval criterion:
  - Coverage percentage of H by the sentence
  - Future work: incorporate better IR scoring functions
- Resource-set & coverage percentage tuned to optimize inference performance
  - Rather than retrieval performance
- Similar flavor as “IR for QA”

Discourse Impact on Inference - Analysis

- Goal:
  - Identify & categorize discourse phenomena that impact inference
  - Prioritize according to phenomena distribution
- Analyzed a sample of the positive examples
  - Marking only relations that are relevant for inferring H
- Results guided our consequent implementation

Incorporating Anaphor Information

- Frequently, H includes the *antecedent* of an anaphor in T

⇒ Identifying the coreference relation needed to infer H

- Available tools miss many of these relationships

- Entailment knowledge resources may help :

  Kamchatka → eastern Russia

  - .. sometimes such information is missing or uncertain (example soon)

  ⇒ Useful to incorporate semantic knowledge for co-reference resolution

*H: The AS-28 accident happened in **eastern Russia***

T: The bathyscaphe submersible had only 24 hours of oxygen in reserve when it became stuck … in the bay of Kamchatka in far **eastern Russia***

T: The vessel rose to the surface at 4:26 p.m. local time … more than 600 feet below the surface off the Kamchatka Peninsula.

Compensating for Poor Performance of Co-reference Tools

Initial step - our implementation:

- Consider two NPs as co-referring if:
  1. Their heads are identical
  2. No semantic incompatibility is found between their modifiers (Note: relevant for entailment inference too)

- Implemented incompatibility types:
  - Antonymy: *first flight* vs. *last flight*
  - Mismatching numbers: *560 dollars* vs. *1,200,000 dollars*

- Further incompatibility types can be considered:
  - Co-hyponyms
  - Semantically disjoint modifiers
    - *first* vs. *second* ; *747's pilot* vs. *747's flight attendant*

Co-references Involving Verbal Predicates

- Out of the scope of most available co-reference tools
  - V-V or V-N
- Incorporating knowledge:
  - Considering the relatedness between *retreat* and *melt* can help identify the coreference relation
  - Not necessarily an entailment relationship
- Not addressed yet in our implementation

*H: The ice is **melting** in the Arctic*

T: The **melting** ice may also affect polar bears, and whales, who live off the sea life beneath the ice.*

T: "Everyone wants to know: Is the ice retreating because of global warming?

Implicit Information Required to Infer H

- Many entailing sentences refer **implicitly** to information required for inferring H
  - May be viewed as bridging anaphora [Thanks, CELCT]

- A prominent case - **"Global"** information:
  - Mentioned at the beginning of the document (title / first few sentences)
  - Assumed known from that point on

- Initial implementation:
  1. Identify key terms in each document - TFIDF
  2. Add top-k terms as nodes directly attached to the root of T

  ⇒ A global term found in the hypothesis is lexically matched in each sentence
  - Even if not explicitly mentioned

*H: Mine accidents cause deaths in **China***

T: TWO MORE MINE ACCIDENTS IN **CHINA** BRING WEEK'S DEATH TOLL TO 60*

T: So far this week, four mine disasters have claimed the lives of at least 60 workers and left 26 others missing

Cross-documents Coreference Resolution

- Quite often, cross-document co-reference resolution is needed for inferring H
  - Not available in typical co-refernce tools
- Usually involved alternative names of the same object
  - *Xena : ub313*
  - *Submarine : AS-28*
  - (Once identified) can be solved by a substitution of terms
- Not addressed yet in our implementation

Locality of Entailment

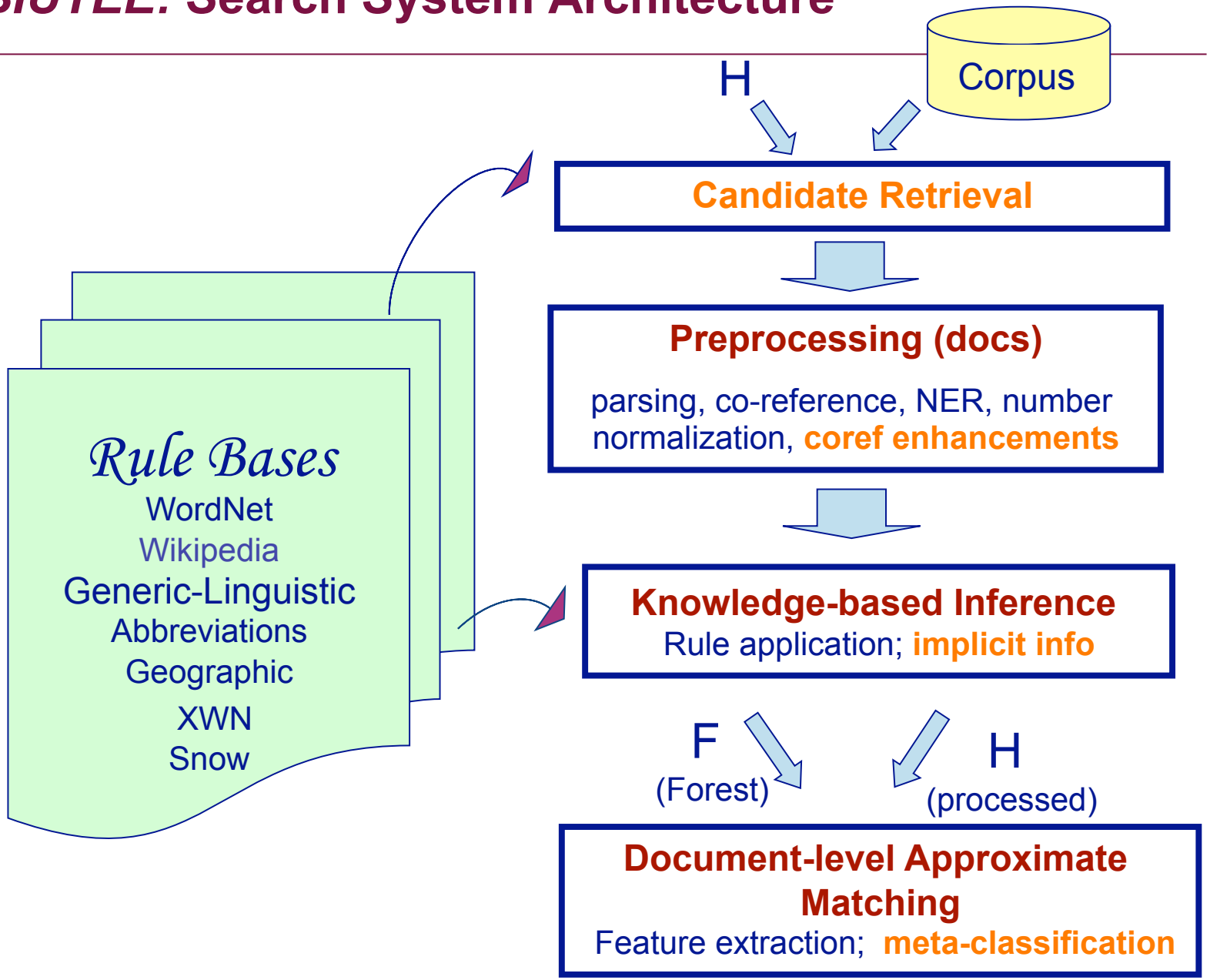
- **Assumption**: Entailing sentences tend to come in bulks
  - For discourse coherence, discussion of a specific issue is continuous
    - Especially in long documents

$\Rightarrow$ If a sentence entails *H*,

adjacent sentences are more likely to entail it as well

- Addressed by a **meta-classifier**
  1. Base classifiers make initial entailment decisions
  2. Meta-features computed to "smooth" classification positions and reflect bulks of entailments
     - Used by the meta-classifier in a 2nd classification pass

BIUTEE: Search System Architecture



Submissions

BIU1: Lexical Coverage

- Determine entailment purely based on term coverage of *H* by *T*
  - using the retrieval system's output directly
- Experimentally picked Wiki resource with a 50% coverage threshold
  - Overall, resources for lexical entailment rules did not contribute much

BIU2: BIUTEE at sentence-level

- Single classifier, with all knowledge resources
- Features extracted for each sentence separately
- Test-set sentences pre-filtered by the retrieval system
  - no resources for expanding retrieval
- Include "globally prominent" words in each sentence

BIU3: BIUTEE at document-level - Our complete system

- *BIU2* +
  - Document-level features
  - Meta-classifier, SVM & Naïve-bayes

Results

- Micro-averaged results:

Run	Suggested Sentences	P(%)	R(%)	F1(%)
Search-BIU1	1199	37.03	**55.50**	44.42
Search-BIU2	946	40.49	47.88	43.87
Search-BIU3	1003	**40.98**	51.38	**45.59**

Conclusions

- First step towards addressing the search task
  - Identified key issues, initial solutions
- *Major contribution*: analyzing discourse impact on inference, identifying needed research in:
  - Discourse technology to support inference needs
  - Inference technology to incorporate discourse information
- Complete system just slightly surpassed lexical baseline
  - Simple lexical methods are initially (yet again) difficult to beat
  - Still, document-level processing is helpful
- Open questions
  - Can we improve lexical match by entailment expansions?
  - Can we surpass lexical methods in summarization search?

Future Work

- Analysis , analysis , analysis
  - Resources, features, components
- Lexical methods
  - Incorporate IR/QA know-how
  - Improve expansion algorithms
- Reconsider our approximate matching component
  - May improve syntactic/semantic inference contributions
- Discourse:
  - Co-reference: better performance, incorporate verbal expressions, identify implicit references
  - Inference: utilize the above info

Thank you!

Questions?