Combining Deterministic Dependency Parsing and Linear Classification for Robust RTE

Alexander Volokh, Günter Neumann, Bogdan Sacaleanu

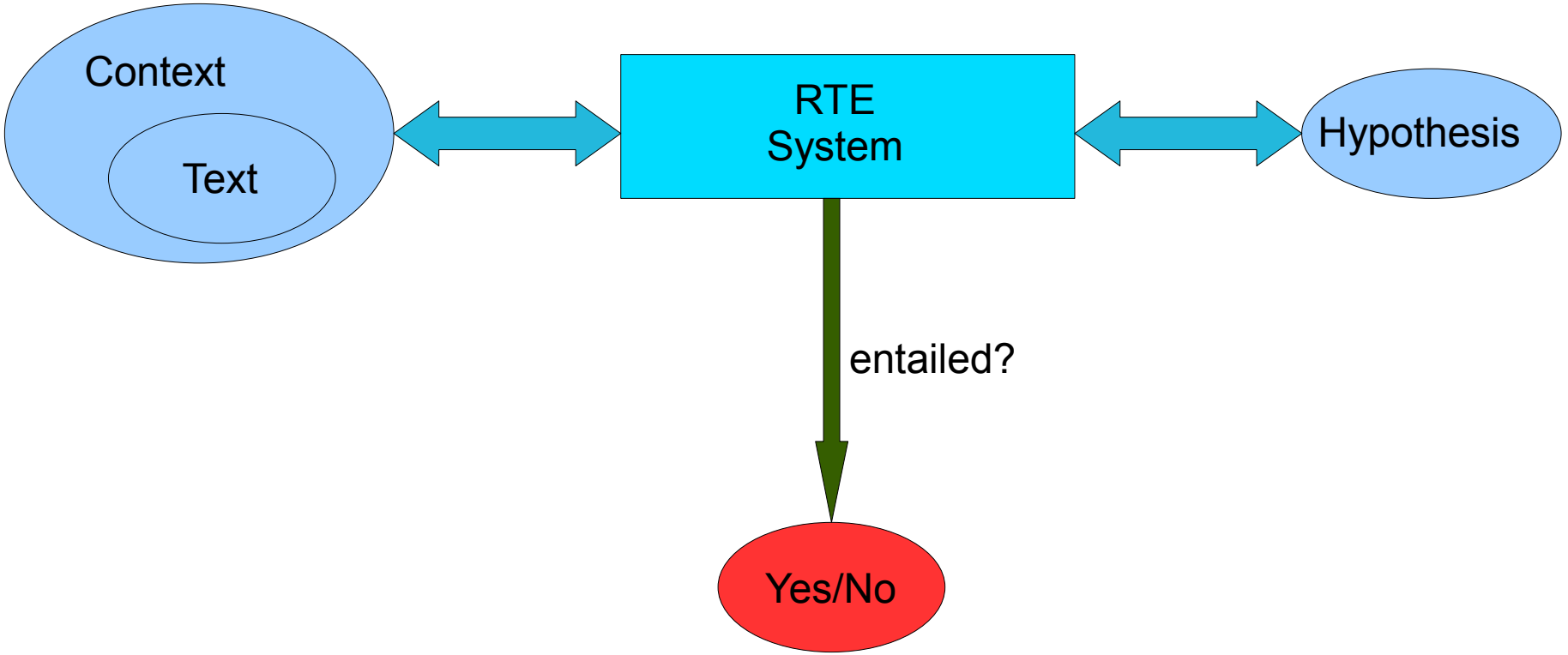
DFKI, Saarbrücken, Germany

German Research Center for Artificial Intelligence GmbH

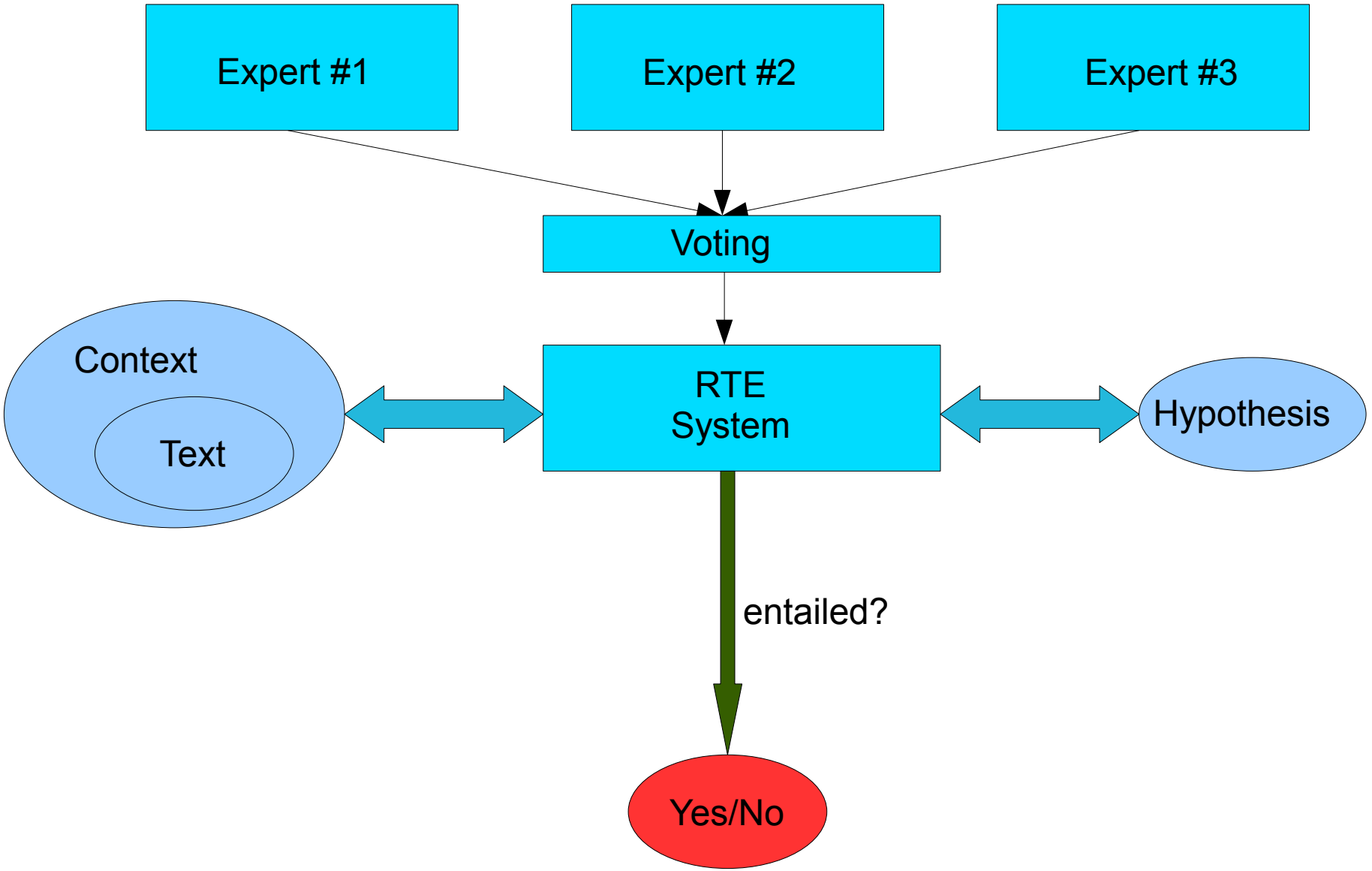
Outline

- RTE6 Main Task:
  - Former DFKI approaches for RTE
  - Feature-based approach of this year
- Parser comparison via RTE

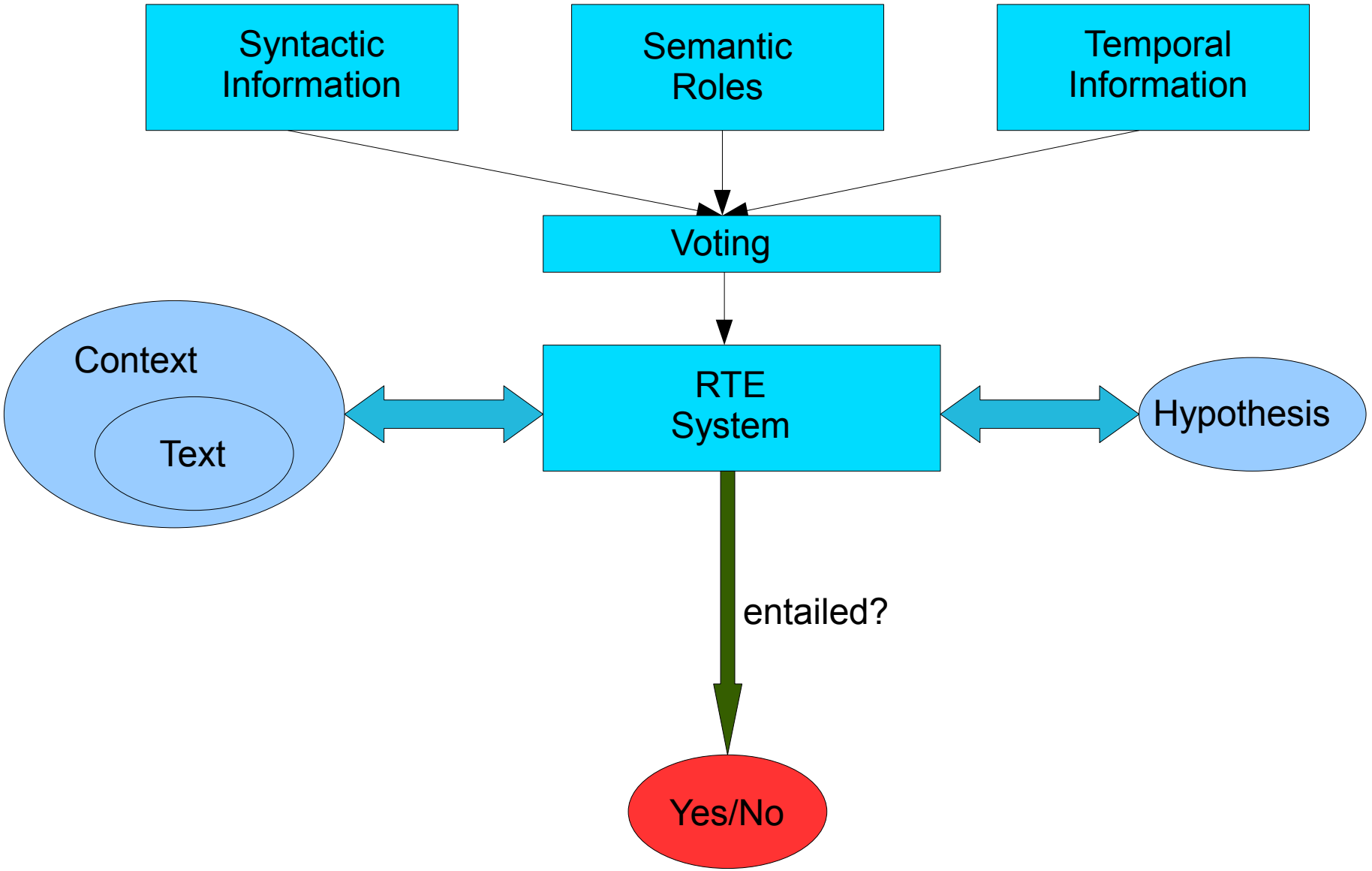
General RTE Architecture



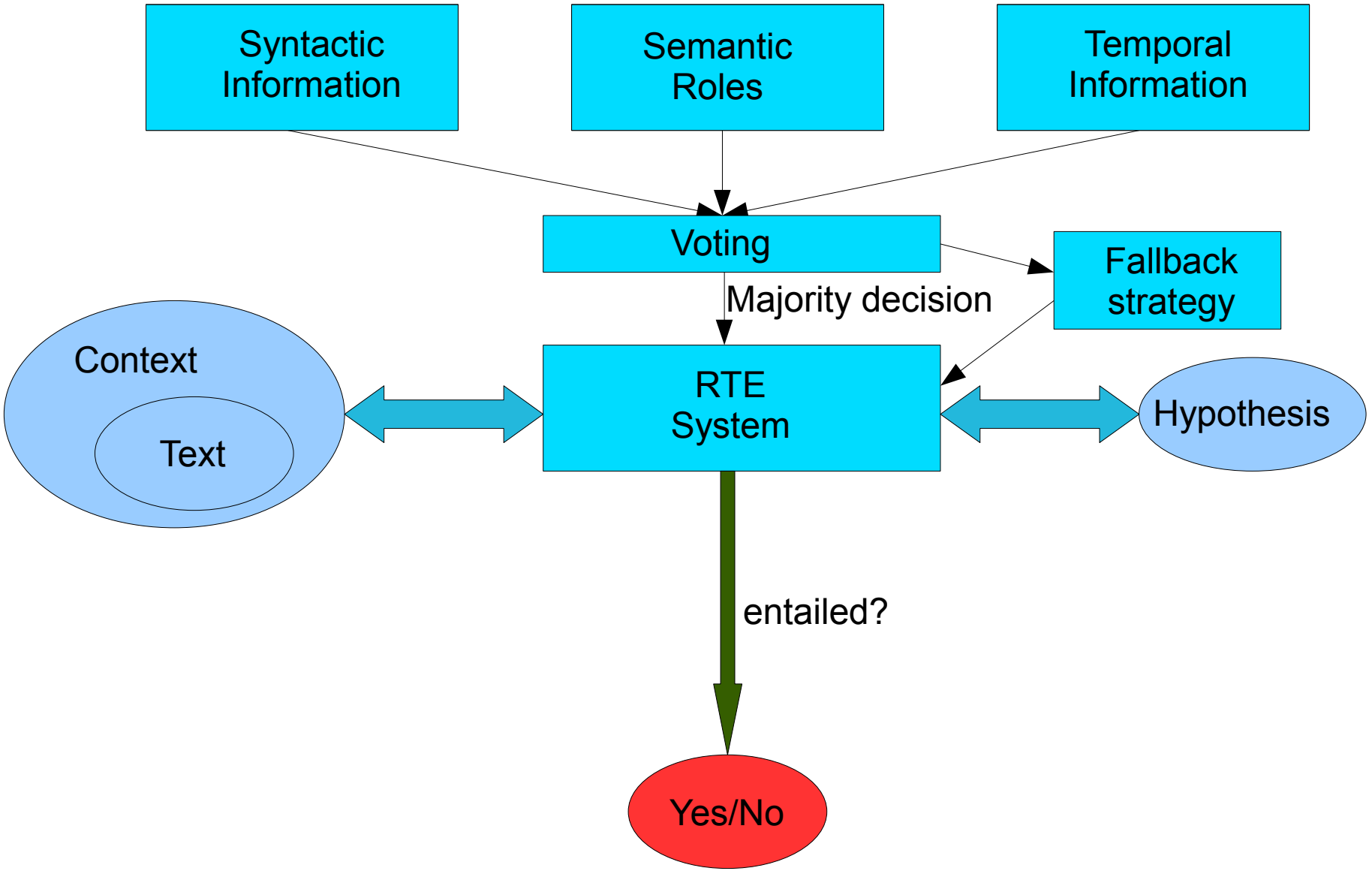
Architecture with Voting



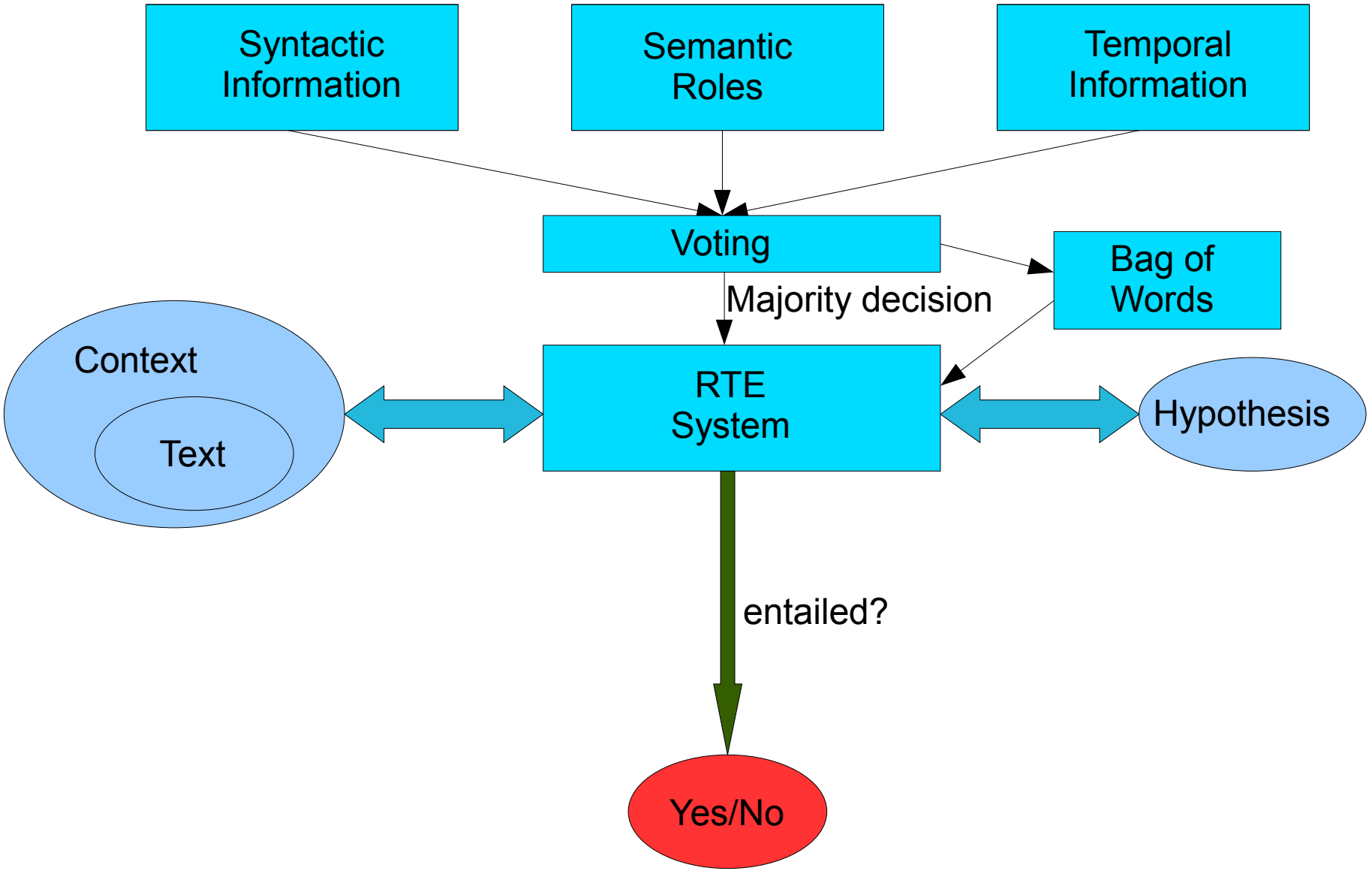
Architecture with Voting



Architecture with Voting



Architecture with Voting



RTE-6 Changes

- This year's RTE task:

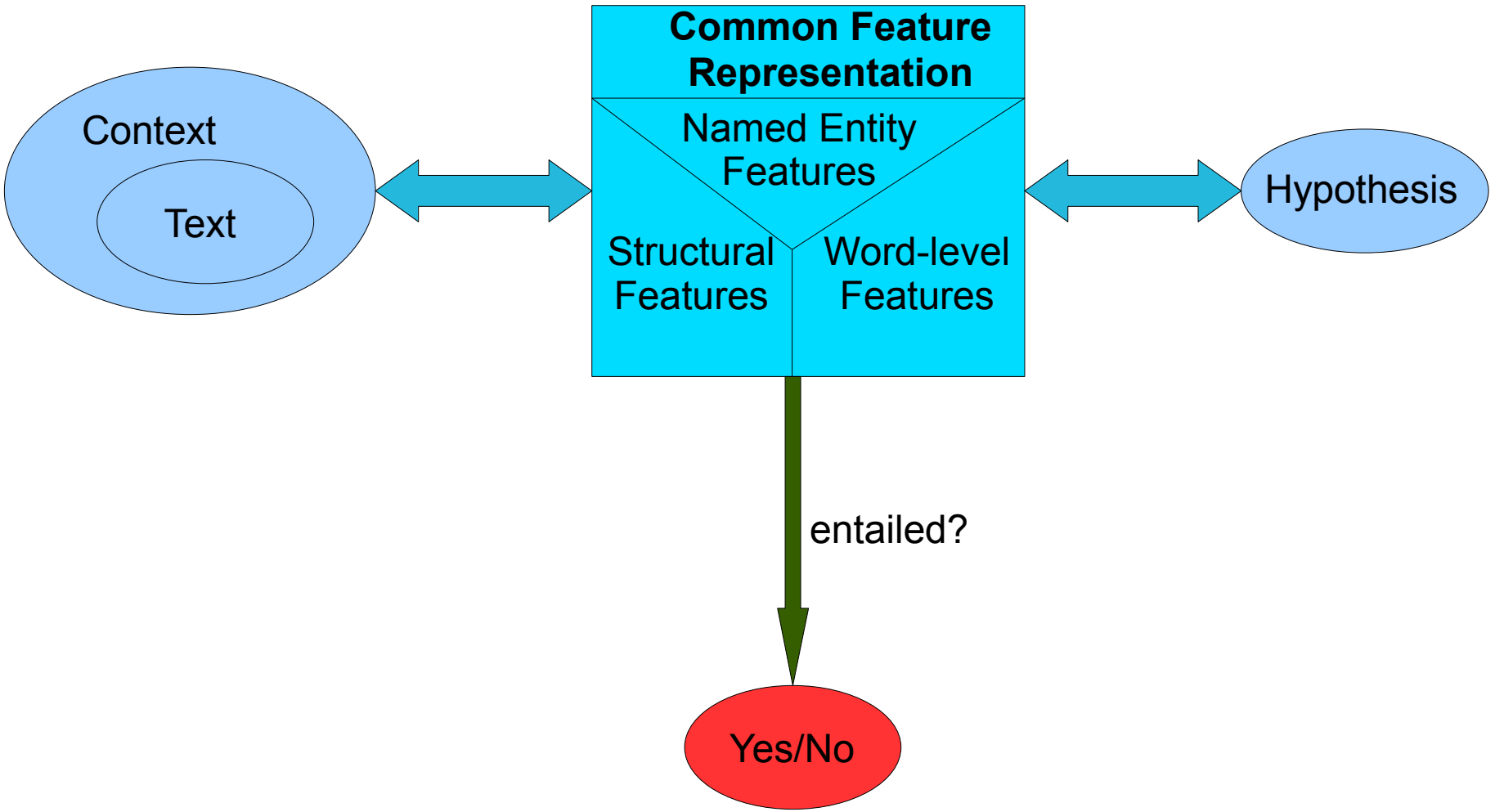
  - Explore the influence of the different knowledge sources

  - Each of the “arrows“ is a factor when computing the contribution

  - Main / Fallback strategies make the computation even more difficult

  → Principal change of architecture

Feature-Based Architecture



Ablation Tests

- **Official F-Score:** 38.26
- **Features:**
  - Dependency structure comparison(1,2,3,4)
  - Overlap and similarity on word level(4,5,7)
  - Named entity overlap and type similarity(6)
  - Coreference resolution(8)
- **Classifier Parameters:**
  - Threshold for YES/NO classification

Test #	F Measure	Impact	Left out Features
1	35.31	2.95	1 (root features)
2	39.11	-0.85	2 (depth 1 features)
3	38.54	-0.28	3 (depth 2 features)
4	33.27	4.99	4 (word form and pos features)
5	19.22	19.04	6 (content word features)
6	36.04	2.22	7 (named entities features)
7	38.49	-0.23	5 (WordNet similarity features)
8	36.72	1.54	Coreference resolution features. No additional features were introduced or left out, but the content of all T-H-pairs was first processed with the LingPipe coreference resolution tool.
9	39.10	-0.84	Threshold 0.15
10	39.11	-0.85	Threshold 0.13

PETE

- Parser Evaluation using Textual Entailments (PETE), task #12 at SemEval-2, 2010 (Yuret et al.)

  - **subject-verb** dependency

  "John kissed Mary." --> "John kissed somebody."

  - **verb-object** dependency

  "John kissed Mary." --> "Mary was kissed."

  - **noun-modifier** dependency

  "The big red boat sank." --> "The boat was big."

- Natural evaluation instead of LAS/UAS
- Formalism independent comparison

Parser Comparison

- Two parsers: MaltParser(Nivre et al.), MDParser

	Parsing Time	Sentences per Second	Tokens per Second
MDParser	73.188s	46.128	1015.55
MaltParser	1954.684s	1.73	38.02

- RTE results:
  - MDParser: f-score = 38.26
  - MaltParser: f-score = 39.81

Summary

- We thank RTE6 for the opportunities:

  - implementation of a robust RTE approach

  - parser comparison in a real-world application

- RTE6 shows that no specific data is necessary to evaluate parsers

  - on the contrary to PETE data more than syntax was necessary, but the comparison worked