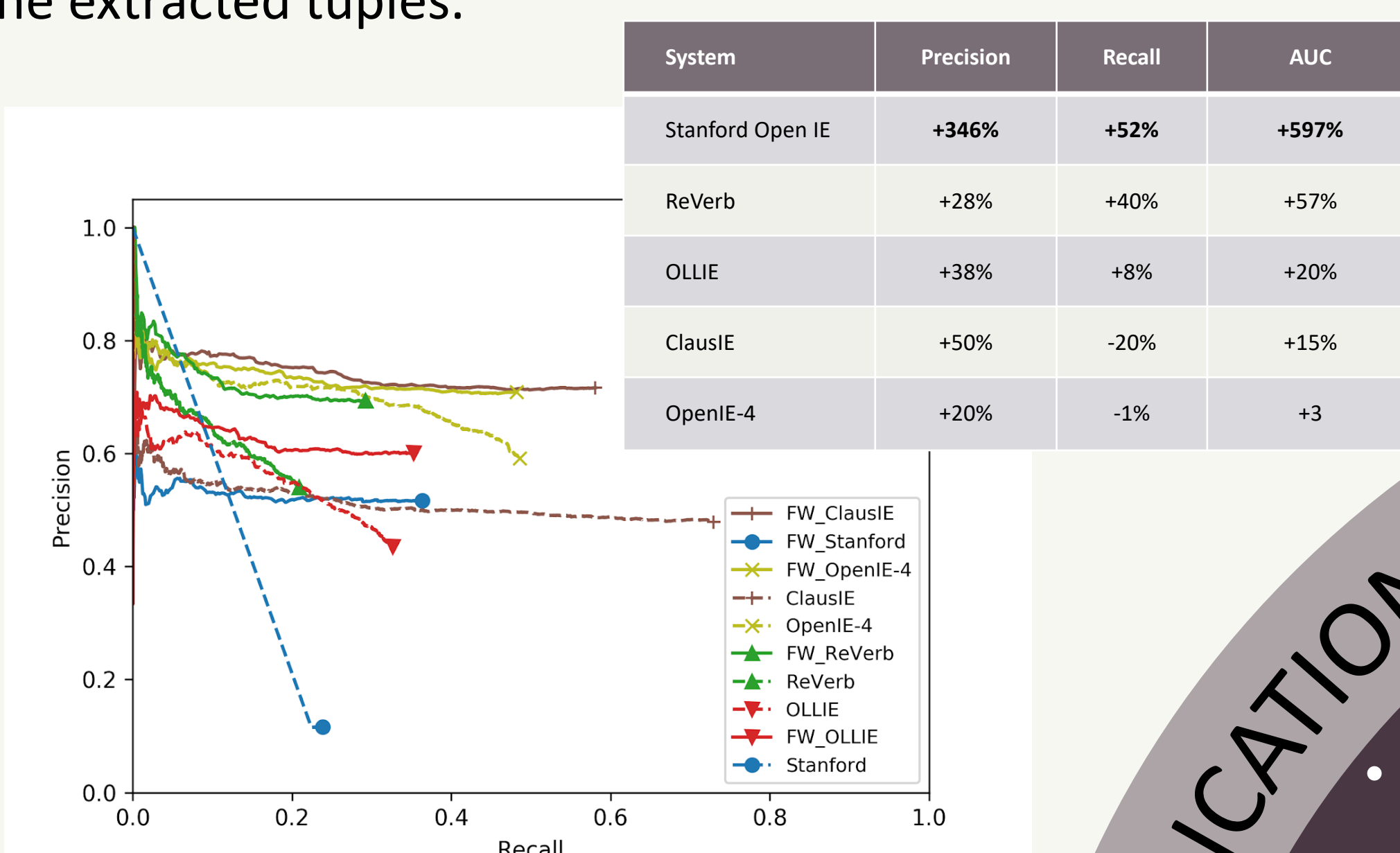


“A fluoroscopic study known as an upper gastrointestinal series is typically the next step in management, although if volvulus is suspected, caution with non water soluble contrast is mandatory as the usage of barium can impede surgical revision and lead to increased post operative complications.”

- hard to comprehend by **human readers**
- difficult to analyze by **semantic applications**

improves the performance of SOTA Open IE approaches in terms of **precision** (up to 346%) and **recall** (up to 52%), i.e. leads to a **lower information loss** and a **higher accuracy** of the extracted tuples.



Lightweight semantic representation can be used to **facilitate and improve** a variety of AI tasks.

Supervised-OIE (alone):

- (1) (A fluoroscopic study; known; as an upper gastrointestinal series)
- (2) (caution with non water soluble contrast; is; mandatory as the usage of barium)
- (3) (as the usage of; barium can impede; surgical revision and lead)
- (4) (; to increased; post operative complications)

Supervised-OIE (using DisSim for preprocessing):

- (5) #1 0 (A fluoroscopic study; is; typically, the next step in management)
- (6) ELABORATION #2
- (5b) CONTRAST #3
- (6b) #2 1 (This fluoroscopic study; is known; as an upper gastrointestinal series)
- (7) #3 0 (Caution with non water soluble; is; mandatory)
- (7a) CONTRAST #1
- (7b) CONDITION #7
- (7c) BACKGROUND #4
- (7d) BACKGROUND #5
- (7e) BACKGROUND #6
- (8) #4 1 (The usage of barium; can impede; surgical revision)
- (8a) LIST #5
- (8b) LIST #6
- (9) #5 1 (The usage of barium; can lead; to increased operative complications)
- (9a) LIST #4
- (9b) LIST #6
- (10) #6 1 (The usage of barium; to increased; post operative complications)
- (10a) LIST #4
- (10b) LIST #5
- (11) #7 1 (Volvulus; is suspected;)

Semantic Hierarchy: enriches the output with important **contextual information** that is needed for a **proper interpretation** of complex assertions.

- DisSim outperforms the SOTA in structural text simplification:
- It generates a **fine-grained output with a high level of grammaticality and meaning preservation**.
- It achieves an improvement of 5%, 4% and 6% with regard to the SAMSA score [3] against the second-best performing approach.
- It can be applied in a **domain independent** manner.
- The full evaluation methodology and detailed results are reported in [4].

A comparative analysis with the annotations contained in the **RST Discourse Treebank** [5] demonstrates that

- DisSim captures the **contextual hierarchy** between the split sentences with a **precision of 90%**, and
- DisSim reaches an **average precision of 70%** for the **classification of the rhetorical relations** that hold between them.

- **35 (English)** and **29 (German)** hand-crafted grammar rules
- Patterns encode **syntactic and lexical features** that can be derived from a sentence's parse tree

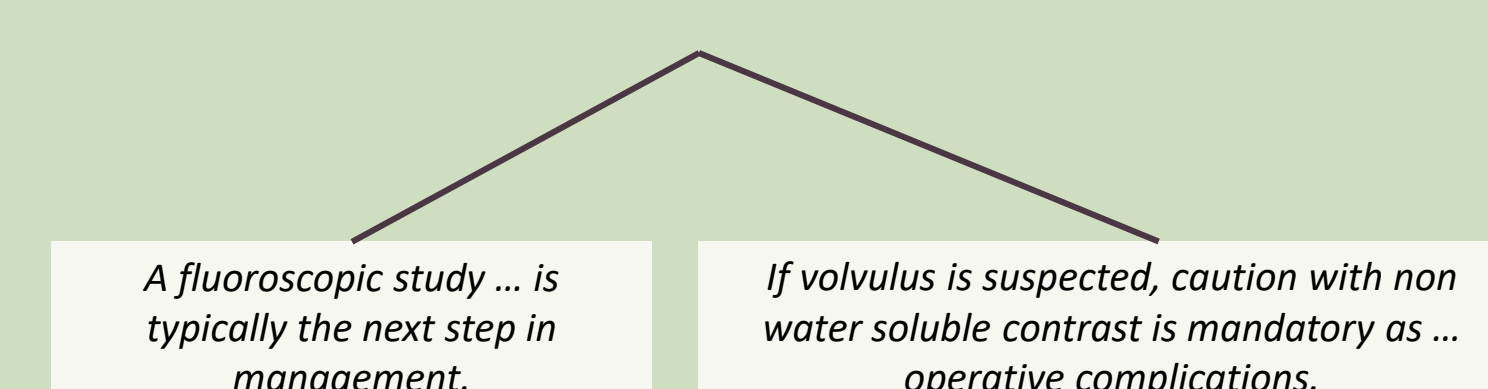
RULE	TREGEX PATTERN	EXTRACTED SENTENCE
SubordinationPostExtractor	ROOT <<: (S < (NP \$.. (VP < + (VP) (SBAR <, (IN \$+ (S < (NP \$.. VP)))))))	S < (NP \$.. VP .

1. How to **split up and rephrase** the input?
2. How to set up a **semantic hierarchy**?

Split into Minimal Propositions

a sequence of **sound, self-contained utterances**, with each of them presenting a **minimal semantic unit** that cannot be further decomposed into meaningful propositions

Simplification rules encode both the **splitting points** and **rephrasing procedure** for **disembedding clausal and phrasal components** and transforming them into syntactically simplified stand-alone sentences.



simpler, more **regular structures**

easier to +

loose arrangement of self-contained units that **lack important contextual information**

impedes interpretability

Establish a Semantic Hierarchy

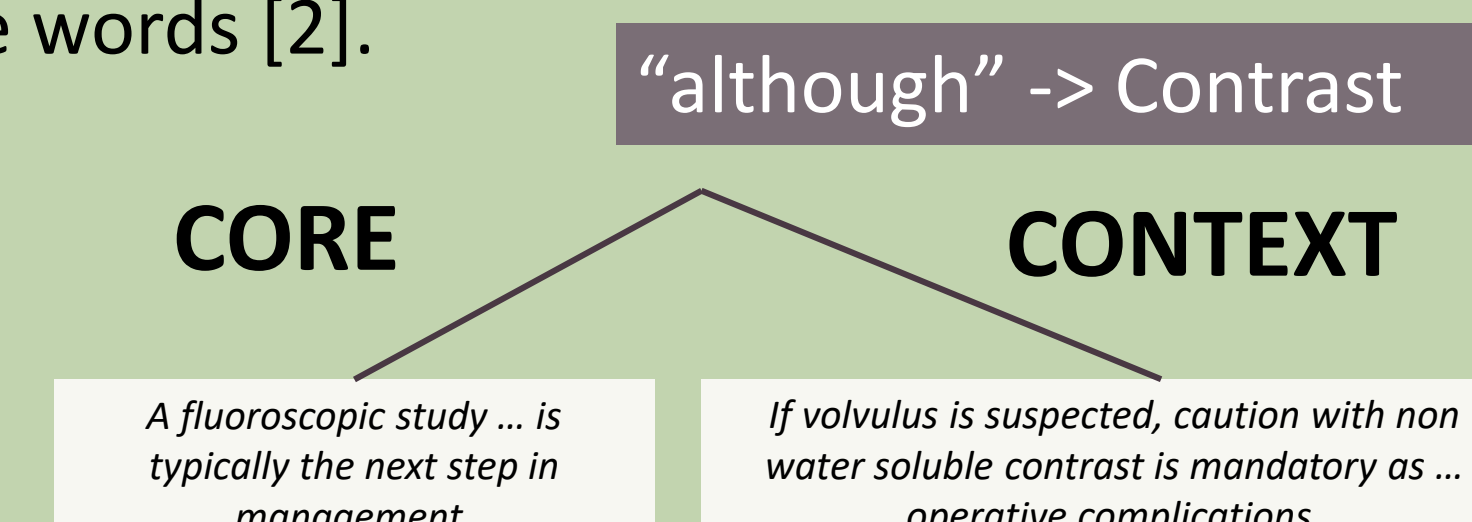
establishes a **contextual hierarchy** between the split sentences by connecting them with information about their hierarchical level (**core vs. context**) based on the concept of **nuclearity in RST** [1].

hierarchically ordered sentences

- restores the **semantic relationship** between a pair of split sentences by **identifying and classifying the rhetorical relation** that holds between them.

**semantically
interconnected
sentences**

- is based on syntactic and lexical features to extract **cue phrases** from the parse tree which are mapped to rhetorical relations using a predefined list of rhetorical cue words [2].



logical structure of utterances that present a simplified syntax

OUTPUT