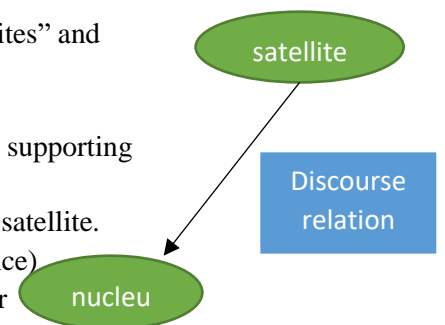TEXT COHERENCS

- Human languages do not normally consist of isolated,unrelated sentences,but contains collocated,structured,linked groups of sentences.
- We refer to such structured group of sentences as a discourse
- The word coherence refers to the relationship between sentences which is not just random assembling of sentences.
- Example :jane took train from paris to Istanbul. she had to attend a conference
- The sentence#2 provides a reason to the sentence #1.
- The coherence relation connecting two sentences is "Reason".
- There could be many coherence relation that exit between sentences, paragraphs and even between documents
- Rhetorical structure theory(RST) is one such theory that provides these coherence relations.
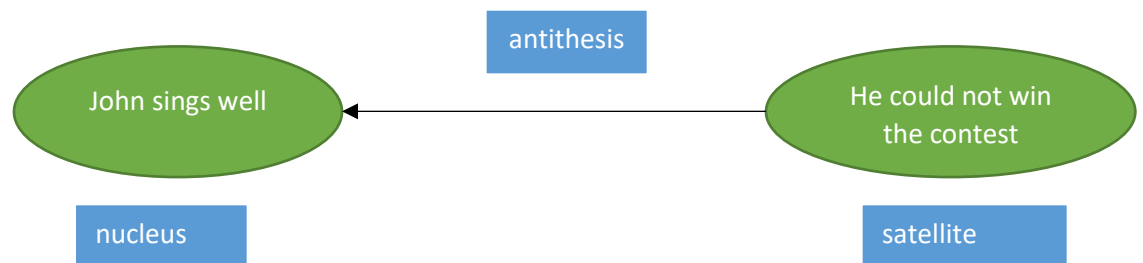
RST BASICS

➢ RST is a descriptive theory which focuses on the organization of the natural language.
➢ RST has proved to be a powerful tool in analyzing the disourse of a text since 1987.
➢ RST mainly captures the coherence in the fragments.
➢ The coherence is found between two/more text spans in the text.
➢ A text span could be a sentence , paragraghs, and even documents.

- The building blocks of rst are "nucleus","satellites" and "discourse relations"
- The nucleus carries the necessary information
- The satellites carries the additional information supporting the nucleus.
- The discourse relation connect the nucleus and satellite.
- Nucleus-discourse relation -satellite(nrs sequence)
- The RST based discourse structure are used for Many NLP application .
- Summary generation
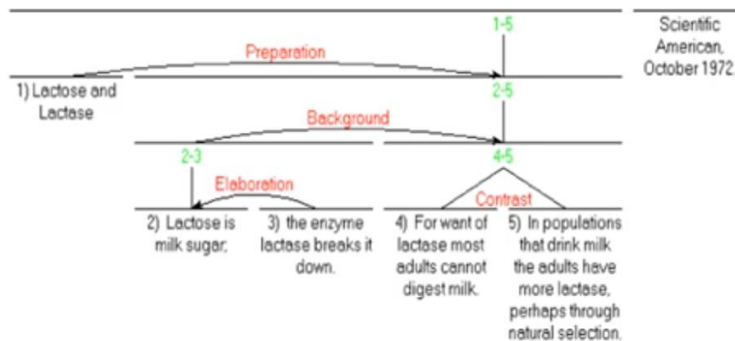- Question answering system
- Information retrieval system

satellite

Discourse relation

nucleu

John sings well but he could not win the contest

antithesis

John sings well

He could not win the contest

nucleus

satellite

RST Based Discourse Structure
Figure Taken from
:https://www.sfu.ca/rst/01intro/intro.html

HOW TO BUILD DISCOURSE STRUCTURE

- Given a sequence of sentences, we can automatically determine the coherence relations between them.(i.e)building discourse structure
- This task is called discourse parsing.
- Discourse parsing is done in two stages.
- 1.identification of elementary discourse units(EDU)
- 2.labelling the coherence relation.
- Standard data sets
- Penn discourse tree bank (PDTB)-the final dataset contains roughly 18000 explicit relations and 16000 implicit relations.
- RST discourse treebank-makes a global tree structure making a higher level discourse structure from PDTB.

k