Use Hans Rosling video from 2:30 to 10:00

<https://apps.dtic.mil/sti/pdfs/ADA625019.pdf>

At item [1], Rosling begins by highlighting the ignorance many of his students have regarding Demography and Demographic Economics. He emphasises the pressing need to communicate the data related to issues such as life expectancy, child health and economic inequality. He explains that the software being used in his presentation is designed to make the data easier to understand. This is an example of the preparation relation, as the audience is informed of the types of data to be discussed in the presentation. The audience are told what to expect and are informed of the information’s importance.

The contrast relation, items [2]-[3], summarises the most pronounced and statistically significant distinction between wealthy, developed societies and poorer, third world societies. In the developed world, people have smaller families and live longer lives, whilst in poorer societies, people die younger and have more children. [2] is treated as the nucleus and [3] is the satellite by virtue of [2] being mentioned first, though the distinction is immaterial in this context.

Items [5]-[8] are a sequence of facts as they directly follow one another, and each serve an identical literary purpose. These items are also satellites that constitute evidence for claim [4], which is the nucleus.

Similarly, [9] and [10] both include similar information, as they both provide explanations for the statistical changes alluded to in item [4] and evidenced in [5]-[8]. Items [9]-[11] constitute evaluations of item [4], as they assess the information contained in and related to the nucleus item [4]. These items are extrapolations Rosling made by comparing the data in the presentation with knowledge of historical events and government policy decisions which the demographic changes can be attributed to. Item [12] provides background, using Vietnam as an example to illustrate the fact highlighted in item [11].

**Rules for heuristic**

Highest priority will be given to the Joint relation. There are no constraints on the join relation, meaning the purpose of any join will be subjective, depending entirely upon its context and the intentions of the writer. The only assumption that can be made is that the author intends any joint text spans to be delivered together. The lack of constraints makes the Joint relation unpredictable and it is therefore safer to prioritise Joints over other relations unless specifically directed to do the contrary.

<>

Preparation: SECOND HIGHEST PRIORITY. Satellite comes before nucleus. Likely to start the presentation. Should it depend on how many relations the nucleus of the preparation has? More other relations the nucleus has, the more important, probably

Background: Nucleus before satellite, though it can work both ways, it will require more complex algorithms to determine the best word order for this relation. Typically, it is safer to start with the nucleus because sometimes the information contained in the satellites will be harder to understand or will seem less important without the nucleus

Contrast: THIRD HIGHEST priority, as speaker wants the two items to be juxtaposed, in order to prove a point

Sequence: FOURTH highest priority. What ever happens after the sequence will be determined by the relations in the last entry in the sequence

Joint: HIGHEST PRIORITY

Evidence: Nucleus comes before the satellite as the purpose of evidence is to improves the audience’s belief. As with background there is a chance the information won’t make sense or will seem less important if the audience isn’t first aware of why they are being presented with it

Evaluation: Nucleus comes before satellite. Evaluation should come after evidence

**Defining Rhetorical Structure Theory**

Thomson and Mann (1987) define Rhetorical Structure Theory (RTS) as consisting of schemas which govern how text spans, blocks of text, can be built of other portions of text. These schemas serve as relations between each block of text and help to dictate the text’s meaning.

Many of these relations can be described as nucleus-satellite relations. This relation involves one clause, the satellite, providing information that complements that contained in the main clause, the nucleus.

A relation which involves a satellite clause dependant on a nucleus clause is known as a mononuclear relation. An example of a mononuclear relation is ‘I bought this laptop, because it has a good processor’ (Hoogervorst et al 2016). Demonstrably, the latter clause cannot exist without the preceding clause and is therefore of inferior importance to it.

A parataxis or multinuclear relation includes two nuclei, both of equal importance to the overall meaning of the text block. Examples of these relations include the Contrast and Concession relations.

RST Relations can also be categorised by their purpose. Presentation relations exist to influence the feelings and attitudes of the readers about a particular subject. These include providing justifications or motivations for or a restatement of the sentiments expressed by the nucleus.

Subject Matter Relations differ from Presentation Relations in that their effect is solely to provide additional information to the reader (Wark & Nowina-Krowicki 2015).

Each RST relation will enumerate four criteria. The relation must impose constraints on the nucleus, constraints on the satellite, constraints on the combination of nucleus and satellite and must achieve a desired effect on the reader.

|  |  |  |
| --- | --- | --- |
| **Name of Relation** | **Description** | **Example** |
| Solutionhood | Satellite text span poses a problem which the nucleus span addresses | What if you're having to clean floppy drive heads too often?  Ask for Syncom diskettes, with burnished Ectype coating and dust-absorbing jacket liners |
| Elaboration | Nucleus span is accompanied by additional information | As members of the University's staff, you are cordially invited to attend the 1983 Annual Staff Breakfast presented by President James Zumberge and the Staff Assembly.  The continental breakfast and get-together will be held in the Town and Gown Auditorium (on Main Campus) at 8:30 AM on Thursday 11/3. |
| Background | Satellite span provides information which adds to user’s understanding of the nucleus span | I am having my car repaired in Santa Monica (1522 Lincoln Blvd.) this Thursday 19th.  Would anyone be able to bring me to ISI from there in the morning or drop me back there by 5 pm please? |
| Contrast | The idea presented in the nucleus is distinguished from the idea presented in the satellite | I recently purchased a text which purported to be a guide to Pascal for engineers.  It totally ignored the subtleties of the language and made no bones about it |
| Concession | There is a potential contradiction between the nucleus and the satellite but they are intended to be compatible with each other | Your kind invitation to come and enjoy cooler climes, is so tempting, but I have been waiting for the outcome of medical diagnosis and the next 3 mon |
|  |  |  |

**Figure 1**: A list of Relations defined by Thomas and Mann (1987), including examples taken from the same article

**Applications of RST**

The use of RST in automated generation of text has been explored and its effectiveness has greatly varied depending on its application.

One example is the STOP software, which generates messages encouraging smokers to quit smoking (Reiter, Stripada & Williams, 2003). Readers undertake a questionnaire before being provided with a letter formulated to appeal to their own personal circumstances.

A weakness of this experiment was the inability to gather a large quantity of information from a survey. As participants cannot realistically be expected to spend large amounts of time completing the survey, there is a limited number of questions that can reasonably be asked. Important factors are noticeable absent from the survey.

For example, participants experiencing stress can be reminded of the link between tobacco and stress. However, other factors, such as participants turning to tobacco to deal with grief after the death of a loved one, are not accounted for.

It is worth noting that studies have also asserted that non-tailored letters were more effective at encouraging smokers to quit, in comparison to subjective tailored letters (Lennox et. al, 2001).

Another example of the application of RST is the SumTime-Mousam project, which is used to automate the formulation of weather reports (Reiter et al, 2005). The automation process consisted of three stages: Document Planning, Microplanning and Surface Realization.

Document Planning determines the lexical structure of the report. This structure includes the information the report will highlight, such as wind speeds, temperature and likely changes in weather data throughout a given time period.

Microplanning relates to how sentence structure, grammar and vocabulary should be used throughout the report. Microplanning involves lexicalisation, aggregation and referring expressions generation. Lexicalisation converts non-linguistic data (e.g.: statistics) into words. Aggregation relates to how the information should be divided throughout the blocks of text in the article and include factors such as sentence lengths and the number of sentences. Referring expressions generation involves determining how to refer to entities once they have been introduced. This could include the use of shorthand acronyms or pronouns.

Surface Realisation will create the weather report based on the decisions made in the Document Planning and Microplanning stages. This process requires domain communication knowledge to clarify the appropriate grammar rules and vocabulary for the relevant domain. The domain means the genre or topic at hand, in this case weather reporting (Kittredge et. al, 1991). Weather reporting is performed by humans using a sublanguage called ‘weatherese’ and must be simulated through AI in the same manner.

In the field of automated text generation, domain communication knowledge is particularly important when using relations with minimal restrictions and flexible applicability (https://journals.sagepub.com/doi/pdf/10.1177/1461445606064836). The least constrained relations are among the most difficult to manage through AI and include the joint relation, which connects two nuclei for a subject-dependant reason. It will be the domain and relevant sublanguage, in this case ‘weatherese’, which will inform the program which text blocks should be connected through a joint.

Another example of a practical use of RST is the Aspect-Based Sentiment Analysis (ABSA), which is used to analyse reviews of products and services. ABSA involves two stages: defining the important characteristics of the subject under review and determining which words in the text are describing those characteristics.

<More info about RST>

<http://cgi.sfu.ca/~mtaboada/docs/publications/Hoogervorst_etal_ICWE_2016.pdf>

<Cite> splits clauses into Elementary Discourse Units (EDUs)

Words contained in a relation are typically more significant than ones that aren’t

Steps

Construct the discourse tree: Sets the context and determines which parser to be used; need to consider the inter-sentence relationships

**References**

Hoogervorst, R, Essink, E, Jansen, W, Van den Helder, M, Schouten, K, Frasincar, F, Taboada, M 2016, *Aspect-Based Sentiment Analysis on the Web using Rhetorical Structure Theory*, Simon Fraser University, viewed 25th January 2021, <<http://cgi.sfu.ca/~mtaboada/docs/publications/Hoogervorst_etal_ICWE_2016.pdf> >.

# Lennox, A. S, Osman, L. M, Reiter, E, Robertson, R, Friend, J, McCann, I, Skatun, D, & Donnan, P. T 2001, ‘Cost effectiveness of computer tailored and non-tailored smoking cessation letters in general practice: Randomised controlled trial’, *British Medical Journal*, vol. 322, no. 7299, pp. 1396-1400, viewed 15th January 2021, <https://psycnet.apa.org/record/2001-07256-001>.

Reiter, E, Sripada, S, Hunter, J, Yu, J & Davy I 2005, ‘Choosing words in computer-generated weather forecasts’, *Artificial Intelligence*, vol. 167, no. 1-2. Pp. 137-169, viewed 15th January 2021, <http://www.sciencedirect.com/science/article/pii/S0004370205000998>.

Reiter, E, Sripada, S & Williams, S 2003, ‘Acquiring and Using Limited User Models in NLG’, *Proceedings of the 9th European Workshop on Natural Language Generation*, University of Aberdeen, pp. 87-94, viewed 25th January 2021, <http://oro.open.ac.uk/12455/1/enlg03.pdf>

Taboada, M & Mann, W 2006, 'Rhetorical Structure Theory: looking back and moving ahead', *Discourse Studies*, vol. 8, no. 3, pp. 423-449, viewed 15th January 2021, <https://journals.sagepub.com/doi/pdf/10.1177/1461445606061881>.

# Thompson, S & Mann, W 1987, 'Rhetorical Structure Theory: A Framework for the Analysis of Texts', *IPrA Papers in Pragmatics*, vol. 1, no. 1, pp. 79–102.

Wark, S & Nowina-Krowicki, M 2015, *Intelligent Multi-Media Presentation Using Rhetorical Structure Theory*, Australian Government, viewed 29th January 2021, <https://apps.dtic.mil/sti/pdfs/ADA625019.pdf>.