

TOPIC-COMMENT STRUCTURE OF TEXTS
(AND ITS CONTRIBUTION TO THE AUTOMATIC PROCESSING OF TEXTS)

Ferenc Kiefer

KVAL, Stockholm,
116 45, Södermalmstorg 8

Abstract

In general, topic is informally defined as that part of the sentence which the sentence is about and comment as what is said in the sentence. This is, of course, not a formal definition but so far no one has succeeded to provide a viable formal definition for these notions. In the analysis of sentences and texts these notions can nevertheless be used with considerable success.

Typically, topics are noun phrases and comments are open sentences (that is, sentences containing a variable). For example, the sentence John is writing a letter is about John and it says that he is writing a letter. Formally,

(1) $John_x$: (x is writing a letter)

where John is a variable binding expression and x is writing a letter is an open sentence in which the variable x is bound by John.

A sentence may have more than one topic. The sentence John saw Mary in the garden may be a statement about John and Mary and what is said about them is that he saw her in the garden. Formally,

(2) [$John_x$ & $Mary_y$]: (x saw y in the garden)

If a sentence has only one topic this topic is very often (but not always) the subject of the sentence. Consider the sentence John wants to see Mary with heavy stress on John (the sentence can, for example, be considered as an answer to the question Who wants to see Mary?) This sentence is evidently about Mary and says that John wants to see her.

Texts are built up from sentences and the topic-comment structure of a text will consist of the topic-comment structures of its constituent sentences. The topic-comment structure of texts is, however, not simply a cumulative function of the topic-comment structures of the constituent sentences. I shall present in my paper a series of observations concerning the topic-comment structure of texts on the basis of a legal text (Rules of the Arbitration Institute of the Stockholm Chamber of Commerce).

(i) Each text has a set of terms which are not defined (not introduced) in the text: they are taken for granted. These terms belong to the stock of common knowledge with respect to the given text (SCK). SCK defines the frame of the text (or its scenario). Thus, for example, in the text in question the following terms belong to SCK: Stockholm Chamber of Commerce, The Swedish Law of Arbitration, the parties, the arbitrator(s), the (arbitral) tribunal, the claimant, the respondent, the request for arbitration. It is a characteristic feature of texts that topics belonging to SCK are empty (non-informative) topics. If such a topic is found it should in general be replaced by a more informative topic. The latter can be reconstructed on the basis of the preceding text. For example,

(3) Each party shall be given an opportunity to state his views.

Taken in isolation, this sentence could be analyzed as having the topic each party and the comment it shall be given In the present context, however, the sentence is about the interpretation and correction of an award. Sentence (3) must thus be reconstructed as (4).

(4) Each party shall be given an opportunity to state his views on the award (or the interpretation of the award).

The topic of (6) is, then, the views on the award and it is stated about this topic that each party shall be given an opportunity to express it.

(ii) In several cases the topic of the sentence is given by the heading of the whole paragraph. On the surface the topic of sentence (5) is the Institute.

(5) The Institute may request a party to amplify any submission to the Institute.

However, as indicated by the heading of the paragraph, what is at stake here is the amplification. The question which one would ask about (5) is (6):

(6) What about the amplification of a submission?

(iii) Typically, the topic of a sentence is a definite noun phrase and a term introduced in the comment is an indefinite noun phrase. Taken in isolation this observation may lead to incorrect analyses. Consider, for example,

- (7) The claimant shall pay a registration fee and both parties shall deposit with the Institute a sum to cover the costs of the proceedings.

In the given context, this sentence is not about the claimant and the parties but rather about the registration fee and the deposit.

(iv) Conditional sentences raise special problems. Take

- (8) If an arbitrator appointed by a party dies,
...

Such a sentence may be an answer to the question (9).

- (9) What happens if an arbitrator appointed by a party dies?

Formally, conditional sentences can often be

analyzed as implications: $A \supset B$. In such sentences, as shown by (8) and (9), the if-clause ($=A$) is the topic and the comment is that this clause (i. e. the underlying proposition) entails the consequent clause (i. e. the underlying proposition), that is, we get the structure (10):

$$(10) A_x: (x \supset B)$$

In connection with if-topics several problems must be solved: the function of if-topics in texts, the domain of if-topics, the internal structure of if-topics. Each of these problems shall be given ample consideration in my paper.

On the basis of the above observations I am going to stipulate a general strategy for the determination of the topic-comment structure of texts. In addition, I shall briefly report on an experiment, made at KVAL, concerning the implementation of a model based on topic-comment structure. In this model questions are answered by means of topics. Finally, I shall present some arguments in favor of the approach chosen.