

A PRAGMATIC CONCEPT OF THEME AND RHEME FOR MACHINE
TRANSLATION

Christa Hauenschild

SFB 99 - E2, Universität Konstanz, Postfach 5560,
D-7750 Konstanz

The concept of theme and rheme we want to propose in this paper is pragmatic in two senses of the word: the theme/rheme structure (TRS) is viewed as a strategy of communication and thus belongs to pragmatics, and it seems to offer a pragmatic solution for a problem of machine translation.

The starting-point of our considerations is defined by several arguments claiming the necessity to preserve the TRS of sentences (or utterances) in the process of translation, as stated in Koenitz, 1978:

- the exhaustive-listing reading of the rheme
- the presuppositional character of the theme
- the connexion of the TRS with the degree of textual coherence
- the correlation of themeness and definiteness
- general considerations in a framework of communication theory (readability of texts etc.)

Further arguments might easily be added to this list (e.g. the relation between the TRS and the scope of operators). In principle all of them hold for machine translation as well, at least if it is pursued within a theoretical framework, as e.g. in the case of the project SALAT (cf. Hauenschild, 1978) and of the successor project "Übersetzungsbezogene Kontexttheorie" (Context theory for translation) in the spirit of which this paper was written.

These considerations lead to the question whether it is necessary to compute the TRS of every translated sentence, which would be very difficult - if at all possible - given the high degree of ambiguity of written texts with respect to the TRS.

We want to propose a somewhat simplified procedure which nevertheless ensures the preservation of the TRS. Theoretically, this procedure is based on a concept of theme and rheme that ascribes them a strategic role in communication, controlling, as it were, several other factors (semantic and syntactic). If these latter factors are kept constant during the translation process, we can be sure that we have preserved theme and rheme as well (or that the boundary is irrelevant in the given sentence).

The central hypothesis of this paper is the following: in most cases it is sufficient to preserve the underlying (communicatively conditioned) order of main elements, which is conceived in a way very similar to that of Sgall and Hajičová, 1977/78, or Hajičová, 1980, but without indication of an exact boundary between theme and rheme. The validity of this hypothesis is, of course, limited to certain languages (we treat German, French and Russian) and to certain types of texts, aiming exclusively at an optimal transmission of information. With such a limitation, it is very often possible to derive the underlying order in a straightforward way from the surface order of the main constituents of a sentence.

It might seem surprising that this should hold for such languages as German, French and Russian, which are normally viewed as exhibiting different types of word-order regularities. But if we look at texts of the mentioned kind (e.g. reports for international organizations), we find that in all three languages they are organized in a similar way, obeying more or less to some kind of "linearity principle": sentences

normally start with the elements connecting them to the preceding context, the elements carrying "new information" are situated at the end. Such organization seems to be most effective in the sense of an optimal transfer of information.

Thus we propose a procedure that derives the underlying from the surface order of the main constituents using some transformations for special cases (e.g. cleft-sentences). The synthesis will yield the corresponding surface order of the target language, again with a minimal set of transformations.

The procedure is to be controlled by several semantic factors that are closely related to the TRS (compare the arguments for the preservation of the TRS in translation mentioned at the beginning), but have to be kept constant on independent grounds:

- anaphoric relations
- definite or indefinite (or generic) interpretation of noun-phrases
- scope of quantifiers, negation and other operators

References

- Hajičová, Eva: A dependency based specification of topic and focus I; background and motivation. *Journal of Linguistic Calculus*, 1980: 1-2, pp. 93-109.
- Hauenschild, Christa: SALAT: System for Automatic Language Analysis and Translation. Paper read at the 7th International Conference on Computational Linguistics, Bergen (Norway) 1978.
- Koenitz, Bernd: Die Thema-Rhema-Gliederung - Invariante bei der Translation. *Linguistische Arbeitsberichte* 19, Leipzig 1978, pp. 75-91.
- Sgall P. and Hajičová E.: Focus on Focus (Part I and II). *The Prague Bull. of Math. Linguistics* 28, 1977, pp.5-54, and 29, 1978, pp.23-41.