

THE TRANSFER OF FINITE VERB FORMS IN A MACHINE
TRANSLATION SYSTEM

Bente Maegaard

Institut for anvendt og matematisk lingvistik,
Københavns Universitet
Njalsgade 96, DK-2300 København S, Denmark

This paper is based on work done jointly by Hanne Ruus, Ebbe Spang-Hanssen and the author, all of the University of Copenhagen. The work was done within the framework of EUROTRA and sponsored by the Commission of the European Community. Work on this paper was begun whilst the author was at ISSCO, Geneva.

Throughout this paper we shall imagine a machine translation system which performs the following: correct analysis of source language, correct structure of target language sentence, correct choice of words in the target language. In this environment we shall discuss some aspects of the problem of constructing the correct finite verbal form. The languages involved are the languages of the European Community. It is obvious that it is not possible simply to use the 'same' morphological form in the target language as in the source language: verbal tenses are not the same in all languages (not even in languages as closely related as those considered) and even where the same tenses exist they do not necessarily have the same distribution. So a mapping has to be defined.

As EUROTRA is in its conception multilingual this mapping should not be defined on a binary basis (between pairs of languages), but should rather consist in a definition of some basic and global interpretations of the morphology of verbs in the languages considered.

The study can not be restricted to morphology proper. Meanings that are expressed morphologically (by auxiliaries or flexives) in one language might be expressed lexically (by the use of particular lexical units) in other languages. As an example, take the progressive aspect in English: En. they are writing → Fr. ils écrivent / ils sont en train d'écrire, → → Danish de skriver / de sidder og skriver / de er i færd med at skrive. I.e. the progressive aspect which is expressed in English by morphological means, is in French and Danish either not expressed or expressed by certain lexical units.

It is very important to realize the impact of this. There is often a tendency to think that in the analysis of language a borderline can be drawn between morphological information and lexical information. This is true to a large extent when only one language is considered, but it is certainly not true in a multilingual environment. Roughly speaking, the system should take into account all pieces of information which can be expressed morphologically in at least one language.

The proposed system for the representation of verbal forms.

The system as such will contain the global information, i.e. the information which has to be transferred to other languages. In this paper only the transfer of active finite verb forms with respect to tense will be considered, i.e. modality, voice, etc. is not taken into account.

Apart from this global information the system can for each language be expanded with all kinds of local information which is useful for analysis or for generation. Some morphological phenomena are only grammatically bound and do not carry any semantic meaning that should be coded for the use of other languages, e.g. the use of the subjunctif in French subordinate clauses following il faut que. Information about this subjunctif can be regarded as internal to French and should not be expressed in the global system.

The variables proposed to describe the semantic content of tense and aspect are the following (Note that tense is

used in a very general meaning. Tense is determined by values of time and aspect. It should also be noted that the use of the concepts aspect and Aktionsart has been defined with respect to the languages involved and consequently is not necessarily the traditional one):

TENSE ::= (AXIS, PROSPECTIVE, RETROSPECTIVE, ASPECT,
AKTIONSART)

AXIS ::= NUNC | TUNC

PROSPECTIVE ::= PROSPECT- | PROSPECT+

RETROSPECTIVE ::= RETRO- | RETRO+ | CONTINUOUS | ...

ASPECT ::= null | EMPHATIC PROGRESSIVE

AKTIONSART ::= DURATIVE | NON-DURATIVE

Suggestions of other values (for English, for the examples below):

AKTIONSART ::= DYNAMIC | STATIVE or perhaps rather
ACTIVITY VERB | PROCESS VERB | TRANSITIONAL
EVENT VERB

For adverbs a value TIMEWHEN.

In the paper the use of the variables and their values will be explained. Here only a few remarks can be made.

1. The system is viewed as operating in all languages on two time axes, one for the tenses that are anchored in the present, and a parallel one for the tenses that are anchored in the past (NUNC and TUNC, respectively).

2. The RETROSPECTIVE value CONTINUOUS is used for events that are still going on: Fr. il demeure là depuis un mois and En. He has lived there for a month will get the same values: (NUNC, PROSPECT-, CONTINUOUS, null, DURATIVE).

3. More values for the RETROSPECTIVE variable is to be foreseen.

4. In English it is not possible to have the perfect tense with an adverbial of the type TIMEWHEN. So a rule of the generation grammar of English will say that (NUNC, PROSPECT-, RETRO+ ...,...) + TIMEWHEN → (TUNC, PROSPECT-, RETRO-, ...,...)

5. Certain verbs in English tend to prefer the progressive

aspect. Consequently the AKTIONSART must be expanded as to be able to express these facts.

A simulation of the system made on some Community texts shows good results.

A text is shown in the appendix in French (source) and in English. Both versions are official Community texts. For each verbal form which is finite in French and has a finite equivalent in English is shown how the system works.

a adopté (NUNC, PROSPECT-, RETRO+, null, NON-DURATIVE) +
+ TIMEWHEN --> adopted
a été (NUNC, PROSPECT-, RETRO+, null, DURATIVE) +
+ TIMEWHEN --> was
a augmenté --> rose

(NB. The TIMEWHEN value from the previous sentence must be used, otherwise we get the perfect tense)

a été -> was (because of TIMEWHEN)

ont été -> have been

ont baissé (NUNC, PROSPECT-, RETRO+, ..., DURATIVE) +
+ TRANSITIONAL EVENT VERB -> have been falling

s'est détériorée as above, with the AKTIONSART being PROCESS
VERB -> has been deteriorating

sont tombée à (NUNC, PROSPECT-, RETRO+, ..., NON-DURATIVE) ->
-> have dropped to

(NB. the English version in fact has dropped to, but as there is no TIMEWHEN this is not possible to obtain with the system. However it is possible to choose both translations from the French original).

Note furthermore, that tomber in the meaning 'decrease' is DURATIVE, while tomber à is NON-DURATIVE).

To our knowledge no other system has been proposed which works for more than one language. Reichenbach's system is made for English and shows deficiencies even for English. Bruce's system repairs these deficiencies and works for English, but relies heavily on the structure of English verbal forms and cannot easily be brought to work for other languages as well.

The system which we have proposed seems promising, but still a good deal of work is needed to define adequate values for the different variables.

References

- Bruce: A Model for Temporal References and Its Application in a Question Answering Program. Artificial Intelligence 1972, vol 3, pp 1-25.
- Greenbaum, Quirk, Leech, Svartvik: A Grammar of Contemporary English. Longman. London 1976.
- Maegaard, Ruus, Spang-Hanssen: Investigation of the Verbal System for EUROTRA, Report for the EEC, 1981.
- Reichenbach: Elements of Symbolic Logic. Macmillan. New York 1947, pp 287-298.

APPENDIX

Charbon

2.1.96. La Commission a adopté le 19 avril 1979, après consultation du Comité consultatif CECA (Bull. CE 3-1979, ~~point~~ point 2.3.75), le rapport annuel sur le marché charbonnier de la Communauté en 1978 et ses perspectives pour 1979 (JO C 120 du 14.5.1979).

Avec quelque 287 millions de tonnes, la consommation de charbon a été environ la même en 1978 qu'en 1977, cependant que la consommation du charbon destiné à la production d'électricité a augmenté de 5% environ pour atteindre 159 millions de tonnes environ en 1978 contre 151 millions de tonnes l'année précédente. Avec 238 millions de tonnes environ la production de charbon de la Communauté en 1978 a été inférieure de 1% à celle de 1977; elle se caractérise surtout par une légère augmentation de la production au Royaume-Uni et par de légers reculs ailleurs.

Les hausses de prix du charbon de la Communauté ont été modérées. Toutefois, comme les prix du marché mondial exprimés dans la plupart des monnaies de la Communauté ont baissé du fait de l'affaiblissement du dollar américain, la position concurrentielle du charbon de la Communauté s'est détériorée considérablement. Néanmoins, les importations de charbon en provenance de pays tiers ont diminué de 1,5% et sont tombées à 45 millions de tonnes, dont plus de la moitié de charbon vapeur.

Coal

2.1.96. After consulting the ECSC Consultative Committee (Bull. EC 3-1979, point 2.3.75), on 19 April the Commission adopted the annual report on the Community coal market in 1978 and the forecasts for 1979 (OJ C 120 of 14.5.1979).

At around 287 million tonnes coal consumption was about the same in 1978 as in 1977, while consumption for electricity generation rose by about 5% to reach roughly 159 million t in 1978 as compared with 151 million t the previous year. Community coal production in 1978 was about 238 million t or some 1% lower than in 1977, the main features being a slight rise in production in the UK and slight drops elsewhere.

Price increases of Community coal have been moderate. However, as world market prices expressed in most Community currencies have been falling because of the weakening of the US dollar, the competitiveness of Community coal has been deteriorating considerably. Nevertheless, coal imports from third countries dropped by about 1.5% to 45 million t, of which over half was steam coal.