THE SPECIFICATION OF TIME MEANING FOR MACHINE TRANSLATION

Frank van Eynde - Catholic University Leuven
Blijde Inkomststraat, 21, 3000 Leuven, Belgium

Louis des Tombe - Utrecht State University
Trans, 14, 3512 JK Utrecht, Holland

Fons Maes - Catholic University of Tilburg
Postbus 90153, 5000 LE Tilburg, Holland

In this paper, we put forward some ideas on the representation of time in a machine translation system. In such a system, we usually have the following four representations:

- source text
- source representation
- target representation
- target text

In an interlingual system, there is no difference between source and target representation; in a transfer-based system, the step between the two is usually called transfer, and this step is meant to be as simple as possible.

The research described was originally done in the framework of the EUROTRA MT project, which is transfer-based. However, it can be used in other MT systems as well; in fact, it is very well suited for interlingual systems.

The problem with time meaning is that it is expressed in natural languages in a way that is non-universal and, moreover, not very perspicuous prima facie. As a consequence, it is difficult to find rules for the translation of the tense form of the verb.

In this paper, we propose a conceptual calculus in which the meanings of language specific temporal expressions can be represented in an interlingual way, so that the translation of the latter can be achieved via the corresponding conceptual representations.

The exposition will consist of three parts.

First, we define a time axis model, i.e. a model in which temporal concepts can be understood.

Second, we establish two types of general constraints:

(i) Constraints on possible time meaning representations, resulting in a restricted class of meanings for time and related phenomena in terms of this model.

(ii) Constraints on the relations between syntactic/morphological forms and time meanings, resulting in a non-arbitrary relation between form and meaning.

Third, we show how the calculus can be used for the interlingual analysis of the tense forms of verbs.

1. The time axis model.

The model is a temporal structure $\langle time, < \rangle$, where

(i)
time is a set of elements called time-points;
(ii)
$<$ is a binary relation that linearly orders time (and can be interpreted as 'precedes');

An interval (I) is a subset of time that does not contain 'gaps', i.e.: $\forall$ t1,t2 $\in$ I $\forall$ t3 $\in$ time (t1<t3<t2 -> t3 $\in$ I).

We now turn to the time meanings and their representations.

First, we want to separate the expression that represents time meaning from the rest of the sentence. The instruments we use are based on Dowty (1979):

(i)
A two-place operator AT that takes an interval and a formula to yield another formula, with the following interpretation:

W(AT(I,0)=1 at whatever time t iff W(0)=1 at the interval I.

(ii)
Temporal predicates that take an interval to yield a formula, e.g.,

W(yesterday(I))=1 iff the interval I is a subset of yesterday.

(iii)
Temporal relations that take two intervals to yield a formula, e.g.,

W(before(I,J))=1 iff
∀ t,t' ∈ time (t ∈ I & t' ∈ J ---> t<t')

(iv)
λ-abstraction to separate the temporal expression from the basic proposition, so that the representation of the temporal expresssion takes the following form:

(1)
λ p ∃ I_1, I_2, ...⊆ time ($Rel_i(I_j,I_k)$ & ... & $Pred_m(I_n)$ & ... & AT (I_o, p))

where the I_i are intervals, the Rel_j are binary relations between intervals like 'before', the $Pred_k$ are predicates like 'yesterday', and p is a basic proposition, from which all time-relevant parts have been removed.

The category of expression (1) is t/t; it can be applied to a basic proposition in a functional way.

The interpretation of (1) is the set of propositions that are true at some given interval I_o. This is similar to Kripke's definition of the notion of 'possible world': 'A possible world is given by the descriptive conditions we associate with it....' (1972, p. 44). Analogously, a time interval can be identified with the collection of propositions that are true at it.

2. A theory of time meanings.

In many discussions of time meaning, a distinction is made between an internal and an external temporal system. The external system represents the temporal relation between the state of affairs as described by the basic proposition and the time at which the utterance takes place. This system always refers to the speaker or writer, and consequently it is a deictic system. The internal system is about such things as whether the state of affairs expressed in the basic proposition is described as going on, having just started, having been completed, etc. This type of information is often called aspectual.

In this paper, we adopt the following three basic principles for the representation of time meanings:

(I)
Each time meaning representation contains exactly three time intervals:
- the time of speech or narration (S)
- the time of event (E), i.e. the interval at which the basic proposition is said to be true
- one time of reference (R)

The S-interval consists of one point only: it is a singleton. The R- and E-intervals are non-empty subsets of time.

(II)
The deictic part of time meaning is represented by a binary relation between S and R and optionally by one predicate over R.

(III)
Aspect is represented by a binary relation between R and E, and optionally by one predicate over E.

Principles (I), (II), and (III) together imply that the general form of a time meaning representation can be somewhat simplified. It will now be:

(2)
λ p ∃ S,R,E ⊆ time ($Rel_1(R,S)$ & $Pred_1(R)$ & $Rel_2(E,R)$ & $Pred_2(E)$ & AT(E,p))

Apart from the constraints on possible time meaning representations there are some constraints on the relation between the time meanings and the language specific morphosyntactic forms for expressing those meanings:

(IV)
The predicates over R are those time adverbials that can be used as answers to when-questions, such as
(3) yesterday, now, next week, on Tuesday

(V)
The predicates over E are (a.o.) the duration time adverbials, such as
(4) for an hour, five weeks, since Christmas, until June

(VI)
The relations between R and S and between E and R are determined by the interaction of the verbal tense forms and the time adverbials in ways to be specified and exemplified in section three.

We will now present the deictic and the aspectual components of the temporal system in some detail.

2.1. The deictic system.

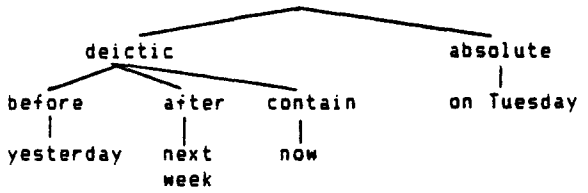
As possible relations between S and R we will take

(i) before (R,S), defined as in 1.

(ii) after (R,S), defined analogously

(iii) contain (R,S), defined as follows :
∀ t ∈ time (t ∈ S ---> t ∈ R)

The specifiers of the reference time are the when-adverbials. A classification of the latter that appears to be relevant for the assignment of deictic values in particular cases is the following one :



The deictic when-adverbials define the position of the reference time with respect to the time of speech, and cannot be combined with all possible tenses. An after-adverbial is, for instance, not compatible with the simple past:

(5) * he came next week

The absolute when-adverbials determine the position of the reference time independently from the speech time. Depending on which tense they are combined with they can either specify a reference time that precedes the speech time, as in

(6) she came on Tuesday

or a reference time that follows the speech time, as in

(7) she comes/is coming/will come on Tuesday

Since there is only one reference time in the representation (= principle (I)) and since the when-adverbials always specify the reference time (= principle (IV)), it is predicted that a proposition can contain at most one when-adverbial.

At first sight this prediction seems to hold: cf. the ungrammaticality of

(8) a. * He left yesterday one week ago
b. * In 1990 he will have arrived in 1998
c. * In 1955 he had died in 1944

There are, however, some problem cases, such as

(9) He left on Tuesday at 9 o'clock
(10) Last year he used to arrive at 9 o'clock

(9) contains two when-adverbials, but notice that they can be used together as an answer to one when-question, and this indicates that 'on Tuesday at 9 o'clock' is just a complex specification of one and the same interval.

(10) is a more serious case. Here the two adverbials cannot be considered to specify the same interval: 'last year' denotes the time of his habit to arrive at 9 o'clock and 'at 9 o'clock' denotes the time of each of his arrivals of last year. What we have in (10) is, in fact, an iterative interpretation, and for such interpretations we need a more complex representation format. This will not be developed in this paper, but see Van Eynde (forthcoming).

2.2. The aspectual part.

There is much discussion in the literature about what aspect is. A description that is not very precise, but has the merit of being independent of linguistic form, is the one given by Comrie (1976, p. 3):

'As the general definition of aspect, we may take the formulation that "aspects are different ways of viewing the internal temporal constituency of a situation".'

In an article on the general theory of aspect Friedrich distinguishes three possible aspects :

(i) punctual, completive, perfective, etc;
(ii) durative, continuative, etc;
(iii) stative, perfect, etc.

(cf. Friedrich 1974, p. 36)

The same three aspects turn up in the work of Comrie, Johnson, Hopper, and others.

We will call them respectively perfective, imperfective, and retrospective.

The intuitions about the three are basically the following:

(i) perfective
This aspect presents a situation 'as a single unanalyzable whole' (Comrie, o.c., p. 3).

(ii) imperfective:
This aspect 'looks at the situation from the inside' (Comrie, op. cit, p.4), and focusses on beginning, continuation, or ending of it.

(iii) retrospective:
This aspect 'expresses a relation between two time-points, on the one hand the time of the state resulting from a prior situation, and on the other the time of that prior situation.' (ibid., p. 52).

In order to make these notions more precise, and -at the same time- to integrate them into our representation format, we will adopt the following proposal by Johnson:

'What I am proposing concerning the semantics of the aspect forms is that they specify the relation between reference time and event time in an utterance.' (Johnson 1981, p. 153)

As applied to the different aspects this gives the following results :

(i) perfective:
In this case we take the relation between E and R to be one of containment (during (E,R)),where the latter is defined as follows:
during (x,y) iff ∀ t ∈ time (t ∈ x ---> t ∈ y)
The fact that E is contained in R is meant to be the formal counterpart of the intuition that E is seen as a single unanalyzable whole from the point of view defined by R.

(ii) imperfective:
This is subdivided into three classes:

(ii.i) durative: contain (E,R), defined as in 2.1.
(focus on the continuation)

(ii.ii) inchoative: since (E,R), definition:
since(x,y) iff x ∩ y ≠ ∅
& ∃ t' ∈ time ∀ t ∈ time (t ∈ x & t' ∈ y ---> t'<t)
& ∃ t ∈ time ∀ t' ∈ time (t ∈ x & t' ∈ y ---> t>t')
(focus on the beginning of E)

(ii.iii) terminative: Until (E,R), definition:
until(x,y) iff x ∩ y ≠ ∅
& ∃ t ∈ time ∀ t' ∈ time (t ∈ x & t' ∈ y ---> t<t')
& ∃ t' ∈ time ∀ t ∈ time (t ∈ x & t' ∈ y ---> t'>t)
(focus on the ending of E)

(iii) retrospective:
The relation is simply before (E,R).

Some authors also distinguish a socalled 'prospective' aspect (cf. Comrie 1976). It seems to be less common than the other ones, and there is some disagreement on the issue of what its language specific counterparts are ('to be going to' ?), but conceptually it can be defined fairly easily, namely as the complement of the retrospective aspect:

(iv) prospective: after (E,R)

The interval E can be specified by adverbials. One class of E-specifiers is the class of duration adverbials. The reasons for treating these adverbials as E-specifiers are the following ones :

1. they always denote the interval at which the basic proposition is said to take place; in that respect they are different from the when-adverbials, since the latter can also denote a time that does not coincide with the event time (cf. the non-perfective aspects).

2. they cannot be combined with all possible propositions; they are, for instance, not compatible with momentaneous events:

(11) they reached the summit for a while

The ungrammaticality of (11) can be explained if we take the duration adverbials to specify the event time, since the latter cannot be both a moment (as required by the proposition) and an interval of some duration (as required by the adverbial).

3. they never have a deictic function: they are not used for specifying the relation between some interval and the moment of speech.

As in the case of the when-adverbials it is possible to have two duration adverbials in the same clause:

(12) he has been studying two hours a day since his childhood now

Notice, however, that (12) has an iterative interpretation, and since the treatment of such interpretations requires a more elaborated representation scheme anyway, we can stick to the principle that a clause contains at most one E-specifier. In this case the E-specifier is 'since his childhood'; 'two hours' is another type of specifier (cf. Van Eynde, forthcoming).

2.3. The calculus as a whole.

In the preceding sections it has been stipulated that there are three possible relations between S and R, and six possible relations between R and E. At first sight that seems to be rather arbitrary, but a careful analysis of the concepts involved shows that they, in fact, exhaust the range of logical possibilities :

For any two intervals x and y ⊆ time,
either x ∩ y = ∅
and then either before(x,y)
or after(x,y)
or x ∩ y ≠ ∅
and then either x ⊆ y, i.e. during(x,y)
or ~(x ⊆ y)
and then either x ⊇ y,
i.e. contain(x,y)
or ~(x ⊇ y)
and then either since(x,y)
or until (x,y)

These are the six aspectual values. The reason why the deictic system has only three possible values is that the speech time - unlike the reference and the event time - is always a singleton, and if one of the intervals involved is a singleton, then the relations 'since' and 'until' and either 'during' or 'contain' cannot hold by definition. It appears, thus, that both the deictic and the aspectual distinctions are not only mutually exclusive but also exhaustive within their respective domains.

Together they form the core of the temporal calculus. This core has to be extended in various ways if one wants to take into account the phenomenon of iterativity, the sequence of tenses in complex sentences, and the relevance of the event type of the basic proposition (cf. Vendler's distinction of

states, activities, accomplishments, achievements). Part of this has already been incorporated in the formalism, but in stead of presenting those extensions we think it more useful to round off this paper with a demonstration of how the calculus can be used for the interlingual analysis of verbal tense forms.

3. The interlingual analysis of tenses.

For the interlingual analysis of the verbal tense forms we adopt the following principle:

(VII)
The interlingual representations of verbal tense forms are pairs consisting of one deictic and one aspectual value.

As the number of possible combinations of deictic and aspectual values is 18 (3x6), it follows that each tense form can have at most 18 different interlingual representations. In order to determine which values a given tense can actually have one has to examine its compatibility with the different types of time adverbials.

As for the deictic subpart, it is not so difficult to invent a criterion:

(i)
If tense X is compatible with a deictic Y-adverbial, where Y ∈ {after, before, contain}, then the tense X can have the value Y.

For the aspectual subpart the criteria are a bit more complicated:

(ii)
If tense X can be used in a sentence with a when-adverbial in which the event is said to take place before or after the interval denoted by that when-adverbial. then the aspectual value of X can be either 'before' or 'after', i.e. X can be used to express either retrospectivity or prospectivity.

(iii)
If tense X can be used in a sentence which contains both a when-adverbial and a duration adverbial that denotes an interval that is larger than the interval denoted by the when-adverbial, then tense X can be used to express the durative aspect.

Similar criteria have to be stated for the other aspects (inchoative, terminative, and perfective). As far as we can see now the perfective aspect might well be considered to be the default value: from a conceptual point of view the least marked situation is the one in which the event time is contained in or identical with the reference time (E ⊂ R or E = R).

As an illustration of how these criteria can be used in practice we give an interlingual analysis of the Dutch 'Voltooid Tegenwoordige Tijd (VTT)'. This tense is expressed by the combination of an auxiliary ('hebben' or 'zijn') and the perfect participle of a lexical verb.

The VTT can be combined with all kinds of when-adverbials:

(13) nu heb ik het gevonden
now-have-I-it-found

(14) morgen heb ik het gevonden
tomorrow-have-I-it-found

(15) gisteren heb ik het gevonden
yesterday-have-I-it-found

In (13) and (14) the time of event precedes the time denoted by resp. "nu" and "morgen"; hence, the aspectual value of the VTT in these sentences is the retrospective one. In fact, (13) and (14) belong to a paradigm of retrospective tenses. The other members of the paradigm are the "Voltooid Verleden Tijd" and the "Voltooid Toekomende Tijd", as in

(16) gisteren had ik het al gevonden
yesterday-had-I-it-already-found

(17) morgen zal ik het gevonden hebben
tomorrow-shall-I-it-found-have

(14) and (17) even have the same meaning and, hence, the same interlingual representation, namely the combination after - before.
(13) has the value contain - before, and
(16) the value before - before.

In (15) the situation is different: here, the time of finding does not precede the interval denoted by "yesterday" (as in (16)), but is rather contained in it. The aspectual value of the VTT in (16) is hence the perfective one, and the interlingual representation in that case is before - during.

It can further be shown that the VTT cannot be used to express a durative aspect. Compare

(18) gisteren ben ik de hele dag ziek geweest
yesterday-am-I-the-whole-day-ill-been

(19) * gisteren ben ik drie dagen ziek geweest
yesterday-am-I-three-days-ill-been

In (18) the event time denoted by the duration adverbial "de hele dag" is a subset of the interval denoted by "gisteren" (= perfective aspect); in (19), on the other hand, the event time (three days) is said to be longer than the reference time (one day). Since this

combination leads to ungrammaticality (in Dutch), it follows that the VTT cannot express durativity.

If these analyses are correct, it follows that the Dutch VTT can have three distinct interlingual representations: contain - before, after - before, and before - during.

The general idea now is that this information is contained in the lexicon, and that for the assignment of temporal representations to particular sentences one first looks in the lexicon to see which interlingual representations the tense used in that particular sentence can have, and then singles out that subset of representations which is compatible with the time adverbials used in the sentence.

If that subset contains exactly one member the sentence may be said to be unambiguous with respect to the temporal calculus; if the subset contains more members, the sentence is said to be temporally ambiguous; and if the subset is empty, the sentence is simply not well-formed.

As a conclusion to this section we give the representations of some of the discussed sentences

(13) $\exists$ S,R,E $\subseteq$ time (contain(R,S) & nu(R) & before(E,R) & AT (E, ik het vinden))

(15) $\exists$ S,R,E $\subseteq$ time (before(R,S) & gisteren(R) & during(E,R) & AT(E, ik het vinden))

(18) $\exists$ S,R,E $\subseteq$ time (before(R,S) & gisteren(R) & during(E,R) & de hele dag(E) & AT(E, ik ziek zijn))

4. Prospects.

In this paper we have concentrated on the definition of a conceptual calculus for the representation of time meanings in natural language. We have also given principles (IV,V,VI,VII) and criteria (i,ii,iii) for relating the concepts of the calculus to language specific morphosyntactic categories. Given these tools, it should be possible to analyse the tenses of the different languages in such a way that the results of the analysis are comparable and, indeed, identical iff they express the same concept.

It goes without saying that the actual analysis of all possible tenses cannot be carried out in a paper of this size, but we have the feeling that we have at least cleared the ground for such an enterprise.

References

Bruce, Bertram (1972) 'A model for temporal references and its application in a question answering program', in Artificial Intelligence 3, 1-25.

Comrie, Bernard (1976) Aspect: an introduction to the study of verbal aspect and related problems, Cambridge University Press, Cambridge.

Dowty, David (1979) Word meaning and Montague grammar, Reidel, Dordrecht.

van Eynde, Frank (forthcoming) Meaning and translatability, doctoral dissertation, Leuven.

Friedrich, Paul (1974) 'On aspect theory and Homeric aspect', in International Journal of American Linguistics 40, memoir 28.

Johnson, Marion (1981) 'A unified temporal theory of tense and aspect', in Tedeschi & Zaenen (eds.), Syntax and semantics. Volume 14. Tense and Aspect, Academic Press, New York.

Kripke, Saul (1972) Naming and necessity, Harvard University Press, Cambridge Mass.

Reichenbach, Hans (1947) Elements of symbolic logic, University of California Press, Berkeley.