A COMPOSITIONAL APPROACH TO THE TRANSLATION OF
TEMPORAL EXPRESSIONS IN THE ROSETTA SYSTEM

Lisette Appelo
Philips Research Laboratories
Eindhoven, The Netherlands

Abstract

This paper discusses the translation of temporal expressions, in the framework of the machine translation system Rosetta. The translation method of Rosetta, the "isomorphic grammar method", is based on Montague's Compositional Principle. It is shown that a compositional approach leads to a transparent account of the complex aspects of time in natural language and can be used for the translation of temporal expressions.

0. Introduction

It is a well-known fact that the translation of temporal expressions in natural languages is not a simple mapping of verbal tenses. In (1) the Dutch Simple Present Tense is used while in (2) a Present Perfect Progressive Tense is the most appropriate tense to represent the time meaning of (1) in English. A more literal translation of the Dutch Simple Present is sometimes considered to be ill-formed, as illustrated by (3). But in other cases translation yields a perfect result, as shown by the pair (4)/(5).

(1) Jan woont hier al 20 jaar
(Dutch: John-lives-here-already-20-years)
(2) John has been living here for 20 years
(3) * John is living here for 20 years
(4) Jan woont hier (John-lives-here)
(5) John is living here

In this paper I will deal with this kind of problem from the perspective of machine translation. More specifically, I will sketch a solution within the framework of the Rosetta machine translation project. In this project translation systems are developed between Dutch, English and Spanish, using the "isomorphic grammar" method. According to this method, the grammars of the languages are attuned to each other in such a way that a sentence s in a translation equivalent of a sentence s' has the same meaning and syntactic and morphological components together define sentences.

I will ignore this component in this paper.

3) The semantic component

M-grammars obey the Compositionality Principle, which states that the meaning of an expression is a function of the meaning of the parts of that expression. The basic S-trees and all other S-trees can be given a model theoretical interpretation: the basic expressions correspond to semantic values in a semantic domain and the rules to semantic operations.

The process of making an expression can be represented by a syntactic derivation tree (D-tree) with the (names of the) basic expressions at the terminal nodes and the names of the rules that are applicable at the nonterminal nodes.

2) The morphological component

The morphological component relates lexical S-trees (the leaves of the surface trees) to strings. So the syntactic and morphological components together define sentences.

In section 1 I will give a short exposition of the isomorphic grammar method. Section 2 presents a theory of time in the Rosetta framework. In section 3 I will sketch isomorphic grammars for temporal expressions and illustrate them by some examples. Possible extensions will be discussed in section 4.

1. The Isomorphic Grammar Method

1.2. M-Grammars

M-grammars can be seen as a computationally viable and syntactically powerful variant of Montague Grammar. An M-grammar consists of three components: 1) a syntactic component, 2) a morphological component and 3) a semantic component.

1) The syntactic component

The syntactic component defines a set of S-trees (surface trees) whose leaves correspond to words, in surface order. An S-tree is an ordered tree with nodes which are labelled with syntactic categories and attribute-value pairs. The branches are labelled with syntactic relations, e.g. subject, object, etc.

In the rest of this paper I will abbreviate them by specifying the top node and a characterization of the rest of the tree, as:

CAT(attribute: value,...) (string)

The syntactic component defines S-trees by specifying:

(i) a set of basic S-trees (also called basic expressions)
(ii) a set of syntactic rules:
3) The semantic component

M-grammars obey the Compositionality Principle, which states that the meaning of an expression is a function of the meaning of the parts of that expression. The basic S-trees and all other S-trees can be given a model theoretical interpretation: the basic expressions correspond to semantic values in a semantic domain and the rules to semantic operations.

The process of making an expression can be represented by a syntactic derivation tree (D-tree) with the (names of the) basic expressions at the terminal nodes and the names of the rules that are applicable at the nonterminal nodes. (cf. figure 1)

M-grammars must satisfy certain conditions to allow for effective analysis next to generation. For more details the reader is referred to Landebergen (1982, 1984).

1.2. Translating with isomorphic M-grammars

The translation relation between two (or more) languages is defined by attuning their grammars as follows:

(i) For each basic expression of a grammar there is at least one corresponding basic expression of the other grammar with the same meaning.
For each syntactic rule of a grammar there is at least one corresponding syntactic rule of the other grammar, with the same meaning.

The correspondence between rules is only required for "meaningful" rules. Syntactic transformations can be added for each language separately and do not occur in the derivation trees.

In terms of derivation trees: for each syntactic derivation tree of a grammar there is at least one syntactic derivation tree of the other grammar with the same geometry and labelled with corresponding basic expressions and syntactic rules. These syntactic derivation trees correspond to the same semantic derivation tree and are called isomorphic derivation trees.

Two sentences are defined as translations of each other if they have the same semantic derivation tree and therefore corresponding derivation trees. (Note that the definition is given for isolated sentences on the basis of linguistic knowledge only: in fact a "possible translation" relation is defined, possible in some context.)

Grammars that are attuned in this way are called isomorphic M-grammars if the corresponding sets of rules satisfy certain applicability conditions, such that for each well-formed syntactic derivation tree in a grammar there is at least one well-formed syntactic derivation tree in other grammars. (A derivation tree is well-formed if it defines a sentence, i.e. if the rules are applicable.)

This time model $T$ is a partially ordered set of which the elements are called time points. The ordering relation is "<", meaning "earlier than". For this model we define a notion interval. Intervals are subsets of $T$ without any gaps or branches.

One of those points is called $S$, the moment of speech or narration.

The "objects" to be located in $T$ will be called events. An event is something that can be located in time, dependent on the temporal ingredients of the sentence that refers to the event. For example in order to establish the truth value of John has been living here for 20 years, we need to locate the event "John is living here" in $T$.

An event is assumed to correspond to an interval $E$ of $T$. We will say that the event is located in $T$ when the relation between $E$ and $S$ can be established. This relation can be complex in the sense that more intervals than $E$ and $S$ can be involved. Such intervals are called reference intervals.

Intervals can be characterized by properties indicating e.g. the "length" or "duration" of an interval, a particular relation to $S$ or a part of the calendar. These properties are expressed by adverbials or special (auxiliary) verbs.

The event which corresponds to the interval $E$ has temporal properties which are often called "Aktionsart" in the literature. Both the main verb and its arguments of the clause which expresses the event may play a role in the determination of the Aktionsart (cf. Verkuyl (1972)).

Usually four types of Aktionsart are distinguished (cf. e.g. De Vuyt (1983)) as is illustrated in the following examples:

(6) This book belongs to me (stative)
(7) John is working (activity)
(8) John wrote a letter (accomplishment)
(9) He reached the end of the street (achievement)

The event in (6) is durative; it can take place at an
arbitrarily long interval. In (7) the event is
durative, but it cannot be claimed to take place only
at a time point (i.e. a minimal interval) in the
model; it is dynamic which implies some progress or
change. The event in (8) is terminative, because the
result or end is indicated; it can be looked at from
the "outside" as a unit, but it cannot be claimed to
take place at a time point in the model; it is
dynamic when looked at from the "inside" (activity). In
(9) the event is terminative, because the end or
result is indicated but it is also claimed to take
place at a time point in the model and therefore also
called momentary.

There are two important types of relations between
Intervals:
1) the deictic relation: relation between a reference
interval and E
2) the aspectual relation: relation between E and a
reference interval

These relations are expressed by morphological and
periphrastic tense.

2.3. Time in Rosetta

For defining the translation of temporal expressions
in the Rosetta framework, we have to write isomorphic
compositional grammars for them, which boils down to:
1) specifying for each language:
(a) temporal expressions (time adverbials etc.)
expressing properties of time intervals,
(b) syntactic rules (e.g. tense rules) that indicate
how temporal expressions must and can be combined,
expressing relations between these intervals,
2) attuning these expressions and rules of the
languages involved to each other in the way described
in section 1.

Before specifying these grammars in section 3 I will
briefly discuss the motives for the strategy followed
in these grammars and for the particular choice of
reference intervals.

1. Tenses and adverbials cannot be translated inde-
dependently:
- Sometimes the translation of a tense is only
correct if it occurs with a certain time adverbial.
Consider for example the pair (10)/(11):
(10) Jan werkt hier al 3 jaar (Pres. Tense)
(11) John has been working here for 3 years
(Pres. Perf. Prog. Tense)

The corresponding tense of the Dutch Present Tense
in the context of the time adverbial al 3 jaar in (10),
is in English a Present Perfect Progressive Tense as
in (11). But without that time adverbial the trans-
lation is different as is shown in the examples
(12) - (14):
(12) Jan werkt (Pres. Tense)
(13) John is working (Pres. Prog. Tense)
(14) John has been working (Pres. Perf. Prog. Tense)

- Adverbials are not always translated into adverb-
ials. Consider for example the pair (15)/(16):
(15) English: He has just arrived.
(16) Spanish: Él acaba de llegar.

In (15) the adverb just expresses the "near past",
but in (16) a special verb acabar de which could be
considered as a part of some periphrastic tense is
used. These examples are an indication that the grammars
for temporal expressions should take into account
tenses and time adverbials together.

2. Van Eynde et al. (1985) give a specification of
time meaning representation for machine translation,
based on a time model with three intervals, E, R and
S, which results in time meaning representations that
do take into account time adverbials.

In their time model, however, the set of aspectual
relations, relations between E and R, contains next
to the retrospective relation (E before R) the
imperfective relation (E contains R). This seems
inadequate in view of sentences that can have both an
imperfective and a retrospective aspectual relation
as in e.g. (10) and (11).

The complexity of the aspectual relation is also
recognized in Maegaard (1982) where a special value
CONTINUOUS for the attribute RETROSPECTIVE was intro-
duced for the translation of verbal tenses.
Krauwer and Des Tombe (1985) make similar observa-
tions.

I propose therefore that in the Rosetta framework
1) a time meaning representation obligatorily will
contain an aspectual relation, i.e. a relation be-
tween E and a reference interval R, which will be
called perfective if R is a subset of E and
imperfective if E is a subset of R, and
2) that it can optionally contain a retrospective
relation between R and a time point S, which is
some "local point of evaluation", meaning that S
lasts until R.

This R is an arbitrary point of some reference
interval. Consider for example:
(17) Yesterday John had been living there for 3 years
(18) The 3 years (R) last until some point during
yesterday (R).

If there is no retrospective relation, R will be an
arbitrary point of R.

Between R and S the deictic relation is specified:
Past (R is before S), Present (R is simultaneous
with S) or Future (R is after S).

So a time meaning representation of an event in
Rosetta will consist of:
- properties of E and R
- an aspectual relation between E and R
- optionally a retrospective relation between R and
- a deictic relation between R and S

In the next section grammars for temporal expressions
will be discussed which start with a clause and apply
rules that will first specify the properties for E,
then the aspectual relation and the properties for
R, then optionally the retrospective relation and
finally the deictic relation.

3. Isomorphic grammars for temporal expressions

3.1. Corresponding rules for temporal expressions

To achieve isomorphy of grammars for temporal
expressions, corresponding rules for the languages of
the system have to be written as was explained in
section 1. These rules are applied to a clause which
consists of a verb, its arguments and an attribute
Aktionsart, the value of which has been specified
during the composition of the clause. The result of
the application of the rules is a clause with
specified tense forms, auxiliaries and adverbials.
Semantically, properties of and relations between
time intervals and the event are specified. The rules have one argument, a clause, or two arguments, a clause and an adverbial or an auxiliary verb that we wish to introduce categorically.

I will distinguish five classes of rules. The rules are either obligatory (OB), meaning that exactly one of this class of rules is applied, or optional (OP).

I. Aktionsart rules (OP): the application of these rules results in a new clause with a different Aktionsart value caused by some (auxiliary) verb or adverbial that is inserted into the clause.

II. duration rules (OP): These rules insert a duration adverbial into the clause.

III. aspect rules (OB): These rules insert a reference adverbial and specify the aspectual tense forms (perfective or imperfective) of the verbs.

IV. retrospective rules (OP): These rules are applied to a clause that contains some retrospective reference adverbial. It inserts another, non-retrospective reference adverbial and adds, if necessary, auxiliary verbs and/or adverbials.

V. deictic rules (OB): These rules determine the deictic tense form of the verbs in the clause.

In 3.3. the rules will be discussed in more detail.

### 3.2. The S-trees

A clause is represented as an S-tree with a top node CL that has the following temporal attributes and corresponding value sets in all languages: Aktionsart: {stative, activity, accomplishment, achievement} aspect: {imperfective, perfective, unmarked} deixis: {present, past, future, unmarked} retrospectivity: {-retro, +retro} (the underlined value is the initial value).

A clause contains a VERB node with attribute-value pairs concerning the verb form, which may differ over languages.

A clause represents an event with time interval E. The other temporal expressions may be of various categories: e.g. ADVP, PP, NP or CONJP (a time conjunction and a clause).

They are marked at the topnode for temporal properties by the following attributes: class: {duration, reference} deixis: {present, past, future, unmarked} aspect: {perfective, imperfective} retrospectivity: {+retro, -retro}

Adverbs of the duration class will always have the value unmarked for their deixis and -retro for their retrospectivity attribute. They indicate a property of the interval E. Perfective duration adverbials specify the duration of the event, imperfective the duration of an interval during which the event takes place. For example:

PP(class: duration, aspect: imperfective, deixis: unmarked, retrospectivity: -retro)

(in three hours)

Reference adverbials indicate properties of intervals. If their deixis attribute has the value unmarked, they are called absolute, indicating that there are no restrictions on the relations of the interval with S, otherwise they are called deictic, indicating that the interval has a certain relation with S. If their retrospectivity attribute has the value +retro, they are called retrospective, indicating that the interval has the relation until to the reference point R. For example:

ADVP(class: reference, aspect: imperfective, deixis: past, retrospectivity: -retro)

(yesterday)

### Absence of adverbials

Clauses do not always contain explicit adverbials:

a) In case of isolated clauses without reference adverbials we will assume an abstract deictic reference adverb which indicates the moment of speech S:

REF(class: reference; deixis: present; aspect: perfective; retrospectivity: -retro)

b) In case a reference interval is indicated that has the property that it ranges over the whole time axis until some reference point R, we assume an abstract adverbial:

PAST(class: reference; deixis: unmarked; aspect: Imperfective; retrospectivity: +retro)

### 3.3. The rules

I will now give an informal description of each type of rule containing an account of syntactic and semantic aspects and the differences between Dutch, English and Spanish.

I. Aktionsart rules (OP). They change the Aktionsart and insert (complex) auxiliary verbs or adverbs, and/or determine the form of the verb. Semantically, a new event is derived from the original event. For example, an accomplishment event can be transformed into a non-terminative event. Compare:

Eng: (18) CL (John write a letter) → CL (John is writing a letter)

Du: (19) CL (Jan een brief schrijven) → CL (Jan is een brief aan het schrijven zijn)

This is sometimes called the "locative tense"; the rules insert:

in Dutch: aan het VERB{form: Infinitief}zijn
in Spanish: estar VERB{form: gerundio}
in English: be VERB{form: ingform}

They change the Aktionsart value to stative (due to the auxiliary).

II. duration rules (OP). These rules are applied to an S-tree with a certain Aktionsart value and a duration adverbial with a certain aspect value that is inserted in the clause. The aspect value of the clause will now be perfective.

This rule applies the property denoted by the adverbial, to the interval E.

For example:

(20) CL (John write for three hours)

III. aspect rules (OB). Rules with two arguments: 1) a clause with a certain Aktionsart and aspect and 2) a (possibly anaphoric) reference adverbial with a certain aspect. They determine the imperfective and perfective verb forms, sometimes with auxiliaries. The aspectual verb forms can differ over languages. English seems to have perfective simple tense forms; the imperfective forms are composed with the auxili-
ary be. Spanish has clear imperfective and perfective past tense forms. In Dutch the verb form does not seem to distinguish between imperfective and perfective.

The aspect value of the clause unmarked has to be changed into perfective or imperfective. The reference adverbial is inserted into the S-tree.

Semantically, the property denoted by the adverbial is applied to the interval \( R_S \) and the relation between the intervals \( R_E \) and \( R_S \) is expressed.

Example:

(21) CL (Ayer Juan leyó un libro.) (perfective)  
(Yesterday-John-read-a-book)  
(22) CL (Ayer Juan leía un libro.) (imperfective)  
(Yesterday-John-was-reading-a-book)

IV. Retrospectivity Rules (OP). Rules that have two arguments: 1) a clause with a certain aspect and a reference adverbial that has the retrospectivity value \(-\text{retro}\) (this may be \( \text{PAST} \)), and 2) a (possibly anaphoric) reference adverbial that has the retrospectivity value \(+\text{retro}\). They insert auxiliary verbs such as have (hebben/zijn), in Dutch, haber (Spanish), if necessary. In English have must always be inserted. In Dutch and Spanish it is obligatory if the aspect value is perfective. So these languages have two rules: one for perfective and one for imperfective clauses. The reference adverbial is inserted in the S-tree.

The rules that insert have just (English) zojuist hebben/zijn (Dutch) and acabar de (Spanish), under the condition that the clause contains the abstract adverbial PAST and the aspect value perfective, to express near-retrospectivity, belong to this class.

Semantically, these rules apply the property denoted by the adverbial to \( R_S \) and express the relation until or "near-until" of \( R_E \) (indicated by the retrospective adverbial) to \( R_S \).

Example:

(23) CL (REF John have just read this book.)  
(24) CL (REF Juan acabar de leer este libro.)  
(25) CL (REF Jan dit boek zojuist gelezen hebben)

V. Deictic Rules (OB). These rules are applied to a clause with a reference adverbial (that can be anaphoric), inserted by rules of type III or IV. They determine present, past and future forms of the verbs, sometimes with insertion of an auxiliary verb form, for example will for future in English, after checking if the deictic value of the reference adverbial is compatible.

Semantically, the relation between \( R_E \) (a subset of the interval indicated by the reference adverbial in the clause) and \( S \) is expressed. The deictic relations are: present (\( R_E \) is simultaneous with \( S \)), past (\( R_E \) is before \( S \)) and future (\( R_E \) is after \( S \)). In these rules the abstract adverbials will be deleted.

Example:

(26) John read a book yesterday  
(27) *John has read a book yesterday  
(28) John had read a book yesterday

3.4 Examples

I will now give some examples of parallel derivations of sentences with temporal expressions that are translation equivalents. I will leave out irrelevant specification of nodes.

Example: (29) John has been writing for 2 hours

**Example application of syntactic rules for English**

\[ R_1: \text{(imperfective aspect rule)} \]
\[ \text{CL} \{\text{Aktionsart: activity, aspect: unmarked, deixis: unmarked, retrospectivity: \(-\text{retro}\)} \}
\[ \text{(John write)} \]
\[ + \text{PP} \{ \text{deixis: unmarked, aspect: perfective, retrospectivity: \(+\text{retro}\), class: reference} \}
\[ \text{(for 2 hours)} \]
\[ \rightarrow \text{CL} \{ \ldots, \text{aspect: perfective}, \ldots \} \]
\[ \text{(John be writing for 2 hours)} \]

\[ R_2: \text{(retrospective rule)} \]
\[ \text{CL} \{ \ldots, \text{retrospective: \(-\text{retro}\)} \}
\[ \text{+ REF} \{ \text{deixis: present, aspect: perfective, retrospectivity: \(+\text{retro}\), class: reference}\} \]
\[ \text{(John have been writing for 2 hours)} \]

\[ R_3: \text{(present deictic rule for finite clause)} \]
\[ \text{CL} \{ \text{deixis: unmarked, \ldots} \}
\[ \text{(John have been writing for 2 hours)} \]

**Example: (29) John has been writing for 2 hours**

**Example: (29) John has been writing for 2 hours**

**Example: (29) John has been writing for 2 hours**

**Example: (29) John has been writing for 2 hours**

**Example: (29) John has been writing for 2 hours**

**Example: (29) John has been writing for 2 hours**
(31) Jan is al 2 uur aan het schrijven
(Johk has already 2 hours on the writing)
The isomorphic D-trees for (29) and (31) are shown in figure 3.

But for 2 hours is ambiguous too and can also be a
perfective duration adverbial corresponding to the
Dutch 2 uur. The duration rules (R_4 and R'_5) insert
them. The clause will now have perfective aspect and
the retrospective adverbial PAST will be introduced by
the perfective aspect rule (R_4 and R'_5). In Dutch
the retrospective rule R'_6 (different from R'_5 but
also corresponding to the English R_6) is now applied:
the one for a clause with perfective aspect that
inserts the auxiliary hebben/zijn. The resulting
Dutch translation equivalent of (29) is:

(32) Jan is 2 uur aan het schrijven geweest
(Dutch: John has 2 hours on the writing been)
The isomorphic derivation trees are shown in figure 4.

The conditions in the rules filter out certain
unacceptable combinations, e.g.:

(33) *John died for a while
(34) *John was working in three hours
(33) is ruled out because for a while is a perfective
duration adverbial that cannot combine with achieve-
ments or accomplishments and (34) because in three
hours is an imperfective duration adverbial that
cannot combine with activities and statives.

4. Concluding remarks
It is not possible to treat all temporal expressions
and all translation problems with respect to time in
this paper, but I have sketched a solution as to how
to treat them in the Rosetta framework. I expect that
other "aspectual forms" such as inchoative, termina-
tive etc. can be added at the "Aktionsart level" or
"aspect level" and that the current approach, which
is: [...]

References
Bruce, B. (1972): 'A model for temporal references
and its application in a question answering program'.
In: Artificial Intelligence 3, 1 -25.
Comrie, B. (1976): Aspect: an Introduction to the
Study of Verbal Aspect and Related Problems. Cam-
bridge University Press, Cambridge.
Dowty, D.R. (1979): Word Meaning and Montague
Grammar. Reidel Vol. 7a--. Reidel.
Kneur, S. and L. des Tombe (1985): 'Composition-
ality'. In: De Haan, G. and W. Zonneveld (eds.):
Formal Parameters of Generative Grammar. I. Yearbook
1985. Nordrecht.
Landsbergen, J. (1982): 'Machine Translation based
on Logically Isomorphic Montague Grammars'. In: J.
Horecky (ed.), COLING 82 North Holland, p. 175-182.
Landsbergen, J. (1984): 'Isomorphic Grammars and
their use in the Rosetta translation system'. Paper
presented at the Tutorial on Machine Translation.
Lugano. Philips Research M.S. 12.950, to appear in:
Kings, M. (ed.), Machine Translation: The State of the
Art, Edinburgh University Press.
Manggaard, B., (1982): 'The Transfer of Finite Forms
in a Machine Translation System'. In: Abstracts
COLING 82, p. 190-194.
Oversteegen, L. and H.J. Verkuyl, (1985): 'De
temporele zinsstructuur van het Nederlands: twee
tijdsbanden'. In CLOT.
Reichenbach, H., (1947): Elements of Symbolic Logic.
University of California Press, Berkeley.
Van Eynde, P., L. des Tombe and F. Maes (1985):
'Specification of time meaning for machine transla-
tion'. Proceedings ACL conference, Geneva.
Verkuyl, H.J. (1972): On the Compositional Nature of
the Aspects. Nordrecht.
Vuyst, J. de, (1983): "A Semantic Analysis of Tem-
poral Elements in Dutch and English". Diss. RUU,
Groningen.