THE ANALYSIS OF TENSE AND ASPECT IN EUROTRA

Frank VAN EYNDE

Eurotra - Leuven
Maria-Theresiastraat 21
3000 Leuven
BELGIUM

ABSTRACT

This paper presents a framework for the model-theoretic analysis of tense and aspect forms in discourse. It has been developed for Eurotra, the MT project of the European Community, and has been applied to the nine Eurotra languages: English, German, Dutch, Danish, Greek, Italian, French, Spanish and Portuguese.

The paper consists of six parts. The first presents the problem of translating tense and aspect forms and indicates the type of solution I envisage. The second contains a formalism for the representation of time meanings. The third and the fourth present a theory of tense and aspect respectively. The fifth discusses the issue of compositionality and the sixth is about the use of the system in the Eurotra framework.

1. THE PROBLEM

It is a fact of language that the number and the use of the tense and aspect forms are different for every language. Even for closely related languages the differences tend to be large. As a consequence, it is not possible to state one-to-one correspondences between the tense and aspect forms of different languages. Some examples:

EN he has lived in London for 20 years
(present perfect)
FR il vit à Londres depuis 20 ans
(simple present)

EN he has been watching TV for hours
(present perfect progressive)
FR il a regardé la télé pendant des heures
(present perfect)

Differences like these pose non-trivial problems for machine translation. In general there are two ways in which they can be handled : either by defining complex mappings from source language forms to target language forms in transfer

SL form   ------------------->   TL form
              complex
              mappings

or by defining mappings between language specific forms and interlingual meanings in the monolingual components

meaning   ------------------->   meaning
   ↑   identity |
mapping | | mapping
   | ↓
SL form   TL form

Because of Eurotra's policy to keep the bilingual transfer components as small and simple as possible it has been decided to pursue the interlingual approach.

The resulting system is based on insights from

- interval semantics (cf. *Bennett, Partee, Dowty, Bruce*)

- the Reichenbachian analysis of tense and aspect in terms of time of speech, time of reference and time of event (cf. *Reichenbach, Johnson, Smith*)

- discourse representation theory (cf. *Kamp, Rohrer, Partee*)

- descriptive typological studies (cf. *Comrie*)

2. THE FORMALISM

As a starting point I take the temporal structure $\langle T,<,\cap \rangle$, where T is a set of intervals, < is a binary relation that linearly orders time (precedence), and $\cap$ is a binary operation on intervals (intersection).

An interval is a continuous subpart of the time line (a). It may consist of one single moment of time (b), but it cannot contain any gaps (c):
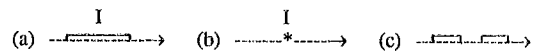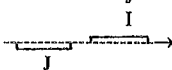
The intersection of two intervals is that subpart of the intervals which they have in common:
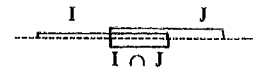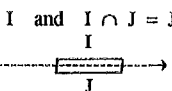
Given the temporal structure $\langle T,<,\cap \rangle$, the number of possible relations between intervals can be determined in a principled way: for any ordered pair of intervals (I and J), it will be the case that

either $I \cap J = \emptyset$

and then   <(I,J)   (preceed)

or   >(I,J)   (follow)

or $I \cap J \neq \emptyset$

and then $I \cap J = I$ and $I \cap J = J$

=(I,J)   (identity)

or $I \cap J = I$ and $I \cap J \neq J$

⊂(I,J)   (part-of)

or $I \cap J \neq I$ and $I \cap J = J$

⊃(I,J)   (inclusion)

or $I \cap J \neq I$ and $I \cap J \neq J$

$$I$$
$$<<(I,J) \quad \text{---} \text{[} \text{---}\text{---}\text{]} \text{---}\rightarrow \quad \text{(left}$$
$$J \qquad\qquad \text{overlap)}$$

$$I$$
$$>>(I,J) \quad \text{---}\text{[}\text{---}\text{]}\text{---}\rightarrow \quad \text{(right}$$
$$J \qquad\qquad \text{overlap)}$$

These are the seven logically possible relations between ordered pairs of intervals on a one-dimensional time line.

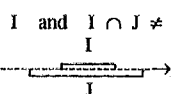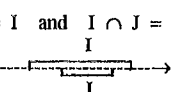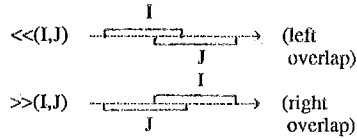For the analysis of single isolated clauses I will use the Reichenbachian notions of time of speech, time of reference and time of event. The time of event (E) is the interval for which a basic tenseless proposition is said to be true, and the function of the tense and aspect forms is to define the relation between that interval and the time of speech (S) via the intermediary time of reference (R).

For the analysis of clauses in context I make use of a generalised model (cf. 3.4.).

3. A THEORY OF TENSE

3.1. the tense meanings

Tense meanings will be defined as relations between a time of reference and a time of speech : Rel(R,S).

The number of possible tense meanings is, hence, equal to the number of possible relations between R and S, which is seven (cf. 2.). However, since the time of speech is generally conceived to be a moment of time rather than an interval of some length, some of these relations cannot hold in principle. The overlap-relations (<< and >>), for instance, can only obtain between two intervals of a certain length, and the proper part-of relation ($\subset$) cannot hold between R and S either, for if S is one moment of time, R can only be a proper part of S if it is smaller than a moment, which is impossible.

Furthermore, there seems to be no linguistic evidence for making a distinction between proper inclusion ($\supset$) and identity (=), since "... languages do not have distinct grammatical categories of tense indicating location in time at a particular point vs. location in time surrounding a particular point." [*Comrie* 1985, 123]

As a consequence, the number of possible relations between R and S can be reduced to the following three:

$$<(R,S) \quad = \text{ anteriority}$$
$$>(R,S) \quad = \text{ posteriority}$$
$$\supseteq(R,S) \quad = \text{ simultaneity}$$

These correspond to the traditional concepts of Past, Future and Present. Notice, however, that the latter is not defined in terms of identity, but in terms of improper inclusion.

The language specific forms for the expression of these concepts are the tense forms and the time adverbials.

3.2. the deictic time adverbials

Typical examples of deictic time adverbials are "now", "tomorrow", and "two weeks ago". Their function is to relate the time of reference to the time of speech. Depending on the kind of relation they express they can be characterised as

simultaneous : now ...
anterior : yesterday, two weeks ago ...
posterior : tomorrow, next summer ...

3.3. the tense forms

In contrast to the tense meanings which are language independent the tense forms are language specific. Their number, names and distribution differ from language to language.

As for the Eurotra languages there seem to be two types of tense form systems: the one of the Romance languages and the one of the Germanic languages and Greek.

An example of the latter type is English:

tense form $\rightarrow$ [+/-ED] (will+infinitive)

yielding

[-ED] $\varnothing$	= Present	play
[+ED] $\varnothing$	= Past	played
[-ED] will	= Future	will play
[+ED] will	= Conditional	would play

There is a bound morpheme [+/-ED] and an optional auxiliary "will". The latter can also have a modal meaning, especially in its past tense form, but in this context I will only discuss its temporal meaning.

An example of the former type is French:

$$\text{tense form} \rightarrow \begin{bmatrix} \text{[+/-R]} & \text{[+/-AIS]} \\ \text{Passé} \end{bmatrix}$$

yielding

[-R] [-AIS]	= Présent	joue
[+R] [-AIS]	= Futur	jouerai
[-R] [+AIS]	= Imparfait	jouais
[+R] [+AIS]	= Conditionnel	jouerais
Passé	= Passé	jouai

In this system the tense forms are combinations of bound morphemes; there are no auxiliaries involved.

As for the meanings of the tense forms they will be defined as elements of the power set of possible tense meanings. This power set contains eight elements :

$$\{\varnothing, \{<\}, \{>\}, \{\supseteq\}, \{<,>\}, \{<,\supseteq\}, \{>,\supseteq\}, \{<,>,\supseteq\}\}.$$

Not all of these combinations can be assigned to particular tense forms, though, for there are a few general constraints.

Bernard Comrie has argued, for instance, that "in a tense system, the time reference of each tense is a continuity" [*Comrie* 1985, 50]. This implies that there can be no tense forms which can express posteriority and anteriority without expressing simultaneity as well. The combination {ante,post} can, hence, be discarded a priori. For the Eurotra languages this restricition appears to hold.

A second restricition concerns the combinations {ante,simul} and {post,simul}. The former is a possible meaning in languages which make a basic distinction between Future ({post}) and non-Future ({ante,simul}) ; the latter is a possible meaning in languages which make a basic distinction between Past ({ante}) and non-Past ({post,simul}). Since a language cannot belong to both types at the same time, it follows that for any given language either the combination {post,simul} or the combination {ante,simul} is ruled out. As far as the Eurotra languages are concerned, they all belong to the latter type.

In order to find out which of the six remaining combinations can be assigned to the tense forms one can make use of a grammaticality test : a tense form X can have a meaning Y (where Y is any of {simultaneous, anterior, posterior}), if and only if it can be combined with a deictic adverbial of type Y.

The application of this test to English and French yields the following results :

English :	Present	→	{post,simul}
	Past	→	{ante}
	Future	→	{post}
	Conditional	→	∅
French :	Présent	→	{post,simul}
	Futur	→	{post}
	Imparfait	→	{ante}
	Conditionnel	→	∅
	Passé	→	{ante}

The conditional tenses get the value ∅ since they do not have a temporal meaning in single isolated clauses (cf. 3.6.).
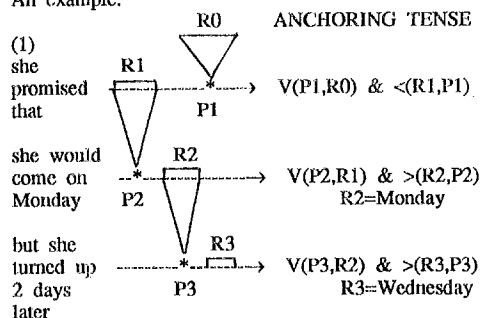
3.4. a discourse model

The model presented so far is useful for the analysis of isolated clauses. For the analysis of texts we need an extension, or rather a generalisation of the original model.

The main extensions concern the introduction of another kind of interval, the point of perspective P (the term is borrowed from [Rohrer 1985]), and the addition of indices to the intervals.

Instead of defining the time of reference with respect to the time of speech I will now define its position with respect to the point of perspective. For any clause i which is part of a discourse, there will be one point of perspective P_i and one time of reference R_i. If the clause is the first main clause of the discourse, then its point of perspective is derived from the time of speech. In other cases the point of perspective will be derived from the time of reference of a dominating or preceding clause.

An example:

(1)
she
promised
that
she would
come on
Monday
but she
turned up
2 days
later

RO   ANCHORING TENSE

V(P1,R0) & <(R1,P1)

V(P2,R1) & >(R2,P2)
R2=Monday

V(P3,R2) & >(R3,P3)
R3=Wednesday

The notation "V(Pi,Rj)" means that Pi is derived from Rj. The interval Rj from which the position of Pi is derived will be called the temporal antecedent of the clause with point of perspective Pi. In the example the temporal antecedent of the third clause is "Monday" (R2), the temporal antecedent of the second clause is the time of her promising (R1), and the temporal antecedent of the first clause is the time of speech (R0=S).

Notice that the temporal antecedent of a clause i need not always be the time of reference of the immediately preceding clause (Ri-1). In the sequence

(2) she promised that she would come on Monday, but then she changed her mind

the temporal antecedent of the third clause is the time of her promising (R1) rather than "Monday", and in

(3) she promised that she would come on Monday, but now it seems that she cannot

the temporal antecedent of the third clause is the time of speech : "now" refers back to the time of speech directly.

The differences between the discourse model and the original temporal model are minor: the tense meanings are now relations between R_i and P_i (instead of between R and S) but, since P_i is always a moment of time (just like S), the number of possible tense meanings remains the same. The expressive power of the formalism, however, has been enhanced considerably. It now provides a formalism for the temporal analysis of all types of clauses - whether embedded or not, whether isolated or in context - and for the description of anaphoric temporal expressions. The latter will be discussed in the next paragraphs.

3.5. the anaphoric time adverbials

There is a class of adverbials which do not refer to the time of speech, as the deictic ones, but rather to the time of reference of a dominating or preceding clause. They can also be grouped in the three subclasses

simultan : the same time, at that moment ...
anterior : two weeks before, previously ...
posterior : one week later, then ...

Together with the deictic adverbials they form the class of relational time adverbials.

For completeness sake I also mention the locational time adverbials, such as "at 2 o'clock", "on Monday" and "in the summer". They do not express any relational information and can, therefore, be combined with all possible tenses.

Common to both the relational and the locational time adverbials is that they can be used as answers to "when"-questions. In this respect they differ from the aspectual adverbials (cf. 4.3.).

3.6. anaphoric tenses

The use of the tense forms in texts is somewhat different from their use in single isolated clauses. This is due to the fact that in anterior contexts the present is often replaced by the past and the future by the conditional. This phenomenon, which is called transposition (cf. *Rohrer* 1985), can be seen at work in the following sentences:

(4) he said that he was ill
(5) he entered the room and fell on his face

In (4) the time of his being ill is simultaneous with his saying that he is ill, and in (5) the time of his falling on his face is posterior to the time of his entering the room. In both cases one would expect a present tense in the second clause, but since the first clause is in the past, transposition applies and results in the use of the past tense.

The discourse diagrams for these sentences look as follows:

(4)
he said
that

$R0=S$

$R1$

$P1$

$V(P1,R0)$ & $<(R1,P1)$

he was
ill

$R2$

$P2$

$V(P2,R1)$ & $\supseteq(R2,P2)$

(5)
he entered
the room

$R0=S$

$R1$

$P1$

$V(P1,R0)$ & $<(R1,P1)$

and fell
on his
face

$R2$

$P2$

$V(P2,R1)$ & $>(R2,P2)$
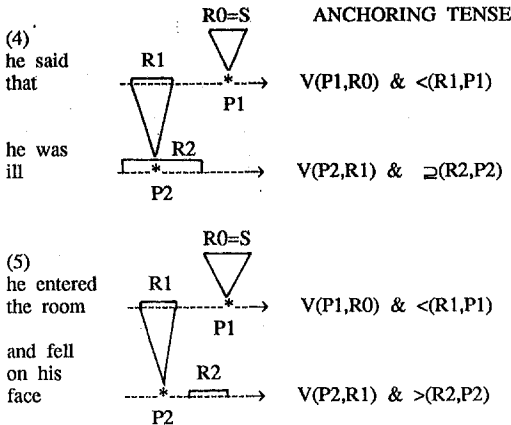
Similar remarks can be made about the use of the conditional in

(6) we all hoped that he would soon recover

From the point of view of analysis there are at least two ways of dealing with the phenomenon of transposition: it can be treated as a syntactic transformation or as an irregularity in the relation between form and meaning. In the former case one first maps the past on the present and the conditional on the future and then applies the normal rules for the assignment of meanings. In the latter case one defines extra rules for the assignment of meanings to the past and the conditional tenses.

The former alternative is more constrained than the latter and, hence, more attractive, but the choice for the one or the other might turn out to be language dependent.

4. A THEORY OF ASPECT

4.1. the aspect meanings

There is a considerable confusion in the literature about the definition of aspect. This is largely due to the fact that many authors donot make a distinction between grammatical aspect and lexical aspect. The former concerns the syntax and semantics of aspectual auxiliaries and adverbials, whereas the latter concerns the semantics of main verbs and propositions (cf. the event/state/process distinction). In this paper I will use the term *aspect* for the former only. The latter will be called *Aktionsart*.

As a general definition of aspect I will adopt the formulation by Marion 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, 153]

Starting from this definition of aspect meanings as binary relations between intervals and combining it with the observation that the number of possible binary relations between intervals is seven (cf. 2.) it is possible to predict that there will be seven aspectual relations. In the following paragraphs I will discuss them in some detail and relate them to the traditional aspectological terminology.

A well-known aspectual distinction is the one between the perfective and the imperfective. The perfective presents a situation as a single unanalysable whole, whereas the imperfective looks at a situation from the inside and focusses on the beginning, ending or continuation of it (cf. *Comrie* 1976, 3-4).

As formal counterparts of these definitions I propose the relations $=(E,R)$ and $\subset(E,R)$ for the perfective :
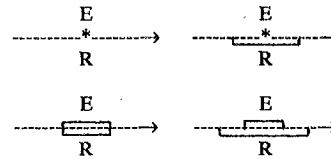
These relations express the intuition that the time of event (E) is seen as one unanalysable whole from the point of view of the reference time. The formal definition of the perfective is, hence, $\subseteq(E,R)$.

For the imperfective I will make a distinction between three types. If the focus is on the continuation, the aspect is durative. For its representation I use the relation of proper inclusion : $\supset(E,R)$.
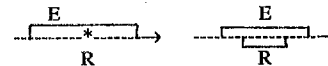
The situation is clearly looked at from the inside : R is in E.

For the two other types of imperfectivity I will make use of the overlap relations :

$>>(E,R)$ $<<(E,R)$
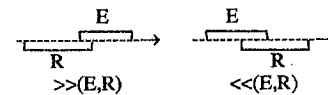
In the case of right overlap the focus is on the beginning of the situation. This aspect I will call the inchoative. In the case of left overlap the focus is on the end of the situation. This aspect I will call the terminative.

Another aspect that is often mentioned in the literature is the socalled perfect ($\neq$ perfective !). In conformity with Reichenbach, Johnson and others I will analyse it in terms of precedence : $<(E,R)$. I will, however, not use the term "perfect" for it, but rather the term "retrospective". The reason for this is that the perfect aspect form should be distinguished from the retrospective aspect meaning : the former is syntactic, the latter is semantic, and the relation between both is not necessarily one-to-one.

Finally, there is the inverse of the retrospective, i.e. the prospective : $>(E,R)$. It is one of the meanings of the English "be going to" and of the French auxiliary "aller".

In short, there are six different aspect meanings. Their language specific counterparts are aspectual auxiliaries and adverbials.

4.2. the aspectual auxiliaries

As for the Eurotra languages the aspect form systems show a larger diversity than the tense form systems.

Some typical aspectual distinctions are the ones between

- perfect and non-perfect (have + past participle)

- progressive and non-progressive (be + present participle)

- go and non-go (go + to-infinitive)

The first distinction is made in all of the Eurotra languages, but the two other ones are not euroversal. They are present in English, but not in German and Danish, for instance, and French has the third distinction, but not the second.

It may be worth stressing that I will only analyse the aspectual auxiliaries. Full lexical verbs, such as "stop", "start" and "continue", and periphrastic forms, such as "être en train de" and "venir de", will not be discussed here.

For English, the set of aspectual auxiliaries can be defined as follows:

aspect form → (have+papa) (be+prpa(go+to-inf))

∅	∅	∅	= Simple	rain	
have	∅	∅	= Perfect	have rained	
∅	be	∅	= Progressive	be raining	
have	be	∅	= Perf Progr	have been raining	
∅	be	go	= Go	be going to rain	
have	be	go	= Perfect Go	have been going to rain	

For French, the definition looks as follows :

aspect form → $\begin{bmatrix} \text{avoir/être+papa} \\ \text{aller+inf} \end{bmatrix}$

∅	= Simple	pleuvoir
avoir/être	= Composé	avoir plu
aller	= Futur proche	aller pleuvoir

As for the assignment of meanings to the auxiliaries I will follow the same procedure as for the tense meanings. The meaning of an aspect form is an element of the power set of possible aspect meanings. This set contains 64 elements.

For the definition of the mappings one can start from the following euroversal scheme (euroversal = common to the nine Eurotra languages) :

simple	→	{perfective}
perfect	→	{retrospective}
go	→	{inchoative}
progressive	→	{durative}

If one of these forms is not present in the language, its meaning may be expressed by another form. In general this will be the form whose basic meaning is the least distant from the meaning to be expressed. For a specification of the notion of distance between meanings I will use the following scheme:

retro	term	perf,dur	incho	pro
<	<<	⊏ , ⊐	>>	>
1	2	3	4	5

The distance between two aspect meanings is equal to the difference of their numbers.

It follows from the principle of minimal distance that a language without a progressive will express the durative by means of the form which expresses the perfective (|3-3|=0), i.e. the simple form. This is indeed true for French, Dutch, German and Danish.

The principle also guides the choice of a form for the expression of the terminative. Some languages have a special form for this aspect. English, for instance, has the perfect progressive for this meaning. Most languages, however, donot have such a form and in those cases the minimal distance principle predicts which forms can be used for the expression of terminativity, i.e the (retrospective) perfect form, the (perfective) simple form or the (durative) progressive form: |2-1|=|2-3|=1 (see also 4.3.).

Taking into account the basic scheme and the principle of minimal distance, and complementing it with language specific observations, one can derive the following mappings for English and French:

English :	Simple	→	{perf}
	Perfect	→	{retro,term}
	Progressive	→	{dur,perf}
	Perf progr	→	{term}
	Go	→	{incho,pro}
	Perfect Go	→	∅
French :	Simple	→	{perf,dur,term}
	Composé	→	{retro,term}
	Futur proche	→	{incho,pro}

4.3. aspectual adverbials

The aspectual adverbials include the duration adverbials and the boundary adverbials.

The duration adverbials specify the length of the time of event. Depending on whether the basic proposition is an event or a state/process they are expressed by an IN-adverbial or a FOR-adverbial:

(7) she ran the mile in five minutes
(8) he has been sleeping for ten hours
(9) we have been in France for a month

They do not express any relational information.

The boundary adverbials specify the beginning and/or the end of the time of event. They are prepositional phrases introduced by "since", "from", "until", "till", "from ... till".

One of these expresses relational information: the "since"-adverbials denote an interval which begins in the past at some specified time, e.g. Christmas in "since Christmas". The end point of such an interval is not specified by the adverbial, but is normally taken to be included in the time of reference. The relation between time of event and time of reference will, hence, be one of overlap:

It follows that "since" adverbials express terminativity and that the compatibility of these adverbials with the aspect forms can be used as a test for deciding whether a given aspect form can be terminative.

What the aspectual adverbials have in common is that they can be used as answers to "how long"-questions. This distinguishes them from the time adverbials.

5. DEGREES OF COMPOSITIONALITY

Tense and aspect forms do not occur in isolation: finite verbs have both a tense form and an aspect form. The meaning of their combination is the relational product of the meanings of the tense form and the meanings of the aspect form. An example: the meaning of the English present perfect progressive is the relational product of the meanings of the present tense with the meanings of the perfect progressive aspect. In other words, the meaning of the present perfect progressive is compositional.

Not all combinations of tense and aspect are compositional, though. In some cases a form can have a meaning which cannot be derived compositionally, in other cases a form may lack a meaning which can be derived compositionally.

An example of the former type is the present perfect in languages like French, Dutch and German. Apart from or even instead of its compositional meanings this form has an <anterior, perfective> meaning, i.e. the meaning of the English simple past. This appears a.o. from their compatibility with anterior time adverbials:

FR   je l'ai vu hier
DU   ik heb hem gisteren gezien
GE   ich habe ihn gestern gesehen
EN * I have seen him yesterday
       I saw him yesterday

An example of the latter type is the French *passé simple*. The simple aspect can have three different meanings in French, but in combination with the *passé*, it can only have the perfective interpretation: the durative and the terminative meaning are expressed by the *imparfait*.

Depending on how many exceptions there are, the tense and aspect system of a given language will be more or less compositional. In case of a low degree of compositionality one could decide to assign meanings to combinations of tense and aspect forms, rather than to tense and aspect forms separately.

6. THE SYSTEM IN USE

Eurotra is a transfer based system. The integration of the given analyses in the Eurotra framework has been achieved as follows.

In analysis the tense and aspect forms are mapped onto their meanings. This mapping is many-to-many and will, hence, result in the assignment of many meanings to one and the same form. Disambiguation is done on the basis of the context. Factors to be taken into account are the temporal adverbials and the Aktionsart of the basic proposition.

In transfer the tense and aspect meanings are simply copied: their representations are interlingual.

In generation the meanings are mapped onto forms. Unlike the mapping in analysis, this mapping is a function.

The system presented in this paper has been applied to the nine Eurotra languages and has been implemented in terms of the unification based Eurotra formalism. Still lacking at this moment are the treatment of the transposed uses of the tenses and the rules for determining the Aktionsart of basic propositions.

REFERENCES

Bennett Michael & Partee Barbara (1972), Toward the logic of tense and aspect in English. Indiana University Linguistics Club.

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*. Cambridge, Cambridge UP.

Comrie Bernard (1985), *Tense*. Cambridge, Cambridge UP.

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

Johnson Marion (1981), A unified temporal theory of tense and aspect. In: Tedeschi & Zaenen (eds), *Tense and aspect*. New York, Academic Press.

Kamp Hans & Rohrer Christian (1983), Tense in texts. In: Bäuerle, Schwarze & von Stechow (eds), *Meaning, use and interpretation of language*. Berlin, De Gruyter.

Partee Barbara (1984), Nominal and temporal anaphora. In: *Linguistics and Philosophy 7*, 243-286.

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

Rohrer Christian (1985), Indirect discourse and "consecutio temporum". In: Lo Cascio & Vet (eds), *Temporal structure in sentence and discourse*. Dordrecht, Foris.

Smith Carlota (1980), Temporal structure in discourse. In: Rohrer (ed), *Time, tense and quantifiers*. Tübingen, Niemeyer.