# ITERATION, HABITUALITY AND VERB FORM SEMANTICS

Frank van Eynde
University of Leuven
Maria-Theresiastraat, 21
3000 Leuven
Belgium

## ABSTRACT

The verb forms are often claimed to convey two kinds of information :

1., whether the event described in a sentence is present, past or future (= deictic information.
2. whether the event described in a sentence is presented as completed, going on, just starting or being finished (= aspectual information.

It will be demonstrated in this paper that one has to add a third component to the analysis of verb form meanings, namely whether or not they express habituality.

The framework of the analysis is model-theoretic semantics.

## BACKGROUND

The analysis of **iteration and habituality in** this paper is part of a comprehensive semantic analysis of temporal expressions in natural language. The research on this topic is carried out in the framework of EUROTRA, the MT project of the European Community. It is reported on extensively in Van Eynde (1987).

The original motive for starting this research was the fact that verbal tenses and temporal auxiliaries do not correspond one-to-one in the languages that EUROTRA has to deal with. Compare for instance

(1) EN he <u>has lived</u> in Copenhagen for 20 years
(2) DK han <u>har boet</u> i Kobenhavn i 20 ar

with their equivalents in the following languages

(3) DE er <u>wohnt</u> seit 20 Jahren in Kopenhagen
(4) FR il <u>habite</u> à Copenhague depuis 20 ans
(5) NL hij <u>woont</u> sinde twintig jaar in Kopenhagen

when translating from English or Danish to German, French or Dutch the present perfect has to be replaced by a simple present.

Differences like these can be handled in one of two ways: either by defining complex mappings from source language to target language forms in transfer or by defining mappings between language specific forms and interlingual meanings in the monolingual components.
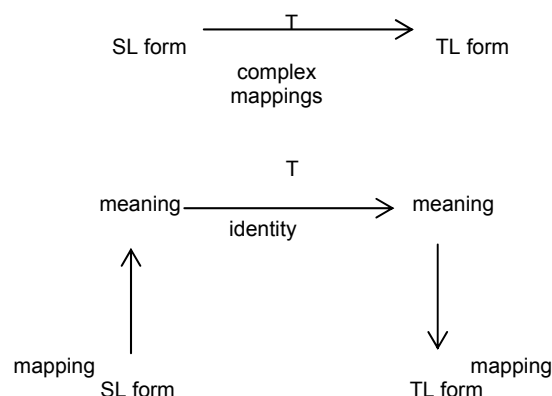


Because of EUROTRA's adherence to the principle of "simple transfer" it was quite obvious from the start that the interlingual approach was the one to opt for. It will, hence, be adopted in this paper as well.

The paper consists of three parts.

In the first I will present a formalism for the representation of time meanings, together with a model for the interpretation of those representations. In the second this formalism will be extended so that it can also be used for the analysis of iteration and habituality. And in the third part I will show how the extended formalism can be used for an interlingual analysis of the verb forms.

## THE CORE FORMALISM

### A Temporal Model

The formalism that will be used here has been defined explicitly in van Eynde, des Tombe & Maes (1985). In this paper I will only give a short informal presentation of the formalism, concentrating on those parts which will be needed in the second part.

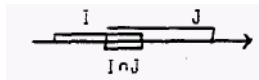The model consists of a set of linearly ordered intervals.

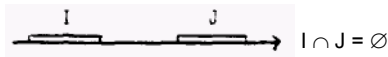An interval is a continuous set of time points or, the time axis :
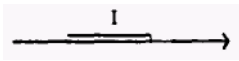
A la limite it might consist of one moment of time :
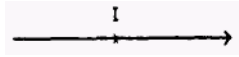
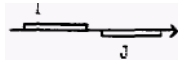For any pair of intervals one can define their intersection as the set of time points which they share :

$I \cap J$

This set might also be empty, as in
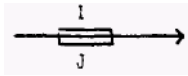
$I \cap J = \varnothing$

It is, furthermore, possible to define some binary relations between intervals, such as

Precedence

    I before J   $<$ (I, J)

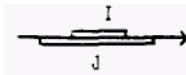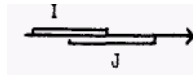    J after I   $>$ (J, I)

identity

    I simul J   $=$ ( I , J )

contain

    I part-of J   $\subset$ ( I,J)

    J contain I   $\supset$ (J, I)

overlap

    I leftover J   $<<$ (I, J)

    J rightover I   $>>$ (J , I)

These relations are also used in Bruce (1972).

**A Format for Representation**

For the semantic analysis of the temporal expressions I will start from the assumption that every sentence can be analysed in two parts : the temporal information expressed by the tenses, auxiliaries and adverbials on the one hand, and a basic atemporal proposition on the other hand.

(6) the cat sat on the mat

will, for instance, be analysed in a basic proposition "the cat sit on the mat" and the information conveyed by the past tense.

The relation between both is established in two steps : the basic proposition is first related to the interval for which it is said to be true, the socalled time of event (E), and then this interval is related to the time of speech (S) :

$\exists E$ [ $>$E, S) & AT (E, the cat sit on the mat)]

This formula states that "the cat sit on the mat" is true at an interval E which precedes the time of speech S.

Following Reichenbach (1947) I will furthermore assume that the relation between the time of event and the time of speech is mediated by a third kind of interval, namely the time of reference (R). So, instead of the simple Rel(E,S) we will have a composite Rel(E,R) & Rel(R,S).

Next to this relational information the temporal expressions can also give specific information about the location or the length of the relevant intervals. This is typically done by means of time adverbials, such as "next year", "in the spring", "for two years", "till Christmas", etc. This information will be represented by means of one-place predicates over intervals : Pred(E) and Pred (R).

An exception is to be made here for the time of speech, whose precise location or length is never specified by linguistic means, but rather by pragmatic factors. A possible way to reflect this in the formalism is to treat it as an unbound variable.

In sum, the general format for the representation of temporal information looks as follows :

$\exists$ R, E [Rel (R,S) & Pred (R) & Rel (E,R) $k$ Pred (E) &
    AT (E,p)]
      where p is a basic atemporal proposition

An example ;

(7) we will visit Moscow next year

$\exists$ R, E [$>$ (R, S) & next year (R) & $\subset$ (E, R) &
    AT (E, we visit Moscow)]

As it stands this format is not adequate yet for the representation of sentences like

(8)  last year they played chess <u>every week</u>
(9)  he was <u>always</u> late

The basic propositions "they play chess" and "he be late" do not hold for one particular time of event E, but rather for a set of intervals which are spread in time in some way specified by "every week" in (8) and "always" in (9).

In the following part I will introduce an extended formalism which can deal with these types of iteration.

## THE EXTENDED FORMALISM

### Cyclic Iteration

Cyclic iteration is marked by adverbials like "daily", "every Monday", "each year", etc. In Quirk e. a. (1972) they are called **periodic frequency adverbials**.

For the analysis of these adverbials I first introduce the notion **frame time**. The frame time is the interval which contains all the instances of the event described in the basic proposition, In

(8) last year they played chess every week

the frame time is last year. In the genera! format the frame time occupies the same place as the time of event in non-iterative interpretations (= the E-interval).

Next, I define a set of distinct, non-overlapping subintervals (I) which are all part of the frame time. In (8), these intervals have a length of one week each. This gives the following (preliminary) representation :

$\exists$ R,E  I < (R, .S) & last year(R) & =(E,R) &
$\forall$ I  [$\subset$ (I, E) & $\cap$ I = $\varnothing$ & week (I) --------->
      AT (I, they play chess)]]
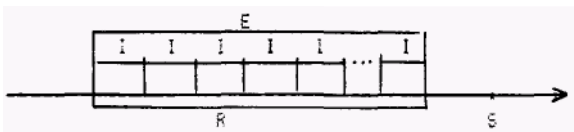


A similar analysis can be found in Stump (1981), where the adverbial frequency adjectives (F) are given the following truth condition :

F $\phi$'  is true in a world w at an interval i
iff   $^\wedge\phi$'' is true in w at non-overlapping subintervals of i distributed throughout i at periods of a specified length l. "

[Stump 1981 , 226 ]

Stump's i-interval corresponds to my frame time, and his non-overlapping subintervals correspond to my I-intervals.

As a representation of (8) this formula is not sufficient, though, since the instances of chess playing do not have to take a whole week for (8) to be true. A more adequate paraphrase is to say that every week contained at least one subinterval (e) during which they played chess :

$\forall$  I [$\subset$ (I, E) & $\cap$ I = $\varnothing$  & week ( 1) ---->
$\exists$  e [$\subset$ (e, I)  &  AT (e, they play chess)]]

An argument in favor of this refinement is that languages have special means for specifying the e-times , In

(10) last year she arrived <u>at 8 o'clock</u> every day

the adverbial "at eight o'clock" denotes the location of the e-interval :
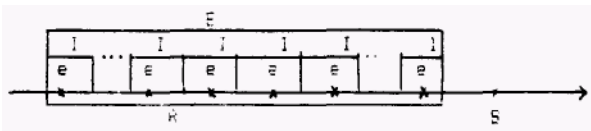


Notice that the properties of e are constant within the frame time :  the adverbial "at eight o'clock" specifies the time of each of her arrivals of last year.

The general format for the representation of cyclic iteration is, hence,

$\exists$ R,E  [Rel (R,S) & Pred (R) & Rel (E,R) & Pred (E) &
$\forall$ I [$\subset$ (I, E)  & = $\cap$ I = $\varnothing$  & P (I)  ---->
$\exists$ e [$\subset$ (e, I)  & M (e)----------> AT (e, p) ] ] ]

where  P is replaced by the head of a periodic frequencv adverbial, specifying the location or the length of I
    M is optionally replaced by a time adverbial, specifying the length of the location of e

An important property of this format is its **chain-like structure :**

R is defined with respect to S : Rel (R,S)
E is defined with respect to R : Rel (E,R)
I is defined with respect to E :  $\subset$ (I, E)
and    e is defined with respect to I :  $\subset$ (e,I)

As it stands, the format does not provide any means tor stating a direct relationship between the intervals inside the frame time (I and e) and the intervals outside the frame time (S and R). As consequence, the formalism predicts that temporal adverbials which are in the scope of a frequency adverbial *(=* the e-specifiers) cannot refer back to the speech time or the reference time: * Rel (e, S) and * Rel (e, R).

A good piece of evidence for this hypothesis is provided by the WHEN-adverbials. In general one can distinguish two kinds of those adverbials : the relational ones, which express a relation between the reference time and the speech time, such as "yesterday" and "tomorrow", and the non-reiational ones, which identify the location of an interval without any reference to the speech time, such as "between 8 and 9" and "at two o'clock".

The interesting thing now is that only the latter adverbials can occur in the scope of a frequency adverbial. Compare

 (11) she arrived every day <u>between 8 and 9</u>
                                      e
*(12) she arrived every day <u>yesterday</u>
                                      e

The fact that the relational WHEN-adverbials cannot occur in the scope of a frequency adverbial provides some positive evidence for not including direct relations between e and S in the formalism. The chain-like structure of the representation format is, hence, linguistically motivated.

## Temporal Quantifiers

The format developed for the analysis of cyclic iteration can also be used for the analysis of the temporal quantifiers, such as "always", "sometimes", "never", "seldom" and "often". The information they provide is less specific than the one provided by the period frequency adverbials, and this should be reflected in their representation.

As a starting point I take the general format for the representation of sentences with a periodic frequency adverbial :

 ...  $\forall$  I [  $\subset$ (I, E)  &  $\cap$ I =  $\varnothing$  & P (I)  ---->

   $\exists$  e [ $\subset$  (e, I)  &  M (e)  & AT (e, p) ] ]

For a semantic analysis of the temporal quantifiers this format has to be generalised.

The most important change is the replacement of the universal quantifier by a variable :

 ... Q I (  $\subset$ (I, E) ...

where Q can be any of the following quantifiers

|  |  |  |
|---|---|---|
| $\forall$ | always |
| $\exists$ | sometimes |
| $\sim\exists$ | never |
| Few | seldom, rarely, now ana then |
| Many | often, frequently |
| Most | usually, mostly, generally |

This sixfold division is taken over from Lewis (1975) .

This analysis accounts for the anomaly of sentences like

? (13) we <u>sometimes</u> played chess <u>every week</u>
              $\exists$                          $\forall$
? (14) they <u>often</u> met <u>every month</u>
            Many              $\forall$
? (15) we <u>always</u> played chess <u>every week</u>
            $\forall$                       $\forall$

These sentences are semantically anomalous because the same kind of information, namely the values of Q, is specified twice. This leads to inconsistency in (13) and (14) where the Q-variable is said to be both universal and non-universal at the same time, and it leads to pleonasm in (15) where the Q-variable is twice said to be universal.

The next question is whether the temporal quantifiers introduce any extra-conditions on those intervals, such as  $\subset$ (I, E),  $\cap$ I =  $\varnothing$  and  P (I).

The first of these conditions appears to be relevant : the temporal quantifiers are indeed interpreted with respect to some given frame time, In

 (9) he was always late

"always" does not denote ALL possible intervals, but only all possible intervals in the past.

The condition that the subintervals may not overlap does not seem to be relevant, though. In

 (16) quadratic equations are always simple

275

the instances for which "quadratic equations be simple" are true are not temporally ordered at all. This might indicate, by the way, that the I-objects are not necessarily intervals, but rather **cases** of **occasions** which can but need not be given a temporal interpretation (cf. Lewis l975).

The third condition concerns the properties of the I-objects. In the case of the periodic frequency adverbials the relevant properties concern the location or the length of the interval. In the case of the temporal quantifiers one could think of specifying a relevance condition, for a sentence like

(9) he was always late

does not mean that he was late at any possible occasion in the past, but rather that he was late on all occasions on which his being late or timely could have mattered.

In Åqvist, Hoepelman & Rohrer (1980) one can find a proposal to incorporate this information in the semantic representation, but I will not adopt this proposal here, since the conditions on the (non)relevance of the occasions are typically determined by pragmatic factors. In

(18) he always leaves at twelve

the relevant occasions (I) could just as well be ail occasions on which he leaves as all occasions on which he leaves for work as all occasions on which he leaves for watching the home game of his favourite football team.

As a result of the foregoing reductions and changes the general format for analysing temporal quantifiers looks as follows :

$\exists$ R, E [Rel R.S) & Pred (R) & Rel (E,R) & Pred (E) &
Q I [$\subset$ (I, E)  --->/&
$\exists$ e [$\subset$ (e, I)  & M (e) &  AT (e, p) ] ] ]

where  Q is replaced by any of ($\forall$, $\exists$, ~$\exists$, Most, Few, Many}
    M is replaced by some time adverbial which specifies the location or the length of e (if there is any)
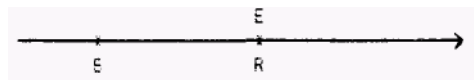
## Habituality

The sentences discussed so far all contain an explicit indication of iteration. The presence of such an indication is, however, not necessary for deriving an iterative interpretation. Take, for instance,

(19) he leaves at twelve

This sentence cannot only mean that he will leave at twelve, but also that he has the habit of leaving at twelve.

In the representation of the former interpretation the time adverbial "at twelve" specifies the time of reference :

$\exists$ R, E [> (R.S) & at twelve (R)  & = (E, R) &
    AT (E , he leaves) ]



In the representation of the habitual interpretation, on the other hand, the time adverbial should be taken to specify the multiple e-time, for the sentence does not report on one of his leavings at twelve, but rather on several of such leaves. As a representation of this interpretation I propose :

$\exists$  R, E [$\supset$ (R.S) &  = (E, R) &
Most I [$\subset$ (I, E)  ---->
$\exists$  e [$\subset$ (e, I)  & at twelve  & AT (e, he leave) ] ] ]
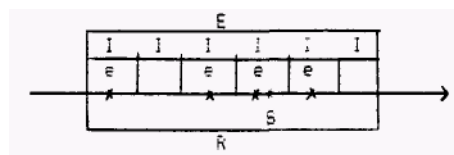


(19) he leaves at twelve

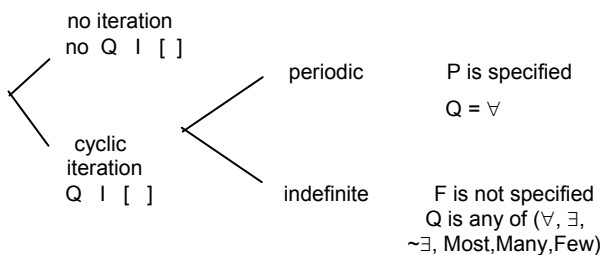is treated as synonymous with

(20) he usually leaves at twelve

If this is felt to be undesirable, one can introduce a special quantifier for marking habituality, but at this moment I do not see any reason for such a move.

The general format for the representation of habitual interpretations is, hence,

$$\exists\ R, E\ [Rel\ (R, S)\ \&\ Pred\ (R)\ \&\ Rel\ (E, R)\ \&\ Pred\ (E)\ \&$$
$$Most\ I\ [\subset (I, E)\ ----->$$
$$\exists\ e\ [\subset (e, I)\ \&\ Pred\ (e)\ \&\ AT\ (e, p)\ ]\ ]\ ]$$

## The Assignment of Representations to Sentences

On the basis of the given analyses one can. distinguish three kinds of sentence meanings :

no iteration
no Q I [ ]

periodic       P is specified
               Q = ∀

cyclic
iteration
Q I [ ]         indefinite      F is not specified
                               Q is any of (∀, ∃,
                               ~∃, Most,Many,Few)

The assignment of these meanings to particular sentences is fairly straightforward when the sentence contains a frequency adverbial or a temporal quantifier, but if there is none of those, then the sentence is ambiguous between a non-iterative and an habitual interpretation (cf. the two interpretations of "he leaves at twelve").

In practice there are some disambiguating factors , though.

1. If the basic proposition (p) denotes a state, then the sentence can not have an habitual interpretation. Compare

(19) he leaves at twelve
(21) he is in jail

(19) can be interpreted as meaning that he has the habit of leaving at twelve, but (21) cannot be interpreted as meaning that he has the habit of being in jail.

2. Certain verb forms can block the derivation of one of the two possible interpretations. Compare

(22) he is drinking coffee
(23) he drinks coffee

(22) can denote a single instance of drinking as well as a recent habit of him to drink coffee (cf. in the sense of "he is drinking coffee nowadays"). (23), on the other hand, can only denote a habit; it cannot be used to report on a single instance of drinking.

This demonstrates the need to distinguish different types of verb forms : the ones that will a:ways elicit an habitual interpretation, the ones that block the derivation of an habitual interpretation, and the ones that admit both kinds of interpretations. The first are unequivocally [+habitual], the second [-habitual] and the last will be given the feature [ + /-habitual].

## THE INTERLINGUAL ANALYSIS OF THE VERB FORMS

### The Meanings of the Verb Forms

In the previous parts I have presented a formalism for the representation of temporal information in sentences. This formalism is especially designed for the analysis of natural language, but not for the analysis of any particular natural language, such as English, Dutch or Kiswahili.

Its main purpose is to provide a conceptually well-defined language for defining and comparing the meanings of temporal expressions in different natural languages. In order to serve this purpose it is not sufficient to have a formalism, though. What is also needed is a general specification of how the semantic representations relate to their language specific counterparts, i.e. the tenses, the temporal auxiliaries and the time adverbials. The former two will further be called verb forms.

For English those verb forms are summed up in the following rule :

Verb form ---> [+/-Past] (will+inf)
               (have + ED)   { be + ING
                               be going to + inf

This rule yields 24 (=2x2x2x3) verb forms.

Their role in the semantic interpretation of sentences can easily be expressed in terms of the given formalism. They specify

1. the relation between reference time and speech time : Rel (R,S) (= deictic information)
2. the relation between event time and reference time : Rel(E,R) (= aspectual information;
3. whether the sentence has an habitual and/or a non-iterative interpretation

The meaning of a verb form can, hence, be represented as a triple <x, y, z> where x and y are substituted for one of the possible binary relations between intervals, and where z is one of the three possible habituality values.

The same verb form can, of course, have different meanings and will, hence, be associated with a set of such triples.

The details of this association have been discussed elsewhere, at least for the x and y values (cf. Van Eynde, des Tombe & Maes 1985). In this paper I will only discuss the z values in some detail.

## The Habituality Value

A good starting point for demonstrating the relevance of the habituality value is provided by the following list of sentences. They are taken from Hess (1985) .

(24) a text editor makes modifications to a text file
(25) a text editor is making modifications to a text file
(26) a text editor made modifications to a text file
(27) a text editor has made modifications to a text file

In (24) it is said "that a test editor makes modifications to a text file in general, almost by definition. We might read this sentence in a reference manual" (Hess 1985, 10).

In (25-27), on the other hand, it is said "that there is, or was, a case of a text editor making modifications to a text file. These remarks might be made by a system operator, watching his screen" (ib.).

Hess concludes from these observations that the quantifier of the subject is universal in (24) and existential in (25-27). However, this conclusion does not follow automatically. In terms of the formalism presented in this paper one could say that (24) has an habitual interpretation, whereas the other sentences have a non-iterative interpretation, In the former case the existential quantifier of the subject will be in the scope of the Most-quantifier, whereas in the latter case it will not be in the scape of any non-existential quantifier, and this accounts for the difference in interpretation without having to postulate two possible meanings for the indefinite article.

Hess's examples are useful in this context, though, because they clearly illustrate the role of the verb form in the interpretation. Since it is the only variable part in the sentences, the differences in interpretation can only be ascribed to them, more specifically to their habituality value.

For the assignment of an habituality value to a given verb form one has to test whether it can or cannot elicit an habitual interpretation in some given context. In testing this one should

1. always use sentences with a non-stative basic proposition, for if the latter is stative the sentence can never be habitual (cf. supra) ;
2. pay attention to the other interlingual values of the verb form. The English simple present, for instance, is unequivocally [+habitual] in its simultaneous meaning, but in its posterior meaning it can be [-habitual] too (cf. the non-iterative interpretation of "he leaves at twelve").

The relevance of the [+/-Habituality]-distinction has so far only been demonstrated from a monolingual semantic point of view. It is, however, possible to give some translational evidence for this distinction as well.

The relevant cases are the ones where the corresponding verb forms have different habituality values. A good example of this is the translation of the Dutch simple present in English.

The Dutch simple present can be both habitual and non-habitual in its simultaneous meaning :

(28) hij drinkt alleen whisky  <simul,y,+habitual>
     "he drinks only whisky"
(29) kijk, hij drinkt koffie  <simul,y,-habitual>
     "1ook, he drinks coffee"

The English simple present, on the other hand, is always habitual in its simultaneous meaning (unless in sentences describing states, of course

(30) he only drinks whisky   <simul,y,+habitual>
* (31) look, he drinks coffee  <simul,y,-habitual>

For the expression of simultaneous non-iterativitv one has to use the progressive :

(32) look, he is drinking coffee

As a consequence, the mapping of (29) to (32). involves a non-trivial tense replacement, and it is one of the merits of the given formalism that it can handle this in an interlingual way.

## REFERENCES

Åqvist Lennart, Hoepelman Jaap & Rohrer Christian (1980), "Adverbs of frequency", in Rohrer (ed.), Time, tense and quantifiers. Niemeyer, Tübingen, 1-17.

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

Hess Michael (1985), "How does natural language quantify?", in Proceedings of the Second Conference of the European Chapter of the ACL, Geneva, 8-15.

Lewis David (1975), "Adverbs of quantification", in Keenan (ed.), Formal semantics of natural language. Cambridge University Press, Cambridge, 3-15.

Quirk Randolph, Greenbaum Sidney, Leech Geoffrey & Svartvik Jan (1972), A grammar of contemporary English, Longman, London.

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

Stump Gregor. (1981), "The interpretation of frequency adjectives", in Linguistics and Philosophy 4. 221-257.

Van Eynde Frank, des Tombe Louis & Maes Fons (1985), "The specification of time meaning for machine translation", in Proceedings of the Second Conference of the European chapter of the ACL, Geneva, 35-40.

Van Eynde Frank (1937), Time. A unified theory of tense, aspect and Aktionsart, An internal Eurotra document *(78* pages). Leuven.