

Inferring aspectuality on French sentences: a minimalist approach

Damien Munch (Munch@Telecom-Paristech.fr) and
Jean-Louis Dessalles (Jean-Louis.Dessalles@Telecom-Paristech.fr)
INFRES, Institut Telecom ParisTech, 46 rue Barrault
75013 Paris FRANCE

Abstract

Current models of temporality in language are either inaccurate or too complex to be cognitively plausible. We present a cognitive model of the computation of aspect in French. Our approach emphasizes the importance of minimalism for cognitive plausibility: structures and computation are kept simple and combinatorial explosion is avoided. Though the model and its current implementation remain partial for now, our approach opens the way to a generic and cognitively plausible method for the determination of aspect.

Keywords: Cognitive minimalism; Natural Language Processing; Temporal aspect; Temporal reasoning

Introduction

Humans are experts in the communication of temporal information. The coherence of discourse relies on the correct expression of time and aspect, both in narratives (e.g. to mark causality) and in argumentative discussions (think of an alibi). Though significant progress has been achieved in modeling temporal processing, current models are either inaccurate or too complex to be cognitively plausible. In the present paper, we stick to the idea that a plausible model should rely on a minimum number of principles. The paper presents new elements in that direction.

Linguists have established various categorizations of aspect, tense and modality (Vendler, 1967; Comrie, 1976; Veters, 1996, among others). They explain variations of meaning by postulating the existence of rich semantic structure stored in lexical entries. For example, Comrie (1976), based on (Vendler, 1967), associates binary attributes such as achievement, accomplishment, semelfactive or activity to verbs. The challenge is to infer aspect, such as repetition and perfectivity, and to predict semantic incorrectness from the combination of attributes when processing a sentence. One problem is to limit the number of attributes that lexical entries may instantiate in their structure. Another problem is to show that the chosen lexical attributes are sufficient.

In addition to fixed attributes attached to the lexical entries, some logicians and computer scientists introduced a *procedural* component in their models of temporal interpretation. To process *tense*, Reichenbach (1947) introduced a minimalist model based on three dynamical coordinates: Event, Reference and Speech. Despite its impressive description power, this model does not account for tense in nested clauses (Hwang & Schubert, 1992) and it fails to explain the behavior of some tenses in other

languages than English (Dowty, 1979; Comrie, 1985). Since then, Reichenbach's model has been steadily improved. The three coordinates have been changed for intervals and/or have been increased in number (Comrie, 1985; Gosselin, 1996; Elson & McKeown, 2010).

There have been attempts to process *aspect* in a minimalist way as well. Recent TimeML versions (Sauri & al., 2009) consider four attributes: Progressive, Perfective, Perfective_progressive and None. Smith (1991) proposes a model based on only three viewpoints: Imperfective, Perfective, and Neutral. Ghadakpour (2004) uses only two viewpoints, called Figures and Grounds.

Lexical models, in which temporal knowledge is stored in static lexical attributes, face the problem of *attribute defeasibility*. For instance, the verb "to hit" is supposed to have the Punctual attribute; therefore, "she hit the wall for one minute before leaving" receives a repetitive interpretation (several knocks); however, "The small galaxy hit (collided with) the Milky Way for ten million years before collapsing" can receive a non-repetitive interpretation, in contradiction with the supposed Punctual attribute of the verb.

Procedural models, in which lexical entries are given computation power, are able to deal with context. For instance, in Gosselin's (1996) and Schilder's (2004) models, the function assigned to *en* (in French) or *in* (in English) checks whether the complement of the preposition involves duration. Schilder's model even checks whether the phenomenon happened in the past or not. Procedural models, however, are not parsimonious as long as they do not set limits to the computational power of words. For instance Person's (2004) implementation of Gosselin's model of French temporality associates a specific computing rule to each tense and each temporal marker (preposition, temporal adverbs, ...). Similar procedural approaches, in which temporal lexemes are given significant computation power, are proposed by others authors like Saussure (2003) and Schilder (2004). Though these models try to remain parsimonious *in fact*, they are not parsimonious *in principle*. Models in which words may have unlimited power (i.e. they may perform any computation like Turing machines) do not qualify as cognitive models, not only because they lack parsimony but also because they cannot explain how children learn the mechanisms of temporality of the surrounding language.

Models of temporality face another problem. The temporal meaning of a sentence cannot be deduced only from temporal information stored in lexemes. Moens and Steedman (1988) have shown that the mental

representations corresponding to events are not reducible to tense and aspect. They are closer to concepts such as causal sequences, preparatory processes, goals and consequent states. According to these authors, temporal attributes stored in the lexicon cannot capture the richness of interpretation that is accessible to humans. Temporal interpretation would involve causal relationships that lie beyond strict linguistic processing. Models such as Event Calculus (Shanahan, 1999), modeled by Mueller (2004), do take background knowledge into account. The problem, for such models, is to circumscribe the effect of context, not only to avoid unrealistic processing time, but also to keep the systematic character of some temporal phenomena.

Our aim is to design a cognitively plausible model of temporality that avoids the previously mentioned difficulties (attribute defeasibility, unlimited computational power and unlearnability, prohibitive processing time, loss of systematicity). We favor a minimalist approach, in which both structures and procedures are kept minimal. In what follows, we will first list a limited set of examples in French that we use as benchmark. Then we will see how Gosselin's and Schilder's models behave on such examples. We chose these two models because they use concepts similar to ours, such as *viewpoint*, *anchoring* and *granularity*. We will then describe our model and its single procedure: *tMerge*. We conclude with a discussion in which we assess the plausibility and the generality of our approach.

Temporal correctness

Table 1 shows a few examples that have been tested for acceptability. We asked thirty-five French native speakers to answer multiple choices questions about forty-one sentences in French. These sentences were designed to test all combinations of tense, event type and time adverbials. They were proposed in random order and participants were allowed to stop whenever they want. We got an average of twenty answers by sentence. All sentences have the same form: verb phrase + prepositional phrase. Participants were asked the following questions:

- a. Is the sentence correct / incorrect ("one wouldn't say that") ?
- b. Does the event occur several times (repetition) / only once (possibly with breaks) ?
- c. Is the event finished / not finished / don't now ?
- d. Is the event taking place during the whole period indicated by the prepositional phrase / during only part of it / don't now ?

Participants could provide two sets of answers for a given sentence if they thought there were two meanings.

Table 1 shows some of the results. The binary values *unique/multiple*, *perfective/imperfective* and *whole/slice* refer to coded answers to questions b-d. Percentages for a given sentence correspond to participants who found the sentence correct. They may add up to more than 100% when several interpretations were given.

Table 1: Tested sentences

- | |
|---|
| <p>(1) Elle mangera du gâteau en février.
<i>She will eat (be eating) cake in February.</i>
(30%) unique/perfective and slice of February
(80%) multiple/imperfective</p> <p>(2) * Elle mangera du gâteau en 30 minutes.
* <i>She will eat (be eating) cake (or from the cake) within 30 minutes.</i>
(30%) unique/perfective</p> <p>(3) Elle atteindra le sommet en février.
<i>She will reach the top in February.</i>
(76%) unique/perfective and slice of February</p> <p>(4) Elle mangera (à la cantine) pendant deux mois.
<i>She will be eating (at the canteen) for two months.</i>
(100%) multiple/imperfective</p> |
|---|

Table 2 shows classical sentences (examples 5-8) that were not included in the test.

Table 2: Sentences variations

- | |
|---|
| <p>(5) Elle atteindra le sommet en 30 minutes.
<i>She will reach the top within (the next) 30 minutes.</i>
unique/perfective</p> <p>(6) * Elle atteindra le sommet pendant 30 minutes.
* <i>She will be reaching the top for 30 minutes.</i></p> <p>(7) Elle atteindra le sommet pendant le prochain mois.
<i>She will reach the top during the next month.</i>
unique/perfective and slice of the next month</p> <p>(8) Elle mangera du gâteau pendant les 30 prochaines minutes.
<i>She will eat (be eating) cake during the next 30 minutes.</i>
unique/perfective</p> |
|---|

The challenge is to account not only for the acceptability or incorrectness of sentences, but also for the judgements about repetition, perfectiveness and wholeness. The next section examines how Gosselin's and Schilder's models perform on this kind of examples.

The computation of aspect

Gosselin's (1996) model represents perfectivity by considering intervals with two different types of boundaries: *intrinsic* and *extrinsic*. These boundaries are retrieved from the aspectual type of events (telicity, punctuality, dynamicity). For instance "manger du gâteau" ("to eat cake") will take extrinsic boundaries because its aspectual type is supposed to be *semelfactive* (this information is provided by some external cognitive processing).

Repetition appears during conflict resolution, when the granularities of two intervals are different. For instance in example (1), “manger du gâteau” and “février” (February) do not have the same granularity; the conflict is solved by iterating the interval of “manger du gâteau”.

Conflict resolution also involves instructions which can move one or both boundaries of an interval. This will lead to shrinking, expanding or moving one of the conflicting intervals. Slices in our examples would result from shrinking the interval of the adverbial phrases (“février”, “le prochain mois”).

In example (2) (“manger du gâteau en 30 minutes”), “manger du gâteau” is represented by an interval [B1,B2] with *extrinsic* boundaries, whereas “30 minutes” is represented by an interval [ct1,ct2] with *intrinsic* boundaries. Step b (Figure 1) succeeds, but step c fails because the two intervals have incompatible boundaries types.

Though Gosselin’s model seems to work fine, it is at the expense of simplicity. Specific instructions are assigned to ‘operators’, that is, to every lexeme with a temporal meaning, such as tenses and temporal prepositions. Figure 1 shows the instructions associated to the preposition “en” + *duration*. The problem is not only the actual complexity of such instructions, but also the fact that this complexity is not bound in principle.

- | |
|---|
| <p>a) associate an interval [ct1,ct2] to the temporal adverbial</p> <p>b) $ct1 < ct2$ (non-ponctual adverbial, boundaries are dissociated)</p> <p>c) [ct1, ct2] CO [B1,B2] (adverbial coincides with the event)</p> <p>d) [I,II] ACCESS [B1,B2] (boundaries of the event must be 'accessible' by the reference interval ; $I \leq B1$ and $II \geq B2$)</p> <p>The interval of the event [B1,B2] must be intrinsic (when “pendant” + duration need extrinsic boundaries).</p> |
|---|

Figure 1: instructions for “en” + *duration*, adapted from Gosselin (1996)

Schilder (2004) uses neither intervals nor boundaries in his model. Events are given one of the four aspectual values defined in TimeML (Sauri & al., 2009): Perfective, Progressive, Perfective_progressive and None.

Schilder’s model can detect granularity incompatibilities, though it is not clear whether they are solved by operations like slice or repetition.

To deal with the examples of Tables 1-2, he proposes two different functions for each temporal preposition, depending on whether the complement is anchored or not. Figure 2 shows instructions for “in” (note that this function applies to the English or German “in”). In example (5), the event “reach the top” occurs at ‘timestamp’ TS, which is the ‘Document timestamp’ (DTS) *plus* the given duration (DUR) “30 minutes”. The granularity (G) of the event is given by

the document timestamp (Figure 2). Note that this computation does not seem to be always valid in English (for instance, in “She will defeat her opponent in 30 minutes”, meaning “She will play during 30 minutes and win”, the duration of the event should be DUR and not DTS).

<p><i>Anchored</i>: TSDUR</p> <p><i>Unanchored</i>: If Tense = Past then DUR else TSP1G where TS = DTS+DUR and G = gran(DTS)</p>
--

Figure 2: function used by “in”, adapted from Schilder (2004)

Contrary to Gosselin, Schilder chooses to assign functions to all lexemes, not only the ‘temporal’ ones. On the other hand, processing is somewhat simpler, as it treats prepositions as unary instead of binary operators, as proposed by Pratt & Francez (2001). However, Schilder model has the same drawback as Gosselin’s: *each lexeme* is given a dedicated function. As long as there is no indication about how to limit the computational complexity of these functions, models cannot be considered minimalist.

A minimalist model

In this paper, we present a model which is minimal in terms of structures, procedure and memory use. We rely on one single fixed-sized semantic structure, called *temporal Semantic Structure* (tSS), and one single non-recursive operation, called *temporal merge* (tMerge). Note that the use of the term *merge* (related to Chomsky, 1995) instead of *unification* is debatable (Jackendoff, 2005). To achieve this reduction, we decided to exclude several operations from the temporal processing proper, in line with (Moens & Steedman, 1988), considering that they required access to other cognitive modules (general knowledge and perception abilities, syntax, determination). These operations include:

- time location (whether a situation is located in time or not)
- temporal granularity (or order of magnitude) consistency checking
- causality and anteriority checking
- self-similarity checking
- phrase syntactic hierarchy

For our purposes, these operations need not be represented in a cognitively realistic way, as they are considered external to the model. We implemented them in a basic form in our perception module. We now describe the two central components of the model, tSS and tMerge.

The Temporal Semantic Structures (tSS)

The tSS is the only structure processed in the model. A tSS is a non-recursive structure. It contains three attributes: an

image reference (ImageID) and two *switches* (*Viewpoint*, *Anchoring*).¹ These attributes may be uninstantiated at a given step of the processing.

ImageID The image identifier is a reference to a perceptive representation. The term 'image' includes any perceptive representation, either stored or constructed. The mere use of *ImageID* in the tSS allows us to grossly simplify several processes which are supposed to be performed in an external module. That module (called '*Perception*') is loosely defined as including anything which does not pertain to syntactic processing or to temporal processing proper (as defined in our model). This includes all forms of memory and all forms of knowledge, such as the granularity or the date of events. It also includes the ability to decide about anteriority and about self-similarity.

Viewpoint The *Viewpoint* switch may take two values, *figure* (*f*) and *ground* (*g*) (Ghadakpour, 2003). It may be defined in the lexical entry (for example, the French 'imparfait' (imperfect tense) requires a *ground*). However, it is most often determined during temporal processing. Viewpoints are a key element of our model. They provide information about how the speaker regards the temporal phenomenon at a given stage of processing: either 'from the outside' and considering the overall event (*figure*), or 'from within' (*ground*) (exactly as for space). These two values correspond to standard aspect values: perfective and imperfective. However, we consider the viewpoint as a *switch* that may change value during processing.

Anchoring The anchoring *switch* indicates whether *Perception* is able to provide some (absolute or relative) location for the perceptive image. The Anchoring *switch* may take two values: *anchored* (*a*) and *unanchored* (*u*). For example "(any) 30 minutes" will be unanchored, but "these 30 minutes" will be anchored. Some lexemes are ambiguous in this respect: "in February" may mean that we deal with a specific (anchored) or a recurring (unanchored) period. Anchoring bears a close relation to determination. "The concert" is likely to designate an anchored time period, whereas "a concert" refers to an unanchored time period.

The temporal merge operator (tMerge)

The tMerge procedure (Figure 3) is launched whenever a phrase is recognized by the syntactic analysis. In other words, the syntactic and semantic analyses are fully synchronous. tMerge receives two tSS as input, plus the indication that one (the head) dominates the other (the complement). It returns a unique tSS as result. These elements appear as H, C and R respectively in Figure 3 (tSS are indicated in square brackets). Even if we deal here with temporal merge operation exclusively, we assume that the semantic merge operation performed in several other specialized modules (spatial relationships, determination, perception).

(1) The essential part of tMerge consists in a *basic merge* operation (bottom of figure 3): corresponding *switches* in the two input tSS are merely matched for compatibility to produce R.

(2) When basic unification succeeds, unification proceeds to the Perception module (see figure 3) where it generates a new image (this process, omitted from our model, merely concatenates images identifiers). The perceptive merge may apply *viewpoint* constraints to the resulting tSS depending of the nature of the phenomenon: *indivisible* events are bound to be *figures*, whereas *self-similar* events must be

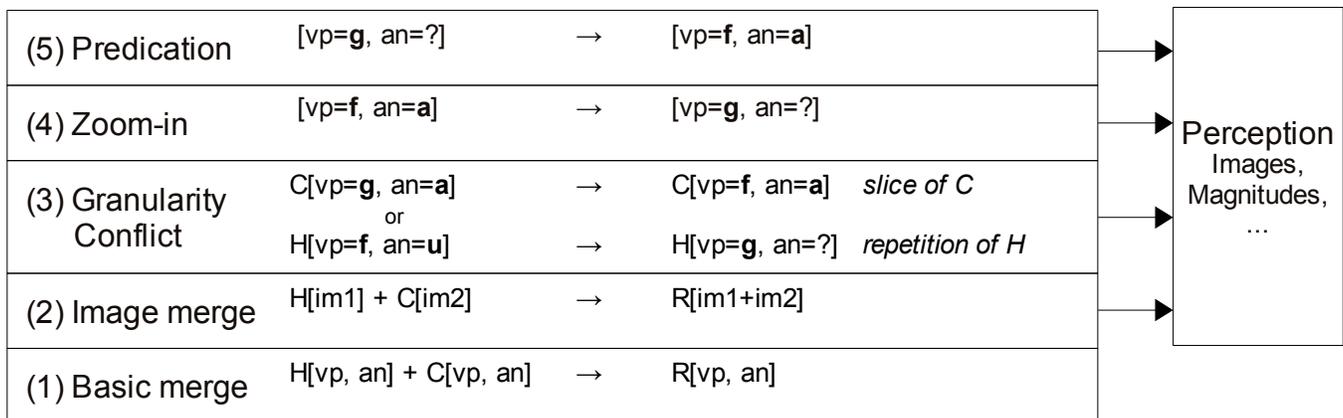


Fig 3: tMerge steps

¹ For the sake of simplicity, we omit *switches* related to tense.

grounds. The perceptive merge also checks for granularity compatibility. All the other operations of figure 3 aim at rescuing basic merge and perceptive merge in case of failure.

(3) The *Granularity conflict* rescue operation is triggered in examples like: “She will eat the cake in February”, where the orders of magnitude are hour *vs.* month. Depending on viewpoint and anchoring constraints, the complement element will be *sliced* (she will eat cake once at some point in February) or the head will be *repeated* (she will be eating cake repeatedly throughout February).

(4) The *Zoom-in* rescue operation may switch a viewpoint that is blocking unification from *figure* to *ground*. It can only apply if the input tSS is *anchored*, if it has an instantiated imageID (for example we can zoom-in on “this month” but not on “30 minutes”), and if Perception is able to create a zoomed image (as in “she reached the top in ten hours”, where one must imagine some definite ultimate climbing phase lasting ten hours).

(5) The last rescue operation is *Predication*. Its effect is to switch one tSS to an *anchored figure*. It requires that *imageID* be instantiated and it can be used only once in a sentence.

The model, characterized by the tSS and the tMerge operation, claims to be minimalist. tSS are not recursive (*i.e.* a tSS does not contain or refer to another structure of same nature, contrary to feature structures like those used in HPSG for instance). A tSS has a fixed length and cannot grow. Moreover, the tMerge operator is ‘amnesic’, which means that the input tSS are lost after the resulting tSS has been computed. This prevents the model from using unrealistic memory resources in uncontrollably growing structures. Many models are monotonic, which means that the structures they process can only grow in size and complexity during processing, becoming unrealistic for large inputs (Ghadakpour, 2003). Our model is non-monotonic and therefore avoids this problem.

Our model has been implemented in Prolog. The program provides all admissible solutions for an input sentence and signals incorrect sentences.

Examples

The model and its implementation account for all sentences of our test, including the examples listed in Tables 1 and 2. It detects “incorrect” sentences; it correctly predicts repetition, slice and perfective and imperfectives aspects. Examples (1) and (2) are detailed below.

(1) Elle mangera du gâteau en février
(*She will eat/be eating cake in February*)

The determiner “du” introduces a *ground* viewpoint. On the other hand, “en” is associated with a *figure*. Let’s consider the step where the two phrases “manger du gâteau” (“to eat cake”) and “en février” (“in February”) are to be unified.

Head: “manger du gâteau” (“to eat cake”)
[im/i_eat_cake, vp/g, an/?]
Complement: “en février” (“in February”)
[i_february, vp/f, an/a]

The *basic merge* (figure 3, (1)) detects a viewpoint conflict. The conflict could be solved either by zooming-in on “en février” (figure 3, (4)) or by predicating “manger du gâteau” (figure 3, (5)), but then the perceptive merge (figure 3, (2)) will detect a granularity difference. We must predicate “manger du gâteau” and zoom-in on “en février” in both case. This leaves us with two solutions.

In the first solution, the *figure* of the head is repeated, and a *ground-ground* merge becomes possible. In the second solution, the complement is sliced and a *figure-figure* merge becomes possible. Slicing is allowed by the fact that “February” is anchored (figure 3, (3)). We thus get the two following interpretations:

Result 1: “manger du gâteau (plusieurs fois) en février” (“to eat cake several times throughout February”)

[i_eat_cake_february, vp/g, an/u]

Result 2: “manger du gâteau” (une fois) en (un moment de) février (“to eat cake once at some point in February”)

[i_eat_cake_february, vp/f, an/u]

(2) * Elle mangera du gâteau en 30 minutes

(* *She will eat (be eating) cake (or from the cake) within 30 minutes.*)

As previously, the tSS of “manger du gâteau” receives a *ground* viewpoint. By contrast with example (1), there is no granularity conflict, but the complement “en 30 minutes” is not anchored. Let’s consider the step where the two phrases “manger du gâteau” (“to eat cake”) and “en 30 minutes” (“within 30 minutes”) are to be unified.

Head: “manger du gâteau” (“to eat cake”)
[i_eat_cake, vp/g, an/?]

Complement: “en 30 minutes” (“within 30 minutes”)
[i_30_minutes, vp/f, an/u]

There is no way to solve the viewpoint conflict: the complement cannot be zoomed-in because it is unanchored. Predication cannot be used to solve the *viewpoint* conflict, since it creates an *anchoring* conflict. The model returns an error, as expected.

Conclusion

We have shown how some of the mechanisms of French aspectuality could be predicted using a minimalist model. We share various notions and mechanisms with Gosselin’s and Schilder’s models, including anchoring, granularity checking and dynamic conflict resolution. Our model departs from theirs by the fact that lexical structures are fixed instead of including algorithms. There is only one procedure in our model: tMerge, which is not attached to the

lexicon and can be synchronized with syntactic analysis. Our model is able to detect and solve aspectual effects such as repetition and slice, to identify the perfectivity and progressivity of events, and to detect incorrect sentences. The output of the model is congruent with the majority judgment among the participants we tested.

The notion of semantic *incorrectness* is relative, as a substantial number of participants considered these sentences as acceptable (e.g. 30% for example (2)). Acceptability seems to depend on several factors that are to be investigated: differences in the kind of computations performed in *Perception*, or differences in judging as correct sentences that wouldn't be uttered by a native speaker but that could still make sense. Another possible source of variation among participants may be judgments of relevance. For instance, "Elle mangera du gâteau en 30 minutes" (understood as: "She will eat cake within a period of 30 minutes") may be acceptable in a context in which any consumption of cake is supposed to require more than thirty minutes. We are currently investigating these phenomena.

Though we are confident in the fundamental principles and in the overall architecture of the model, we need to check its validity against a much larger variety of phenomena, not only in French but also in other languages. For instance, we are currently investigating how the English *progressive* "V-ing" (Deo, 2009) can be explained as a sub-categorization of the *ground* viewpoint depending of perceptive information. These investigations may bring us to adapt and augment the model, while hopefully keeping its minimalist character.

References

- Chomsky, N. (1995) *The Minimalist Program*. Cambridge MA: MIT Press.
- Comrie, B. (1976). *Aspect*. Cambridge university press.
- Comrie, B. (1985). *Tense*. Cambridge university press.
- Deo, A. (2009). Unifying the imperfective and the progressive: partitions as quantificational domains. *Linguistics & Philosophy* 32 (5), 475–521.
- Dowty, D. (1979). *Word Meaning and Montague Grammar*. Dordrecht: D. Reidel.
- Elson, D., & McKeown, K. (2010). Tense and Aspect Assignment in Narrative Discourse. In *Proceedings of the Sixth International Conference in Natural Language Generation*.
- Ghadakpour, L. (2003). *Le système conceptuel, à l'interface entre le langage, le raisonnement et l'espace qualitatif: vers un modèle de représentations éphémères*. PhD Thesis, Ecole Polytechnique.
- Gosselin, L. (1996). *Sémantique de la temporalité en français*. Bruxelles: Duculot.
- Hwang, C.H., & Schubert, L. K. (1992). Tense Trees as the fine structure of discourse. *Proceedings of ACL'1992*, 232-240.
- Jackendoff, R. (2005). Alternative minimalist visions of language. *Chicago Linguistics Society (CLS)*, 41, 189-226.
- Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational linguistics*. 14 (2), 15-28.
- Mueller, E. T. (2003). Story understanding through multi-representation model construction. In Hirst, G. and Nirenburg, S. (eds), *Text Meaning: Proceedings of the HLT-NAACL 2003 Workshop*, (pp. 46-53). East Stroudsburg, PA: Association for Computational Linguistics.
- Person, C. (2004). *Traitement automatique de la temporalité du récit: implémentation du modèle linguistique SdT*. PhD Thesis, Université de Caen Basse-Normandie.
- Pratt, J., & Francez, N. (2001). Temporal Generalized Quantifiers, *Linguistics and Philosophy* 24, 187–222.
- Reichenbach, H. (1947). *Elements of Symbolic Logic*. Free Press.
- Sauri, R., Goldberg, L., Verhagen, M., & Pustejovsky, J. (2009). *Annotating Events in English. TimeML Annotation Guidelines*. Brandeis University. Version TempEval-2010.
- Saussure, L. de (2003). *Temps et pertinence. Eléments de pragmatique cognitive du temps*. Bruxelles: Duculot.
- Schilder, F. (2004). Extracting meaning from temporal nouns and temporal prepositions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 3 (1), 33-50.
- Shanahan, M. (1999). The event calculus explained. In Wooldridge, M.J., Veloso, M. (eds.), *Artificial Intelligence Today*. LNCS, 1600, (pp. 409–430). Berlin: Springer.
- Smith, C. S. (1991). *The Parameter of Aspect*. Dordrecht, NL: Kluwer.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press.
- Vetters, C. (1996). *Temps, aspect et narration*. Rodopi.