

# Study of lexical aspect in the French medical language.

## Development of a lexical resource

Agathe PIERSON<sup>1</sup> and Cédric FAIRON<sup>1</sup>

(1) CENTAL, Collège Léon Dupriez, Place Montesquieu 3, 1348 Louvain-la-Neuve  
 agathe.pierson@uclouvain.be, cedrick.fairon@uclouvain.be

### Abstract

This paper details the development of a new linguistic resource designed to integrate aspectual values in temporal information extraction systems. After a brief review of the linguistic notion of aspect and how it got a place in the NLP field, we present our clinical data and describe the five-step approach adopted in this study. Then, we describe our French linguistic resource and explain how we elaborated it and which properties were selected for the creation of the tables. Finally, we evaluate the coverage of our resource and we present several prospects and improvements to foresee.

## 1 Introduction

Being able to model the chronology of events is paramount in the medical field, especially in electronic health records. Temporal reasoning indeed plays an important role at different stages of patient care: tracking disease status, decision support, prevention of side effects, recognition and discovery of health problems, choice of the appropriate treatment and care quality (Botsis et al., 2011; Chai et al., 2013; Sojic et al., 2016). Because this has yet to be properly implemented in clinical software, it is essential to keep developing NLP techniques and methods that efficiently extract temporal information and medical events found in patient records.

Temporal information extraction has been widely studied. Several methods have been developed to obtain ever more satisfactory results in the extraction of this information, whether by a

statistical method (Li and Patrick, 2012) or by a hybrid method combining rule-based and machine-learning pattern (Lin et al., 2013), statistical and symbolic approaches (Tapi Nzali, Tannier and Neveol, 2015a), neural networks and support vector machines (Tourille et al., 2017).

These studies were mainly concentrated on English data, ‘due to the lack of publicly available annotated corpora’ for other languages, including French (Sun et al., 2013). Recently Campillos et al. (2018) produced a French written clinical corpus with the annotation of temporal entities, attributes, and relations, but it is not yet freely open to the research community.

In this paper, assuming that consideration of aspect could improve the precision of temporal information extraction (see Fig. 1), we undertook to develop an innovative resource which encodes linguistic properties related to the French verbal aspect in medical language. We will test it, within a system of temporal information extraction, in a later paper.

After a brief introduction of the concept of aspect from linguistic and NLP angles and a review of existing resources (section 2), we will define our corpus (section 3) and the method (section 4) used in this research to elaborate this linguistic resource (section 5). We will then assess the coverage of the resource (section 6) and discuss directions for future works (section 7).

## 2 State of the art

### 2.1 The aspect, a linguistic category

In linguistics, the grammatical category of

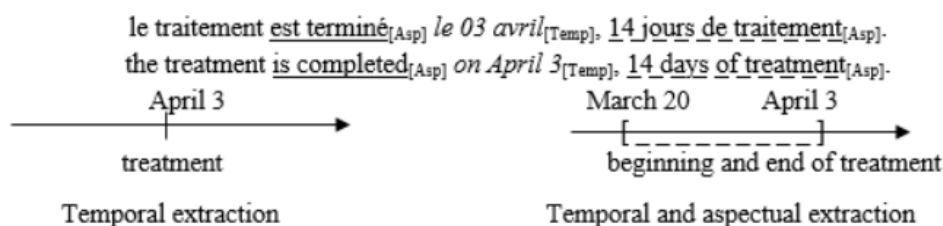


Figure 1. Representation of temporal information extraction vs. temporal and aspectual information extraction

*aspect* refers to “the representation that the speaker makes of the process expressed by the verb (or by the action name), that is to say, the representation of its duration, its course or its completion (inchoative, progressive, resultant, perfective aspects, etc.)” (Dubois et al., 1994: 53).

*Aspect* in French refers to three components that should be dissociated: grammatical aspect, Aktionsart (or semantic aspect) and lexical aspect. The grammatical aspect is “a significant variation of the verb related to a choice of the speaker and, in this respect ‘subjective’” (Confais, 2002: 202). It is realized morphologically through verbal inflections. Smith (1991) has extensively theorized on the *viewpoint* or grammatical aspect in which she distinguishes the perfective (1), imperfective and neutral aspects; after her, Caudal (2006) adds the resultative aspect.

(1) J'ai rencontré cette patiente dont j'ai opéré le mari d'une obésité morbide par Sleeve.

“I met this patient whose husband I performed a sleeve surgery on for morbid obesity.” (Perfective) The Aktionsart (see (2)) describes *eventualities* (Bach, 1986) according to criteria such as telicity, durativity, atomicity or dynamicity (Vendler, 1957; Comrie, 1981; Moens and Steedman, 1988). For the calculation, the verbal diathesis is used, i.e., the verbal kernel extended with its completion and its possible modifiers.

(2) On a surveillé le patient pendant 48h.

“We monitored the patient for 48 hours.” (Activity) The lexical aspect is expressed by the very meaning of the verb (Karolak, 2008; Kozareva-Levie, 2011). This aspect “isolates a moment of the process or specifies the place of the said process in an event chain” (Wilmet, 2001: 63) and presents it as “the result of a selection operation of part of the *constituent time* of this process” (Gosselin, 2011: 149). This aspect can be rendered either by the full verb (3) or by a semi-auxiliary, serving as a support verb (e.g. *continuer* ‘to continue to’).

(3) J'ai préféré interrompre l'examen au niveau de l'angle droit.

“I preferred to interrupt the examination at the right angle.”

Our objective is not so much to end the aspectual controversy, but rather to identify and select the factual elements from this field that are best suited for operationalization and automatic recognition. In this paper, we focus on the internal lexical aspect (Gosselin, 2011: 149-150). Indeed, it seems particularly relevant to us, in a medical linguistics project, to identify events such as the

onset of a disease, the performance of surgery, the resumption of an examination or the end of treatment; all essential aspectual elements to be placed on a patient’s timeline.

## 2.2 What about the aspect in NLP?

NLP researchers have proposed solutions for integrating the aspectual dimension into temporal information extraction system. This integration is made difficult by the complexity of the aspectual phenomenon about which “opinion is often divided as to the appropriate aspectual categories and their realizations” (Bittar, 2010: 33). Among the different methods listed above, we are mainly interested in resources that encode aspectual properties in order to be integrated into information retrieval and information extraction systems. In the late twentieth century, Klavans and Chodorow (1992) developed an aspectual classification system that establishes a distinction between stative and non-stative events using scenario templates. Later, Siegel and McKeown (2001) finalized a statistical method for automatic aspectual classification by prediction on co-occurrence frequencies between verbs and linguistic modifiers.

In 2003, the TimeML specification of event and temporal expressions in natural language text was introduced (TimeML; Pustejovsky et al., 2003). These annotation guidelines include two aspectual levels: i) the aspectual attribute that concerns grammatical aspect; ii) the ALINK that represents the relationship between an aspectual event and its argument event (*initiates*, *continues*, *culminates*, *terminates*).

The following year, the TimeML standard is revised under the name of ISO-TimeML (ISO, 2008) and two aspectual values are added: *imperfective* and *imperfective\_progressive*. This new standard is tested in Task B of the Shared Task ‘TempEval-2’ (Pustejovsky and Verhagen, 2009). In his thesis on the adaptation to French of the ISO-TimeML standard, Bittar (2010) studies the aspectual variants of support verb construction as well as the automatic processing of several aspectual periphrases. He added value to the aspectual attribute, *prospective*, and enriched the ALINK set with the *reinitiates* label.

In the medical field, Styler IV et al. (2014) presented the Thyme-TimeML that integrates contextual aspect attribute – which distinguishes between *intermittent*, *constant* and *new* events –

and permanence attribute – which differentiates between *chronic* and *acute* illness. Campillos et al. (2018) extended the notion of aspect to all that “encode[s] a change (or lack of change) with regard to an entity: Continue, decrease, Improve, Increase, Reccurrence, Start\_Again, Start, Stop, Worsen” (580) when creating an annotated corpus of French medical records.

### 2.3 Aspectual resources

As far as we know, there are no resources in English or in French that are similar to ours, which is focused on the medical language. However, various resources record some aspectual verbs or constructions and detail their functioning. In this way, FrameNet (Baker et al., 1998) labels semantic roles of verbal arguments for English, in order to show the meaning and usage of verbs senses. FrameNet contains some aspectual nuances for which the possible lexical units are indicated (e.g. *Activity\_finish: complete, conclude, finish*), but without any indication of the syntactic construction or type of arguments needed by the verb to take this meaning. The values of *culmination* and *resumption* are not considered here. In the French FrameNet (Candito et al., 2014), no aspectual construction is examined, not even the beginning or the end of an activity.

The lexical resource VerbNet (Kipper-Schuler, 2005) and its French equivalent Verb $\exists$ net (Danlos, Nakamura and Pradet, 2014) have a class of aspectual verbs (class 55) in which the verbs expressing the value of beginning, continuation or end could be found. For each of these values, the verbal valency and syntactic constructions of the prototypical verb (e.g. *begin/commencer* in ‘begin-55.1’) are indicated. We deepen this research in three ways: i) by adding some aspectual subclasses, ii) by describing the lexical units that are synonymous with the verbal prototype and iii) by specifying the semantic class of the arguments.

Two other resources, manually created, exist for French and constitute the theoretical framework of this research: the Lexicon-Grammar (Gross, 1975) and the *Verbes français* (Dubois and Dubois-Charlier, 1997). The first one is based on a descriptive formalism, which we have tried to reproduce to make it automatable; the second one lists a series of aspectual verbs that we have enriched so that it takes into account the medical specificities related to our corpus.

NLP research tackles the notion of aspect. However, there is still a substantial gap to fill, both in French and in the medical field. Our contribution lies in this vast new field of knowledge.

## 3 Data

It is admitted that the temporal information and how it will be presented will depend on the type of documents analyzed (Tapi Nzali, Tannier, Névéol, 2015b). Clinical documents are very heterogeneous and time references are adapted accordingly. To avoid being restricted to one type of document, we collected a corpus which includes five types of clinical texts in the following proportions (see Table 1):

Operating and Review Protocols	55,174
Letters from Doctor to Patient	3,906
Letters from Doctor to Doctor	70,381
Consultation Notes	49,482
Hospitalization Reports	7,833
<b>Total</b>	<b>186,776</b>

Table 1. Types and number of medical texts

These documents (about 54 million tokens), dating from 1996 to 2014, were collected and anonymized as part of a project with six services (abdominal and bariatric surgery, gastroenterology, MRI, scanners, and ultrasound) of a Brussels hospital (iMediate, 2014-2016). We deemed appropriate to use such a considerable amount of data since we focus on a very specific phenomenon, i.e., the lexical aspect conveyed by verbs.

## 4 Methodology

We created the resource following a 5-step method. Firstly, we established the verbal seed terms list that could evoke a lexical aspect. For this purpose, we started from the definition of aspectual relationship in the ISO-TimeML Standard for annotation (Pustejovsky et al., 2010). Five aspectual relationships are considered, here associated with the prototypical French verb: Initiates (*commencer*), Continues (*continuer*), Terminates (*terminer*), Culminates (*accomplir*) and Reinitiates (*recommencer*) (Bittar, 2010: 80). Ideally, categories of verbs expressing the possible outcome of a hospitalization (*healing, improvement, stability, deterioration, and death*; DEFT, 2019) or verbs indicating the *progress* or *decline* of a patient should also be implemented (Elhadad et al., 2015). These new aspectual values require more manual cleaning, which is underway.

Secondly, this list was enriched with the aid of two resources: the consulting of thesauri (TLFi, 2012; Crisco, 2019) and an in-depth exploratory analysis of the medical corpus. At the end of this second step, we end up with a total of 142 verbs and verb phrases that can potentially express an aspectual nuance, as follows.

Thirdly, on this basis, extraction grammars were designed in the corpus analysis software Unitex (Paumier, 2016) and applied to the corpus. These grammars enable the creation of a concordance, an ‘in context’ exploration tool of the aspectual verbs.

Fourthly, using this concordance, we performed a manual vetting which applies three actions on the verbal occurrences: i) the deletion of verbs that never appear (e.g., *to burst into*) or never with an aspectual meaning (e.g., *to break out*) in our clinical language corpus; ii) the removal of non-aspectual verbal constructions (see (4)); iii) the detection of verbs whose aspectual value changes under the negative modality (see (5)).

(4) Par ailleurs, attaque son chir pour abdominoplastie trop serrée.

(5) A arrêté de fumer. > Terminates  
“Quit smoking.”

Malade n’arrête pas de tousser. > Continues  
“Sick person keeps coughing.”

Fifthly, we have undertaken a linguistic analysis of this filtered list (Table 2) to establish verb classes characterized by syntactic and semantic properties and to produce the resource described here.

<u>Aspectual type</u>	<u># verbs before</u>	<u># verbs after</u>	<u># constructions</u>
Initiates	59	27	34
Continues	18	16	23
Terminates	35	25	31
Culminates	23	16	23
Reinitiates	10	6	9
<b>Total</b>	<b>145</b>	<b>90</b>	<b>120</b>

Table 2. Number of constructions and verbs by aspectual type before and after automatic vetting

## 5 Resource

The linguistic resource presented in this paper consists of the description, in tabular form, of 90 verbs and verbal phrases with an aspectual value. After extraction and manual analysis, these 90 verbal expressions lead to 120 different constructions (see Table 2) which are then formally described. For example, the verb *commencer* comes in four forms: *commencer à* (to start doing something), *commencer par* (to start with), *commencer* (+ article) + noun (to start something),

*commencer* (to start). Each of these constructions possesses several syntactic and semantic properties. They will, therefore, fall into different descriptive tables.

The ten tables resulting from this study (see Appendix A) allow us to summarize the syntactic and semantic constraints linking some elements of the sentence, frequently medical events.

The structure of the table was determined based on the six following properties: i) aspectual nature; ii) passivization, iii) pronominalization of the complement, iv) number and type of complements; v) nature of the subject and vi) nature of the object. The first binarization of the 120 structures, based on aspectual nature, separates the structures which are always aspectual (6) from those which are only aspectual in certain instances, depending on the transitive or intransitive nature of the verb (7).

(6) Elle se met à lire des notices pharmaceutiques.  
“She starts reading pharmaceutical leaflets.”

(7) Le Sulpiride coupe les vomissements vs. J’ai coupé le fil résorbable.  
“Sulpiride stops the vomiting.” vs. “I cut the absorbable thread.”

This first binarization based on the aspectual nature criterion applies to the four tables (tables 2, 3, 5 and 6) that respond identically to classification criteria, but differ in their aspectual dimension.

Tests are then applied to the verbs, in the medical context of the corpus, with regard to certain syntactic and semantic properties. Performing these tests allows us to classify the verbs in the tables according to specific criteria. The tests are called *transformations* (Harris, 1957) and are applied to the simple form of the sentence, i.e., a positive assertion, in the active voice and reduced to the Subject-Verb-Object structure.

In terms of syntax, three properties are evaluated and define the assignment of a verbal structure in one or the other table:

- the existence (8) or the absence (9) of an equivalent form in the passive voice;
  - (8) L’examen est abandonné par le médecin.  
“The examination is given up by the doctor.”
  - (9) \*Le malade est prolongé par le médecin.  
\*“The patient(’s life) is prolonged by the doctor.”
- the possibility of pronominalizing the verb complement (10 and 11);
  - (10) Le traitement relance le processus d’hépatite colique. → Il le relance.  
“Treatment revives the process of biliary colic.”  
→ It revives it.
  - (11) Il a cessé toute consommation de bière. → \*Il l’a cessée.

“He ceases the whole intake of beer.” → \*He ceases it.

- the nature of the verb and its complements: verb with a direct object, verb with a prepositional complement, verb with a completive subordinate clause, stative verb and intransitive verb.

Concerning semantic properties, also referred to as *distributional* (Harris, 1954), we essentially examine the subject type (human or thing) and the object type, if relevant. When the subject or the object is an inanimate object, it often corresponds to a medical event. The selection of the type(s) of medical event accepted by the meaning of a verb is indicated in the column corresponding to one of the eleven possible events: administration, restriction or effort (diet, sports, profession, addiction), morphological abnormality, care and treatment, (positive or negative) effect, step, observable entity, findings, surgery, disorder and illness.

Besides, these tables also indicate the aspectual modification of the verb (+) or the permanence of the aspect (-) in the presence of the negative modality. These tables feature two additional columns: the meaning of the verb in this structure and the aspect taken by the structure. Finally, we indicate the presence of an adverbial phrase and its type (of time, of manner, or instrument), when it is recurrent for this meaning.

This resource is available upon request and will shortly be expanded. For an example of the table, see Appendix B.

## 6 Evaluation

To evaluate our resource, we estimated its coverage on 100 new medical texts that serve as a gold standard. These texts come from a different genre since they are 100 post hospitalization reports written by doctors. This different textual genre, still specific to the medical language, should allow a greater generalization of the resource, in that they convey a greater variety of linguistic phenomena.

In these texts, we have identified and annotated all the verbs, and only the verbs, that correspond to one of the five aspects of our typology. We calculated the coverage in terms of aspectual lemmata (Table 3), occurrences (Table 4) and occurrences by aspectual values (Table 5).

	<u>Absolute number</u>	<u>Coverage Rate</u>
Present	34	0,74
Missing	12	0,26
Total	46	1

Table 3. Coverage for aspectual lemmata

	<u>Absolute number</u>	<u>Coverage Rate</u>
Present	104	0,76
Missing	34	0,24
Total	138	1

Table 4. Coverage for occurrences

	<u>Init.</u>	<u>Cont.</u>	<u>Term.</u>	<u>Culm.</u>	<u>Rein.</u>
Present	8	11	8	3	5
Missing	2	5	3	1	0
Total	10	16	11	4	5
Coverage Rate	0,8	0,69	0,73	0,75	1

Table 5. Coverage for occurrences by aspectual values

Results are satisfactory, but probably insufficient for an industrial exploitation. We, therefore, want to integrate the cases identified as missing in the analysis and do this analysis again until reaching coverage of 0.9 by dragging it on a new corpus of 100 texts.

## 7 Conclusion

The development of such a linguistic resource, in which the medical language is formalized according to the lexical aspect of certain verbs, serves two purposes in the NLP field. On the one hand, these descriptions make it possible to distinguish the aspectual senses from the other meanings of a verb. They allow, among other things, to make a decision in case of verbal ambiguity. As a result, it is easier to include the aspect into the clinical patient’s history, which would then take into account some values that were previously omitted. On the other hand, with the semantic tagging of the agent and patient of these verbal structures, it becomes possible to more accurately identify medical events, be it in a vacuum or with regard to their localization and their evolution on the patient’s timeline.

We are now faced with two possible follow-ups. Firstly, an evaluation of the resource efficiency still needs to be done by integrating a temporal information extraction system to see how results can be increased and improved. Secondly, we need to extend the scope of the covered phenomena by listing specific aspectual and medical concepts such as the relapse, the chronicity or the worsening.

## Acknowledgments

We would like to express our special thanks and gratitude to our colleague T. François for his help and comments during the redaction of this paper. We would also thank G. Bardiaux, translator and reviewer for the language.

## References

- Edmond Bach. 1986. The Algebra of Events. *Linguistics and Philosophy*, 9:5-16.
- Collin F. Baker, Charles J. Fillmore and John Lowe. 1998. The Berkeley FrameNet project. In *COLINGACL '98: Proceedings of the Conference*. Montreal, Canada.
- Steven Bethard, Guergana Savova, Martha Palmer and James Pustejovsky. 2017. SemEval-2017 Task12: Clinical TempEval. In *Proceedings of the 11<sup>th</sup> International Workshop on Semantic Evaluation*. Vancouver, Canada, Aug. 3-4 2017. Association for Computational Linguistics, 1052-1062. <http://www.aclweb.org/anthology/S17-2093>.
- André Bittar. 2010. *Building a TimeBank for French: A Reference Corpus Annotated According to the ISO-TimeML*. Ph.D. Thesis, Prof. Laurence Danlos and Pascal Denis, Université Paris Diderot, Nov. 19 2006.
- Taxiarchis Botsis, Michael D. Nguyen, Emily J. Woo, Marianthi Markatou and Robert Ball. 2011. Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. *Journal of the American Medical Informatics Association*, 18(5):631-638.
- Leonardo Campillos, Louise Deléger, Cyril Grouin, Thierry Hamon, Anne-Laure Ligozat and Aurélie Névéol. 2018. A French Clinical Corpus with Comprehensive Semantic Annotations: Development of the Medical Entity and Relation LIMS annotated Text corpus (MERLOT). *Language Resources and Evaluation*, 52(2):571-601.
- Marie Candito, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël De Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathier, Philippe Muller, Benoît Sagot and Laure Vieu. 2014. Developing a French framenet: Methodology and first results. In *Proceedings of the 9<sup>th</sup> international conference on Language Resources and Evaluation (LREC 2014)*. Reykjavik, Iceland.
- Patrick Caudal. 2006. *Aspect*. In Danièle Godard, Laurent Roussarie and Francis Corblin (eds.). *Sémanticlopédie: dictionnaire de sémantique*, GDR Sémantique & Modélisation, CNRS, <http://www.semantique-gdr.net/dico/>.
- Kevin E. K. Chai, Stephen Anthony, Enrico Coiera and Farah Magrabi. 2013. Using statistical text classification to identify health information technology incidents. *Journal of the American Medical Informatics Association*, 20(5):980-985.
- Vincent Claveau, Natalia Grabar, Cyril Grouin and Thierry Hamin. 2019 (à par.). DEFT 2019. Défi Fouille de Textes. Recherche et extraction d'information dans des cas cliniques. TALN 2019, Toulouse, France.
- Bernard Comrie. 1981. *Language universals and linguistic typology: Syntax and morphology*. University of Chicago Press, Chicago.
- Jean-Paul Confais. 2002. *Temps, mode, aspect : les approches des morphèmes verbaux et leurs problèmes à l'exemple du français et de l'allemand*. Presses universitaires du Mirail, Toulouse.
- Laurence Danlos, Takuya Nakamura and Quentin Pradet. 2014. Vers la création d'un VerbNet du français. In *Atelier Fondamental*. TALN 2014, Marseille, France.
- Jean Dubois and Françoise Dubois-Charlier. 1997. *Les verbes français*. Larousse-Bordas, Paris.
- Jean Dubois, Jean-Baptiste Marcellesi, Jean-Pierre Mével and Mathée Giacomo. 1994. *Dictionnaire de linguistique et des sciences du langage*. Larousse, Paris.
- Noémie Elhadad, Sameer Pradhan, Sharon Lipsky Gorman, Suresh Manandhar, Wendy Chapman and Guergana Savova. 2015. SemEval-2015 Task 14: Analysis of Clinical Text. In *Proceedings of the 9<sup>th</sup> International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado, USA, June 4-5 2015. Association for Computational Linguistics, 303-310. <https://www.aclweb.org/anthology/S15-2051>.
- Laurent Gosselin. 2011. L'aspect de phase en français : le rôle des périphrases verbales. *Journal of French Language Studies*, 21(3):149-171.
- Maurice Gross. 1975. *Méthodes en syntaxe. Régime des constructions complétives*. Hermann, Paris.
- Zellig Harris. 1954. Distributional Structure. *Word*, 10(2-3):146-162.
- Zellig Harris. 1957. Co-occurrence and Transformation in Linguistic Structure. *Language*, 33(3.1):283-340.
- Paul Imbs (Ed.). (1971-1994). *Trésor de la langue française. Dictionnaire de la langue du XIXe et du XXe siècle (1789-1960)*. Paris : Gallimard. 16 vol. En ligne : <http://www.cnrtl.fr/definition>.
- Innoviris. 2014-2016. *iMediate: Interoperability of Medical Data through Information Extraction and*



- Term Encoding*. Project under the supervision of Prof. Cédric Fairon.
- ISO. 2008. *ISO DIS 24617-1:2008 Language Resource Management – Semantic Annotation Framework – Part 1: Time and Events*. International Organization for Standardization, ISO Central Secretariat, Geneva, Switzerland.
- Stanislaw Karolak. 2008. L’aspect dans une langue : le français. *Études cognitives*, 8:11-51.
- Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. Thesis, Prof. Martha S. Palmer, University of Pennsylvania.
- Judith Klavans and Martin Chodorow. 1992. Degrees of Stativity: The Lexical Representation of Verb Aspect. In *Proceedings of COLING-1992*. Nantes, France, Aug. 23-28 1992. Association for Computational Linguistics, 1126-1131. <http://aclweb.org/anthology/C92-4177>.
- Yordanka Kozareva-Levie. 2011. *L’aspect grammatical et ses manifestations dans les traductions en français de textes littéraires bulgares*. PhD Thesis, Prof. Jocelyne Fernandez-Vest, Université Paris 3, Feb. 7 2011.
- Min Li and Jon Patrick. 2012. Extracting Temporal Information from Electronic Patient Records. In *AMIA Annual Symposium Proceedings*, Chicago, USA, Nov. 3-7 2012. American Medical Informatics Association, 542-541. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540436/>.
- Yu-Kai Lin, Hsinchun Chen and Randall A. Brown. 2013. MedTime. A temporal information extraction system for clinical narratives. *Journal of Biomedical Informatics*, 46: S20-S28.
- Marc Moens and Mark Steedman. 1988. Temporal Ontology and Temporal Reference. *Computational Linguistics*, 14(2):15-28.
- Sébastien Paumier. 2016. *Unitex 3.1 User Manual*. Electronic version. <https://unitexgramlab.org/releases/3.1/man/Unitex-GramLab-3.1-usermanual-fr.pdf>.
- James Pustejovsky, José Castaño, Robert Ingria, Roser Sauri, Robert Gaizauskas, Andrea Setzer and Graham Katz. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Fifth International Workshop on Computational Semantics IWCS-5*, Tilburg, Netherlands, Jan. 15-17 2003. <http://www.timeml.org/publications/timeMLpubs/IWCS-v4.pdf>.
- James Pustejovsky and Marc Verhagen. 2009. SemEval-2010 Task 13: Evaluating Events, Time Expressions, and Temporal Relations (TempEval-2). In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*, Boulder, USA, June 4 2009. Association for Computational Linguistics, 112-116. <http://aclweb.org/anthology/W09-2418>.
- James Pustejovsky, Kiyong Lee, Harry Bunt and Laurent Romary. 2010. ISO-TimeML: An International Standard for Semantic Annotation. In *Proceedings of the 7<sup>th</sup> International Conference on Language Resources and Evaluation*, La Valletta, Malta, May 17-23 2010. European Language Resources Association, 394-397.
- Eric Siegel and Kathleen McKeown. 2001. Learning Methods to Combine Linguistic Indicators: Improving Aspectual Classification and Revealing Linguistic Insights. *Computational Linguistics*, 26:595-627.
- Carlota S. Smith. 1991. *The Parameter of Aspect*. Springer Netherlands, Amsterdam.
- Aleksandra Sojic, Walter Terkaj, Giorgia Contini and Marco Sacco. 2016. Modularising ontology and designing inference patterns to personalize health condition assessment: the case of obesity. *Journal of Biomedical Semantics*, 7(1):12.
- Jannik Strötgen and Michael Gertz. 2010. HeidelTime: High-quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5<sup>th</sup> International Workshop on Semantic Evaluation*, Uppsala, Sweden, July 11-16 2010. Association for Computational Linguistics, 321-324.
- William F. Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova and James Putstojovsky. 2014. Temporal Annotation in the Clinical Domain. *Transactions of the Association for Computational Linguistics*, 2:143-154.
- Weiyi Sun, Anna Rumshisky and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *Journal of the American Medical Informatics Association*, 20(5):806-813.
- Mike D Tapi Nzali, Xavier Tannier and Aurélie Névéol. 2015a. Automatic Extraction of Time Expressions Across Domains in French Narratives. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, Sep. 17-21 2015. Association for Computational Linguistics, 492-498. <http://aclweb.org/anthology/D15-1055>.
- Mike D Tapi Nzali, Xavier Tannier and Aurélie Névéol. 2015b. *Analyse d’expressions temporelles dans les dossiers électroniques patients*. In 22<sup>ème</sup> Conférence sur le Traitement Automatique des Langues Naturelles, Caen, France, June 22-25 2015. 144-152.

- Julien Tourille, Olivier Ferret, Xavier Tannier and Aurélie Névéol. 2017. LIMSI-COT at SemEval-2017 Task 12: Neural architecture for temporal information extraction from clinical narratives. In *Proceedings of the 11<sup>th</sup> International Workshop on Semantic Evaluation (SemEval-2017)*. Vancouver, Canada, Aug. 3-4 2017. Association for Computational Linguistics, 595-600. <http://www.aclweb.org/anthology/S17-2098>.
- Zeno Vendler. 1957. Verbs and Times. *The Philosophical Review*, 66(2):143-60.
- Marc Wilmet. 1980. Aspect grammatical, aspect sémantique, aspect lexical : un problème de limites. In Jean David and Robert Martin (eds.). *La notion d'aspect*. Actes du colloque organisé par le Centre d'analyse syntaxique de l'université de Metz. May 18-20. Klincksieck, Paris, 51-68.



## Appendices

### Appendix A. Synopsis of the ten tables and their features

<u>Table</u>	<u>Aspectual meaning</u>	<u>Features</u>
Table 1	Always	Aspectual auxiliary verbs; Infinitive subordinate clause introduced by a preposition and bearing the meaning of the process
Table 2.1	Always	Auxiliary <i>avoir</i> ; direct object; passivization and pronominalization
Table 2.2	Occasionally	
Table 3.1	Always	Auxiliary <i>avoir</i> ; direct object; no passivization and no pronominalization; human subject
Table 3.2	Occasionally	
Table 4	Always	Auxiliary <i>être</i> ; adjectival predicative complements
Table 5.1	Always	Prepositional complement; no passivization and no pronominalization
Table 5.2	Occasionally	
Table 6.1	Always	Intransitive; no passivization and no pronominalization
Table 6.2	Occasionally	

Appendix B. Example of one table: Table 2.1.

Sujet humain	Sujet chose	Verbe	Sens	Actif latre avoir	Actif latre être	Attribut	Passation	Pronominalisation	Négation	Complément	Circ constant	Exemples	Aspect
+	restriction ou effort									administratif			
+	restriction ou effort									restriction ou effort			
+	restriction ou effort									anormalité morphologique			
+	restriction ou effort									cure/traitement/médicament			
+	restriction ou effort									effet positif/négatif			
+	restriction ou effort									examen/esthéticien			
+	restriction ou effort									observations cliniques			
+	restriction ou effort									procédure/acte chirurgical			
+	restriction ou effort									troubles/maladie			
+	accomplir	mettre à exécution,		+								jusqu'à accomplir 10 jours de traitement.	Cubitatif
+	achever	finir, parachever		+							instrumental	instrumental qui est faite spontanément	Cubitatif
+	arrêter	stopper		+								Tabac : arrêté il y a 10 ans.	Terminatif
+	clôre	finir, achever		+								je suppose que l'affaire est close.	Cubitatif
+	clôturer	finir, achever		+							instrumental	à revoir pour clôturer le traitement.	Cubitatif
+	commencer	débuter		+								elle commence la phase de sucreries.	Initiatif
+	finir	achever, terminer		+							instrumental	a fini son traitement	Cubitatif
+	freiner	contenir, retenir		+								une couche de fluorine freinant l'évolution.	Terminatif
+	inaugurer	commencer		+								gastro entériente inaugurant les plaintes.	Initiatif
+	incuber	couver		+								On incube une maladie	Initiatif
+	initier	provoquer le début		+								Un suivi urologique a été initié.	Initiatif
+	juguler	arrêter		+								l'antibiothérapie a jugulé le problème	Terminatif
+	mettre en route	mettre en marche, d+		+								Je mets en route ces examens	Initiatif
+	stopper	arrêter, suspendre		+								elle stoppera d'ici 1 semaine ce traitement.	Terminatif
+	suspendre	arrêter, interrompre		+								On suspend l'examen.	Terminatif
+	faire	amener, dissiper		+								Enterol pour faire les suites de l'entérite.	Terminatif
+	terminer	achever, finir		+								Il terminait un traitement antibiotique.	Terminatif