Temporal hints in the cultural heritage discourse: what can an ontology of time *as it is worded* reveal?

Gamze Saygi, Jean-Yves Blaise, Iwona Dudek

UMR CNRS/MCC 3495 Modeles et Simulations pour l'Architecture et le Patrimoine
Campus CNRS Joseph Aiguier - Bat. Z',
31 chemin Joseph Aiguier 13402, Marseille Cedex 20, France.
(gamze.saygi, jean-yves.blaise, iwona.dudek)@map.cnrs.fr
http://map.cnrs.fr

Abstract. Time is an indispensable component of CH information: implementing appropriate knowledge models carry crucial importance in order to provide deeper understanding of heritage elements' evolution, to uncover concurrences, and to weigh *quality* factors. It is a challenging task though due to the uncertain characteristics of temporal data, and to the *wording of time* in the CH discourse. Existing KR models are either not designed for these distinctive characteristics, or spatial aspects tend to upstage the temporal dimension. This research aims at deciphering and proposing a formal representation of the way temporal hints are formalized in historical narratives. An OWL ontology is introduced that provides a core support mechanism allowing for a semantic representation of temporal statements, and for structural analysis. The objective is to facilitate the cross-examination of temporal hints in and across CH collections so that specialists can have extensive *reading* possibilities of heritage information.

1 Introduction

As pointed out by Jurisica et al. (2004), with more and more computer-readable pieces of information, analysts today need to rethink their knowledge extraction strategies. Ontologies offer significant capabilities for knowledge management, especially in large volumes of information (Davies et al., 2003) by providing controlled and consistent vocabularies defined as a set of representational primitives (information types, their properties and relationships) coherent with the meanings and constraints in a domain of knowledge (Gruber, 1993).

On the other hand, *time* is a feature that appears in many pieces of information (Faucher et al., 2010), and ontologies of time can be of concern for various disciplines. In this research, we focus on the concept of time in the *cultural heritage* (CH) discourse: temporal aspects are there an inseparable and central role-player for historical analysis, and in any reasoning task performed on the evolution, transformation, reuse, status of heritage assets. But in the CH discourse past events or facts anchored in time in a large variety of forms (e.g., [...] it goes back to the second half of the 13^{th} or 14^{th} Century, [...] after the Revolution, etc.). These wordings do not fit into "classical" quantification systems such as date formats in DB management systems (e.g., "1942-03-19").

In order to reach a more accurate representation of the temporal dimension of historic narratives, it is important to analyze with care the actual wording of the temporal hints. Only once this analysis has been carried out and confronted to the reality of the historical data sets, can one expect to build a generic formal model, providing interoperable means to decipher and represent temporal structures.

The idea behind the research is that, as analysts of historical evidence, and prior to any interpretation steps, we first need to understand and depict in a structured and sharable manner the nature of the data we handle, in particular of temporal statements.

Such a formalization targets reasoning tasks on verbalization patterns: correlating them to types of information providers, to historical periods they concern, to particular geographic or cultural areas, to particular authors, etc. The background objective of this research is to do a comprehensive analysis of the temporal hints in the CH discourse. To do so, we introduce a hands-on application scenario, encompassing the extraction of "real" hints from "real" data and an experimental implementation using the OWL (Web Ontology Language) / Protégé technological suite (Musen, 2015). Such an ontology not only helps reaching a formal model of time *as it is worded* but also renews the way heritage specialists extract different interpretations from the data they handle.

We however make no claim that we have developed a generic ontological framework that would be suited to historic sciences at large. We do acknowledge the fact that temporalities, and the way they are worded, can be dependent on distinctive parameters such as region, typology or collection. Nevertheless, we highlight particular tendencies of *wording* in the CH discourse and focus on typical challenges while acquiring and formalizing temporal information in the domain. In Section 2, we draw the outline of the research context in a twofold manner; first by outlining how the time concept is handled in the general research context, and second by focusing on what has been experimented so far in the specific context of the CH domain. Section 3 introduces our approach to the representation of time as it is worded: concepts, notions, and their interrelations. Section 4 discusses the technical implementation and its experimental evaluation. In the final section, we list outcomes and shortcomings at this stage of the research, and some lessons learnt from the experiment.

2 State-of-the-art

Useful standards, definitions, specifications and recommendations already exist aiming at diverse concerns and steps of temporal information processing. Time-oriented data analysis is a concern within many research communities, such as the TIME community that deals with temporal theories, logics, representation languages, reasoning and ontologies (Ermolayev et al., 2014). ISO 8601 describes a standardized way of presenting dates and times, whereas ISO 19108 sets the information technology standards for the interchanging of temporal information. ISO TimeML targets a very crucial concern; Natural Language Processing (NLP) tasks for creating controlled temporal expressions from unstructured text. It does not only take into account quantities but also relevant semantic operators. OWL-Time, which is a candidate W3C recommendation, aims at providing a vocabulary for expressing facts about topological relations among temporal instants and intervals. It has been recently extended (Cox, 2016) to support the encoding of temporal reference systems other than the Gregorian calendar. There is a growing interest on extending regular time concept to a wider non-absolute perspective, i.e., research dealing with dirty datasets. For instance, Tao et al. (2010) develop an ontology called CNTRO for representing temporal information in clinical narratives as RDF (Resource Description Framework) triples supporting time oriented queries in semantic web. Anagnostopoulos et al. (2013) draw attention to the frequency of qualitative expressions in temporal expressions (e.g., before, after), and develop a reasoner named CHRONOS for uncovering temporal relations. Golden and Shaw (2016) ease the task of linking among datasets that define temporal periods differently. Poveda-Villalón et al. (2014) highlight the importance of integrating recurrent events, whereas Diallo et al. (2015) consider different granularities in addition to recurrences. Faucher et al. (2010) experiment a pipeline from bottom to top, i.e., for acquiring temporal knowledge from texts in order to populate a constrained computable model.

Nevertheless, uncertainty, vagueness and imprecision in the wording of temporal hints (e.g., late 1980s, end of November) remain tough to represent formally. Besides, none of the above mentioned research works yet take into account contradictory notations or alternative wordings that can be combined, even in a single source (e.g., in the 14^{th} Century, probably around 1380), although such ways of saying are common in the CH discourse.

In the CH domain, the CIDOC CRM, also known as ISO 21127, is a core ontological model aiming at creating semantic glue between different sources of information, such as that published by museums, libraries and archives. In CRM model spatiality and temporality go hand in hand. Particular to CRM based experimentations, some researchers represented temporal periods on 4 dimensional volumes (Papadakis et al., 2014), being associated with spatio-temporalities (Hiebel et al., 2016). Approximate, definite and indefinite bounds of periods are considered. Although, they state the approach could be embedded in an information system such as GIS, there is no solid test-case, yet. Moreover, Papadakis et al. (2014) point out that their model does not allow representing periods which retreat to the same place several times (i.e., recurrent events) or occur at disjoint places (e.g., festive events). This is a quite critical issue in CH, considering for example sets of transformations/additions/extensions shaping a heritage artefact's lifespan. They focus on modeling reality using only material evidence about past periods or events derived from the observation of traces. However although time may leave physical traces on tangible heritage assets it does not always do so, in particular in the case of intangible heritage (e.g., practices, traditions, festive events). Binding (2010) adopt CRM entities and properties for controlled vocabularies, and demonstrate a temporal reasoning method for modelling temporal relationships for archaeological records. They make use of conventional agreements for temporal subdivisions. This allows aligning data records with known time periods and representing the approximate lower and upper bounds of the time periods with numeric values. For instance, they split centuries into years of 01-32 if indicated early, 33-66 for mid or 67-100 for late with reference to advice received from English Heritage. As shown in Section 3.1 our contribution reuses this concept of *conventional mapping (agreements)*, but extends it in terms of granularities, and proposes a level of flexibility that allows for a user-chosen or user-specific *conventional mapping*.

Kauppinen et al. (2010) stress the imprecision of temporal information in the CH discourse. They deal with fuzzy boundaries and exact boundaries of time intervals, and formalize each temporal interval by constraining the earliest and latest possible start and end dates. Nevertheless, they do not take into account open-end indications like before and after statements. Besides they solely analyze the potential overlapping of intervals. Nurminen and Heimburger (2012) discuss representation and retrieval of uncertain temporal information in museum databases with a specific focus on anchored time intervals. They underline the shift from item-centric (i.e., structured around physical entities) to event-centric cataloguing (i.e., concentrating on various events) of museum artefacts. It is particularly important when considering events as clusters linking heritage entities to the cultural actions of human beings and the social setting.

Generally speaking a large number of research efforts on temporal reasoning build on the concepts introduced by Allen (1983, 1991), and there is no doubt that such efforts are welcome for instance in the context of artificial intelligence applications. But there is quite a distance between the way historical hints are actually worded, and a sound application of Allen's formal relations. Our research fits in that gap: temporal statements (both quantitative values and lexical modifiers) first need to undergo a process of extraction in order to uncover their structure, and a fine-grain analysis so that scientists can get a critical understanding of how doubts pervade their reasoning processes. This key challenge in historical sciences was taken on in the context of heritage data cross-examination and visualization Blaise et al. (2016) but an in-depth investigation of how it can be addressed in the KR context remains to be carried out. Efficient temporal reasoning, based on Allen's relations or not, will only be possible if a complete formal model of the specific wordings of historians has been introduced. In order to fulfill this aspect, a basic approach could be following the footsteps of a standard temporal ontology (such as OWL-Time) by extending it or reusing the spatio-temporal ontology by CIDOC CRM. Nevertheless, the way time is verbalized in the historical discourse would require a very significant move away from these standards' original versions. The approach presented in this paper is not to work on the concept of time itself, or on spatio-temporal entities, but to try and assess the potential added-value of a formal representation of temporal statements *as worded* in the CH discourse.

3 Analysis and representation of temporal statements

In this section we discuss the formal model of temporal statements. To start with we define the main notions, the top-level organization and relations, and detail the time-related concepts classification.

3.1 Main notions and general organization of the formal model

A *TemporalStatement* is some sequence of words that tell us when something hap-pened and/or how long something lasted within a temporal reference. A *TemporalStatement*'s semantic field is the notion of time alone, it does not extend to the concept of spatial/physical property which it affects or refers to. Any *TemporalStatement* is composed of one or more time-related concepts (points, intervals, etc.), which may be accompanied by one or more *LexicalOperator*s (during, before, end of, etc.).

There is a set of basic but important elements we associate to a temporal statement (figure 1). First and foremost, granularity describes the mapping of time into conventional units upon human decision. That mapping is basically aimed at dealing with time in an easier way, and can be specified in multiple ways (in larger or smaller units) depending on the needs of the analyst (Aigner et al., 2011). In a *TemporalStatemen*t like "The first quotation dates back to 20 April 1687", the temporal granularity can be fixed to "a day". But the day can be subdivided into smaller segments such as 24 hours or 1440 minutes (and incidentally the “standardization” of time dates back to 1884, with the definition of an “average” hour in response to the multiplicity and variability of solar hours, hence serious interpretation bottlenecks when observing such statements today). The non-decomposable unit for a given granularity is called *chronon*, a term coined by Lévi (1927). For instance, in Java, date class uses milliseconds as *chronon*. Naturally in the historical discourse there can be a disjunction between a statement’s intrinsic granularity and the temporality of the fact that is reported. For instance, in a temporal statement like “A *great fire* damaged the building extensively *in the winter of 1920*”, the granularity of the statement is a season whereas the “great fire” mentioned in the statement most probably lasted some days at most or even some hours. The notion of *UnfoldableTimePoint* (see Section 3.1.3) is a pragmatic answer to that concern.

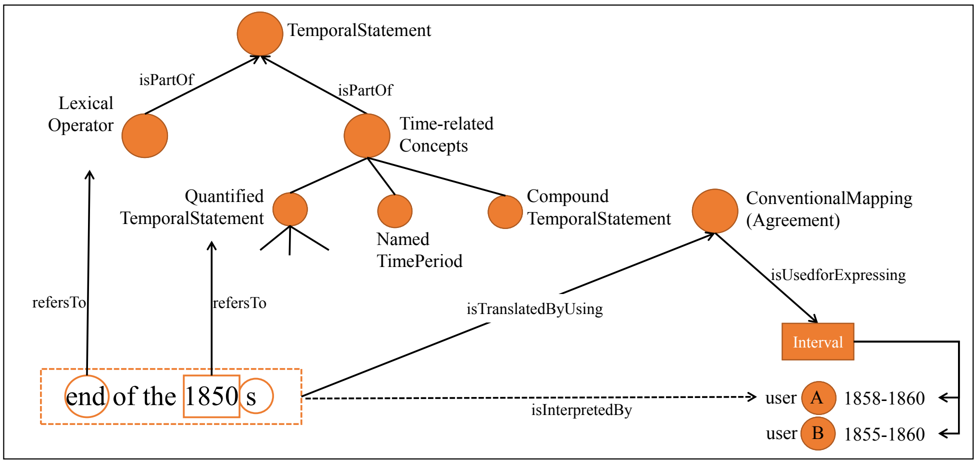


FIG. 1: Components of a *TemporalStatement* and their interrelations.

A difference is made between Temporal Reference Systems (calendars, part of OWL-Time) and the notion of periodization in history, that can be used as a reference in the wording of temporal hints but that is not a systematic discretization of time (e.g., *Gothic* appears before *Renaissance* in European architectural styles, nevertheless we can not specify the exact temporal boundaries of these trends). This notion is mapped in a concept called *NamedTimePeriod*.

Consequently, the proposed model is organized as follows:

— A *LexicalOperator* concept matching the “verbal modifiers” (around, before, etc.),

— A utility concept called *ConventionalMapping* used to interpret qualitative expressions when needed, and turn them into workable quantities,

— Classes that represent the time-related concepts present in the *TemporalStatements*.

3.1.1 LexicalOperator

Lexicals refining the qualitative extent of time and supporting the anchoring of *TemporalStatements* are defined as *LexicalOperators* (table 1). These qualitative components of the hint determine the “extent” of a *TemporalStatement*.

Type of usage	Defines quantity by	Examples
ExplicitOperator covers the words that do not impact the worded quantity.	Pointing at	in || on || at
	Bounding	from-to || between-end
	Defining frequency	every || annually
ImpreciseOperator enlarges or extends the worded quantity.	Anchoring at a median point	around || about || towards
	Bounding at one direction	before || after || since
RefinementOperator narrows the extent of the *TemporalStatement* by creating a subdivision.	Ordering Subdivisions	early || late beginning of || end of middle of
	Natural Subdivisions	seasons || tides

TAB. 1: Types of *LexicalOperator*s acting as potential components of *TemporalStatement*s.

3.1.2 ConventionalMapping

Prior to any conversion of a temporal statement containing a lexical operator into a workable quantity, an ad-hoc convention should be established: this is the role played by *ConventionalMapping* (agreements). The quantification of any *LexicalOperator* is subject to *ConventionalMapping* (CM). For instance, "end of the 15th century" can be translated into a user-chosen quantified time slot as "1480-1500", and the analyst needs to formalize in a sustainable manner such a decision.

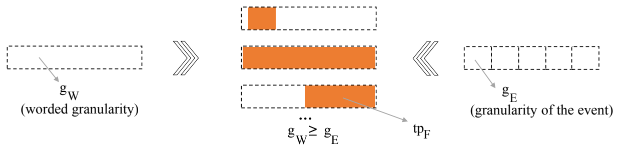
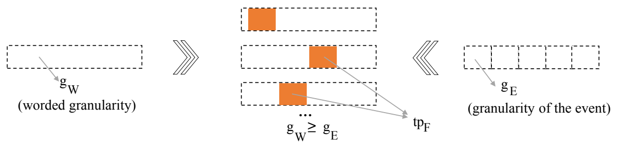
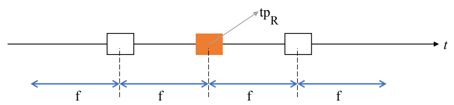
3.1.3 Time-related Concepts

Time-related concepts are represented through three major classes / hierarchy of classes: *NamedTimePeriod*, *QuantifiedTemporalStatement*, *CompoundTemporalStatement*. In this section, we address their definitions, properties and specifications.

NamedTimePeriod. *NamedTimePeriod*s correspond to *TemporalStatement*s that provide an ordered reference rather than a temporal coordinate system (e.g. *after the reign of* rather than *between X and Y*). The sequence of *NamedTimePeriod*s can overlap on each other. It can be related to diverse frame of references such as art movements (e.g., Art Nouveau), political events (e.g., 30 Year Wars) or natural facts. A *NamedTimePeriod*, whether it is accompanied with a *LexicalOperator* or not, implies the use of CM. The concept matches the notion of periodization in historical analyses.

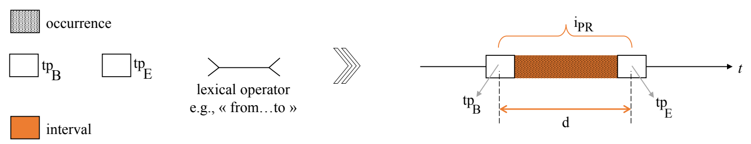
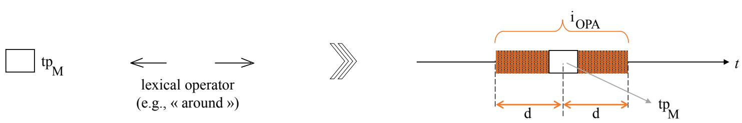
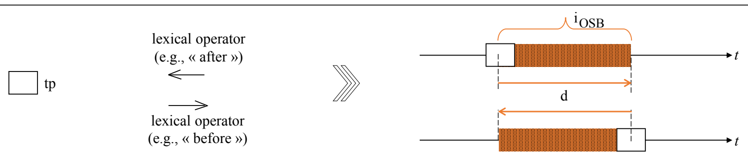
QuantifiedTemporalStatement. A *QuantifiedTemporalStatement* is a temporal hint expressed in numbers (or with universally accepted lexicals such as a decade). It can represent either a *TimePoint*, an *Interval* or an *UnanchoredDuration*. A *QuantifiedTemporalStatement* has two properties: a temporal reference system, and a *chronon*. Although the *QuantifiedTemporalStatement* is the quantified part of a hint, its value should still be interpreted in relation with any *LexicalOperator* present in the statement, e.g., "before 1650" does not point out to "the year of 1650" itself, but to a "time slot that precedes 1650".

TimePoint: definition and subclasses. At the conceptual level a time point represents an instant, with a *zero* length. Depending on the granularity, a *TimePoint* de facto can have a temporal extent. Three concepts refining the classic *TimePoint* (tp) class are introduced to deal with temporal statements that are verbalized as *TimePoint*s (table 2).

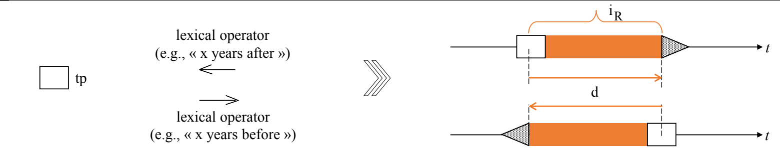
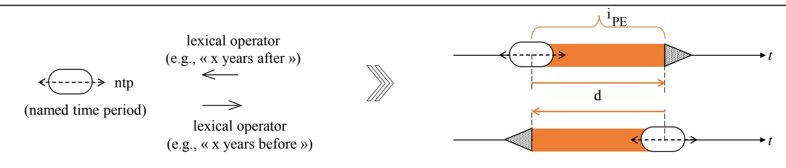
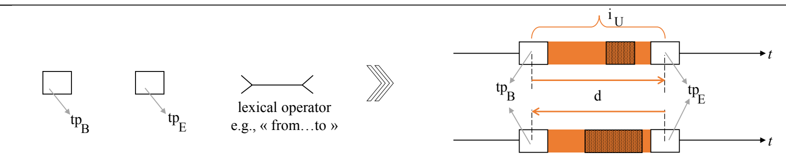
<table>
<tr><td colspan="2">FuzzyTemporalStatement (tp_F) is a wording that refers to an event confined in an ambiguous way to a TimePoint. The fuzziness of the information delivered by such a TemporalStatement comes from a doubt concerning the alignment of the granularity possibly needed to analyze the event and this of the TimePoint.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>Elle a été reconstruite au milieu du 18e siècle.
The edifice was built between 12^{th} and 13^{th} Centuries.</td></tr>
<tr><td>Extraction</td><td>[tp_F] = middle of the 18^{th} century and [g_W]= a century $\geq$ [g_E]= a year</td></tr>
<tr><td colspan="2">UnfoldableTemporalStatement (tp_U) is a wording that refers to an event expressed as a TimePoint. Unlike in the case of FuzzyTemporalStatement the duration of the event is here for sure shorter than the chronon corresponding to the TimePoint.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>La toiture a été emportée par l'avalanche de 1978.
The roof was washed away by the avalanche of 1978.</td></tr>
<tr><td>Extraction</td><td>[tp_U] = the avalanche of 1978 and [g_W]= a year > [g_E]= a minute</td></tr>
<tr><td colspan="2">RecurrentTemporalStatement (tp_R) is dedicated to TemporalStatements for periodic/cyclic occurrences. Its structure is as same as the TimePoint but with frequency descriptors (f).</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>Aujourd'hui, Notre-Dame-de-Vie fait toujours [...] annuel le 15 août.
Today, Notre-Dame-de-Vie is still [...] an annual pilgrimage on August 15^{th}.</td></tr>
<tr><td>Extraction</td><td>[tp_R] = August 15^{th} and [f] = annual</td></tr>
</table>

TAB. 2: Subclasses of *TimePoint*.

Interval: definition and subclasses. At the conceptual level a time interval represents a segment of time with a duration. Six concepts are introduced to deal with temporal statements that are verbalized as time intervals (table 3). They share common features: they are defined by one or two *TimePoint*s and are anchored in time.

<table>
<tr><td colspan="2">ProperIntervalStatement (i_{PR}) is a wording corresponding to an Interval expressed with 2 TimePoints (beginning boundary: tp_B, and end boundary: tp_E). These two TimePoints are connected by an ExplicitOperator.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>[...] Bâti en pierre entre 1815 et 1825.
[...] Built in stone between 1815 and 1825.</td></tr>
<tr><td>Extraction</td><td>[tp_B] = 1815 and [tp_E] = 1825 implies d = ([tp_E]- [tp_B])+ 1 = 11 years</td></tr>
<tr><td colspan="2">OnePointAnchoredIntervalStatement (i_{OPA}) is a wording that refers to an occurrence expressed by a TimePoint, and accompanied by a LexicalOperator such as “around”. The LexicalOperator’s effect is to create two equal durations (d) on both sides of the point. The value of the durations is user-chosen (ConventionalMapping).</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>Bâtiment actuel édifié autour de 1645.
Current building is built around 1645.</td></tr>
<tr><td>Extraction</td><td>[tp_M] = 1815 and [d] is related to CM</td></tr>
<tr><td colspan="2">OneSideBoundedIntervalStatement (i_{OSB}) is a wording that refers to an occurrence expressed by a TimePoint (tp), and accompanied by a LexicalOperator such as “before” or “after”. The TimePoint acts as a boundary (beginning or end) and the LexicalOperator’s effect is to create an undefined duration on one side. A conventional value for the duration can be defined by ConventionalMapping.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>Elle a été construite après 1720 par souscription publique.
It was built after 1720 by public subscription.</td></tr>
<tr><td>Extraction</td><td>[tp] = 1720 and [d] is related to CM</td></tr>
</table>

(Continued on next page.)

<table>
<tr><td colspan="2">RelativeIntervalStatement (i_R) is a wording where a TimePoint is used to anchor an occurrence located at a distance in the past or in the future of the TimePoint. The "temporal gap" between the TimePoint and the occurrence is expressed, but the duration of the occurrence is not.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>en 1706: la chapelle ste Anne [...], bîie depuis plus de 60 ans.
in 1706: the chapel ste Anne [...], built more than 60 years ago.</td></tr>
<tr><td>Extraction</td><td>[tp] = 1706 and [d] = more than 60 years</td></tr>
<tr><td colspan="2">PertinentIntervalStatement (i_{PE}) is a wording where one only boundary is defined, and by an instance of NamedTimePeriod ("WWII", "the great plague", etc.). The concept is used to refine i_{OSB} and i_R when the anchoring cannot be done through a TimePoint, but is done through an instance of NamedTimePeriod.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>[...] remaniements après la Révolution.
[...] rearrengements after the Revolution.</td></tr>
<tr><td>Extraction</td><td>[ntp] = the Revolution and ([d] and [ntp]) are related to CM</td></tr>
<tr><td colspan="2">UnfoldableIntervalStatementt (i_U) is a wording expressed with two TimePoints, and where the duration of the occurrence is for sure shorter than the distance separating the two TimePoints.</td></tr>
<tr><td colspan="2">
</td></tr>
<tr><td>Example</td><td>1160-1164 [...] seigneurs [...] pour y construire une abbaye acte signé [...]
1160-1164 lords [...] to build an abbey [...] deed signed [...]</td></tr>
<tr><td>Extraction</td><td>[tp_B] = 1160 and [tp_E] = 1164 implies [d] = 5 years > [d_{EVENT}]</td></tr>
</table>

TAB. 3: Subclasses of *Interval*.

UnanchoredDuration. An *UnanchoredDuration* represents any temporal statement that mentions a segment of time with a duration but without explicit boundaries (anchored in time). This concept is used to represent hints such as "the renovation lasted for 40 days."

CompoundTemporalStatement. It is quite common to have multiple temporal indications within one *TemporalStatement*, e.g., “it dates back to the 13th or 14th century”. *CompoundTemporalEntity* represents such cases in which the statement contains two or more alternative *QuantifiedTemporalEntities*, may they overlap or not, and may they be consistent in terms of wording or not.

4 OWL Implementation, results and limitations

We implemented an open-source solution, by using OWL 2, built upon the RDF standard in which data is represented by sets of “triples”. We populated the ontology with information extracted from online resources published by different stakeholders (e.g., public archives, cultural societies, etc.) harvested in the context of a research initiative [1] on minor tangible and intangible heritage. These information sets, being verbalized by various parties, correspond to the heterogonous, imprecise and uncertain nature of the temporal information often met in the CH domain. We selected 1576 statements and in addition to standard OWL data types, we reused OWL-time datatypes such as time:generalDay, time:generalMonth.

At this still early stage of the research, the formalization effort does shed light on some significant verbalization patterns, but also on where the effort needs to be consolidated. For instance, 90 *TemporalStatement*s correspond to the concept of *CompoundTemporalStatement* but their classification as such is motivated either by contradictory indications “ *[...]datées du XIème ou du XIIème siècle*”) or by a sort of caution in the wording (“ *[...] doit remonter au XVIIe siècle (vers 1668-1670)*”). Incidentally, the use of wordings classified as *CompoundTemporalStatement* apparently gets more common when occurrences are located farther in the past, which is a rather expected pattern. A majority of the overall statements considered fall into one class: the *FuzzyTemporalStatement* class. This also is rather expected since the granularity alignment between a hint and an occurrence is quite often out of reach in Historical Sciences, and in particular when dealing with Minor Heritage. Some more question-opening tendencies can be observed also, for instance when observing the variety of linguistic figures used to verbalize one same situation: bounding the “beginning” of an occurrence. But it has to be said clearly that tendencies observed could very well be related to the particular dataset we have worked on, or to biases introduced in the model itself, and therefore they should be taken for what they are worth. Our purpose is certainly not to draw out of such early observations a Historical-Sciences related general conclusion but by contrast to use the observations as a mean to question the model. In addition, one of the lessons learnt from the experimentation was that some “contemporary” hints as found in citizen-birthed e-sources harvested on the net, such as “the edifice was reconstructed in last years”, are far from being the easiest to deal with.

Briefly said, our formalization efforts do highlight the extreme diversity of temporal information in the CH discourse, but also uncover some significant patterns such as a relative disconnection between lexical operators and temporal granularity (operators used whatever granularity). As a consequence, it appears clearly that what the research also underlines is the amount of *unsaid* in the reasons why this or that verbalization modality has been chosen by an information provider.

1. The ontology is available on the Territo*graphie* project web site (http://map.cnrs.fr/territographie/), portal of an exploratory research project on citizen science and minor heritage conducted in co-operation with MuCEM (Museum of European and Mediterranean Civilizations) and funded by the région Provence-Alpes-Côte d’Azur authorities.

5 Conclusions and Future Works

In this paper, we analyzed imperfect temporal statements used in the CH discourse, and introduced an ontology for enabling and encoding temporal knowledge. Our contribution, small though it may be, aims at facilitating the cross-examination by analysts of historical evidence of *the temporal evidence* itself, prior to interpretation steps. Our experimental results show that the ontology carries the potential to shed light on (ill-defined) temporal information effectively. Nevertheless, we acknowledge the necessity to validate the ontology on larger datasets, and test its extraction capability in heritage collections where regional and typological parameters are extensively circumscribed.

In addition, we acknowledge the necessity to inspect the proposed ontology by external domain experts. However the services expected, as well as the information providers, correspond to a variety of domains ranging from linguistics ("wording as such") to museology, ethnology, art history (if not historical sciences at large), etc. Hence a robust validation helping to recalibrate the proposition will require a multidisciplinary investigation, and is definitely part of the research agenda we have ahead of us. Future works will include a critical analysis of the applicability of the approach beyond the initial corpus (rules of assignment in particular), a visualization effort, and a deeper attention to the interdisciplinary issue of elicitation: not only how a temporal statement is worded but also why.

References

Aigner, W., S. Miksch, H. Schumann, and C. Tominski (2011). *Visualization of time-oriented data*. Springer Science & Business Media.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Commun. ACM 26*(11).

Allen, J. F. (1991). Time and time again: The many ways to represent time. *International Journal of Intelligent Systems 6*(4), 341–355.

Anagnostopoulos, E., S. Batsakis, and E. G. Petrakis (2013). Chronos: A reasoning engine for qualitative temporal information in owl. *Procedia Computer Science 22*(Supplement C), 70 – 77. 17th Int. Conf. in Knowledge Based and Intelligent Inf. and Eng. Systems.

Binding, C. (2010). *Implementing Archaeological Time Periods Using CIDOC CRM and SKOS*, pp. 273–287. Berlin, Heidelberg: Springer.

Blaise, J.-Y., I. Dudek, W. Komorowski, and T. Weclawowicz (2016). *Architectural transformations on the Market Square in Krakow - A systematic visual catalogue*. Krakow: Oficyna Wydawnicza AFM.

Cox, S. J. D. (2016). Time Ontology Extended for Non-Gregorian Calendar Applications. *Semantic Web Journal 7(2)*, 201 – 209.

Davies, J., D. Fensel, and F. Van Harmelen (2003). *Towards the semantic web: ontology-driven knowledge management*. John Wiley & Sons.

Diallo, P. F., O. Corby, I. Mirbel, M. Lo, and S. M. Ndiaye (2015). HuTO: an Human Time Ontology for Semantic Web Applications. In *Ingénierie des Connaissances 2015*, Rennes.

Ermolayev, V., S. Batsakis, N. Keberle, O. Tatarintseva, and G. Antoniou (2014). Ontologies of time: Review and trends. *IJCSA 11*(3), 57–115.

Faucher, C., C. Teissèdre, J.-Y. Lafaye, and F. Bertrand (2010). *Temporal Knowledge Acquisition and Modeling*, pp. 371–380. Berlin, Heidelberg: Springer Berlin Heidelberg.

Golden, P. and R. Shaw (2016). Nanopublication beyond the sciences: the periodo period gazetteer. *PeerJ Comp Science 2*, e44.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition 5*(2), 199–220.

Hiebel, G., M. Doerr, and Ø. Eide (2016). Crmgeo: A spatiotemporal extension of cidoc-crm. *Int Journal on Digital Libraries*, 1–9.

Jurisica, I., J. Mylopoulos, and E. Yu (2004). Ontologies for knowledge management: an information systems perspective. *Knowledge and Information systems 6*(4), 380–401.

Kauppinen, T., G. Mantegari, P. Paakkarinen, H. Kuittinen, E. Hyvönen, and S. Bandini (2010). Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *Int. J. Hum.-Comput. Stud. 68*(9), 549–560.

Lévi, R. (1927). Théorie de l'action universelle et discontinue. *J. Phys. Radium 8*(4), 182–198.

Musen, M. A. (2015). The protégé project: a look back and a look forward. *AI matters 1*(4).

Nurminen, M. and A. Heimburger (2012). Representation and retrieval of uncertain temporal information in museum databases. In *Information Modelling and Knowledge Bases XXIII*.

Papadakis, M., M. Doerr, and D. Plexousakis (2014). Fuzzy times on space-time volumes. In *eChallenges e-2014, 2014 Conference*, pp. 1–11. IEEE.

Poveda-Villalón, M., M. C. Suárez-Figueroa, and A. Gómez-Pérez (2014). A pattern for periodic intervals. *Semantic Web Journal Oct14*, 1–10.

Tao, C., W.-Q. Wei, H. R. Solbrig, G. Savova, and C. G. Chute (2010). Cntro: a semantic web ontology for temporal relation inferencing in clinical narratives. In *AMIA annual symposium proceedings*, Volume 2010, pp. 787. American Medical Informatics Association.

Résumé

Dans le champ des sciences patrimoniales, la dimension temporelle de l'information joue un rôle à l'évidence majeur tant pour l'interpréter et l'analyser que pour relier des faits isolés. Mais la façon dont cette dimension est verbalisée pose des problèmes de formalisation non triviaux. Pourtant, cette verbalisation, que l'on associe souvent au terme-chapeau d'incertitude, peut être lue en dissociant d'une part le caractère *mal connu* d'un fait documenté, irréductible, et les choix faits par le producteur de l'information pour la *relativiser*. Dans cette contribution nous proposons un modèle formel permettant d'observer et d'analyser de façon systématique cette couche de verbalisation. L'expérience est menée sur des données fortement hétérogènes, souvent d'origine citoyenne, documentant le petit patrimoine matériel et immatériel. Ce cas d'étude est donc limité, mais il apparait néanmoins comme portant une question de fond allant au-delà du cas d'espèce. La contribution détaille d'abord la grille d'analyse d'indices temporels proposée, puis relate l'expérimentation concrète associée (ontologie OWL). Il n'est pas fait état d'une quelconque prétention à un résultat généralisable stricto sensu, mais cette expérience peut contribuer à nourrir de façcon pragmatique un débat nécessaire sur la formalisation d'indices temporels dans les sciences historiques.