Chapter 7

HISTORICAL ONTOLOGIES

Nancy Ide and David Woolner
Vassar College; Marist College

Abstract: Static ontologies cannot capture the relevant contextual knowledge required for search and retrieval of historical documents because the entities in the world and the relations among them change over time. This demands that information represented in the ontology is temporally contextualized and that relations among entities that are relevant during different temporal intervals are available to support user queries. Furthermore, it is necessary to account for the fact that the course of the ontology's evolution and the processes that have effected it are a part of the knowledge that should be brought to bear on the analysis of information at any given time. This chapter outlines a model for historical ontologies that is intended to meet these requirements.

1. INTRODUCTION

Ontologies have played a role in natural language processing (NLP) since the hey-days of symbolic AI and NLP in the 1970's, given that language understanding was assumed, particularly at that time, to require complex information about concepts and the relations among them. Since then, ontologies (and their lesser sibling, the taxonomy) have been incorporated into various language processing applications. However, as McGuinness (2003) and many other authors have pointed out, ontologies have recently become the center of attention of a much broader community, due to the pivotal role they play in the vision of the Semantic Web (Berners-Lee et al. 2001, Fensel et al. 2003). As a result, there has been an unprecedented flurry of activity in the past few years focused on construction of ontologies and the development of tools to edit and reason with them, as well as standards for their representation and access via the World Wide Web.

Most current work involves the rapid construction of ontologies for specific domains by a small team of "ontology engineers", who typically identify the relevant concepts and relations *a priori* and then impose it on the system. The resulting ontologies are typically static, representing a set of objects/concepts, relations, and properties that remain constant. However, in some cases, the domain modeled by an ontology changes over time, and, depending on the time perspective, a different ontology is relevant for retrieval. For example, in order to retrieve all relevant information for a query about Germany, it may be necessary to recognize that instances of East Germany and West Germany in an ontology representing geo-political entities prior to 1989 are each a part of a single entity, Germany, in a later ontology. This demands that information represented in the ontology is both temporally contextualized and that relations among entities that are relevant during different temporal intervals are available to support user queries. Furthermore, it is necessary to account for the fact that the course of the ontology's evolution and the processes that have effected it are a part of the knowledge that should be brought to bear on the analysis of information at any given time.

The FDR/Pearl Harbor project is developing means to support enhanced search and retrieval from a set of documents drawn from the Franklin D. Roosevelt Presidential Library (FDRL). The documents in our collection refer to situations and events over the ten year period prior to the bombing of Pearl Harbor, during which the definitions of and relations among "entities"—especially, geo-political entities—were in a state of constant flux. A single, fixed ontology cannot capture the relevant contextual knowledge for all of the documents in our collection, because the entities in the world and the relations among them—and consequently, the configuration of the ontology—differ depending on the date of the document. Furthermore, a query concerning a given entity such as the U.S. Secretary of State may demand retrieving information concerning the different persons who filled this role prior to the U.S. declaration of war on Japan.

In this chapter, we first provide an overview of the FDR/Pearl Harbor project and a discussion of the requirements for an ontology to support historical research. We then define a framework for historical ontologies intended to address these requirements. Finally, we describe the FDR historical ontology and outline open problems and future work.

2. THE FDR/PEARL HARBOR PROJECT

2.1 Overview

The work reported here was undertaken in the context of the FDR/Pearl Harbor Project[1], which is enhancing a range of image, sound, video and textual data drawn from the Franklin D. Roosevelt Presidential Library (*FDRL*). The data in the FDRL represents one of the most significant collections of historical material concerning the history of America and the world in the twentieth century. The centerpiece of the collection is a body of textual material known as the "President's Secretary's Files" (PSF). The PSF is the most sought after collection of papers held in the FDR Library. The PSF is comprised of 150,000 documents, including letters, diplomatic correspondence, intelligence reports, memoranda, newspaper clippings, photographs, and other historical materials. The FDR/Pearl Harbor Project is concerned with a collection of 1,446 internal administration documents concerned with US-Japanese relations between 1931 and 1941, including memoranda of conversations, letters, diplomatic correspondence, intelligence reports, and economic reports.

The documents in our collection are used by historians, political scientists, policy-makers, diplomatic and intelligence analysts and others studying Japanese-American relations over the ten-year period leading up to the Japanese attack on Pearl Harbor. Research on Japanese-American relations during this period focuses on the nature of the diplomatic, military/strategic, and economic relations between the two nations—not only in isolation, but also in terms of the interactions among them. Historians and political scientists therefore rely on the documents in the FDRL as a primary resource to study the interplay of the dialogue between the two countries and between high-level officials in the Roosevelt Administration. Piecing together this type of information, however, is very time consuming. The vast majority of material an historian looks at is of no use to his or her research—it represents the chaff that must be sifted through before the researcher finds a kernel of information that is of real value. Through this lengthy process of elimination, the historian slowly builds his case; often with the result that it can sometimes take years to sift through all of the information required to arrive at a conclusion.

Information retrieval and extraction provide the potential to sift the information in the FDRL documents in ways that will greatly enhance the speed and utility of the research process. In a large body of data, enhanced

[1] Supported by U.S. National Science Foundation grant ITR-0218997.

search and retrieval techniques might uncover new evidence (or a body of evidence not previously viewed in its totality) that when analyzed could result in a shift in the historical interpretation of a given event. The Japanese decision to pull out of the London Naval Talks in December 1935, for example, might take on new importance in the history of US-Japanese inter-war diplomacy if the data were to show that this decision played a significant part in the deterioration of relations between the two states.

As the preceding example shows, historians and others studying historic documents seek not only to uncover facts, but also attitudes and opinions that result from, and/or lead to, events within the historical period during which they were produced. In particular, they want to know what the "actors"—the people, governments and military leaders and policy-makers—were thinking at the time. The FDR Project is therefore concerned with identifying evidence of such attitudes in the wording of documents in the corpus, and attributing this information to the appropriate person or entity. Attitude analysis can provide provocative information for the historian or political scientist to explore further by examining the documents themselves. For example, a quick analysis of a literal transcription of a statement by FDR to Japanese Ambassador Nomura on August 17, 1941 concerning the Japanese use of force in Southeast Asia shows a significantly higher percentage of words denoting strong (emphatic) language and power than a transcription of his remarks to Nomura on December 2, five days before the attack on Pearl Harbor (see Figure 2). The same is true to a lesser degree for the language in Secretary Cordell Hull's report of Nomura's replies in the same memoranda. One might have expected that conversations would become more heated as the attack on Pearl Harbor approached, but in fact the language of both FDR and the Ambassador appears to become more conciliatory. At the same time, FDR's language on December 2 contains more "positive" words than on August 17, while the reverse is true of Hull's report of Nomura's language. Thus the suggestion of the analysis runs counter to the widely held public perception that FDR "conspired" to force the United States into the war by provoking the Japanese in the final months leading up to the attack on Pearl Harbor.

CATEGORY	FDR/AUG17	FDR/DEC2	NOMURA/AUG17	NOMURA/DEC2
positive	5.76	7.61	4.76	3.49
strong	17.07	9.78	14.29	8.72
power	7.98	2.17	7.94	5.23
negative	2.44	3.26	.79	2.33
hostile	1.55	1.45	0	1.74

Figure 2. Percentage of words in several categories

2.2 Entity and Event Annotation

The FDR/Pearl Harbor Project corpus has been annotated for a wide range of linguistic phenomena and entities, including persons, titles, dates, locations, and organizations (military, government, civilian, etc.), as well as documents, treaties, policies, ships and other military apparatus, raw materials, monetary references, etc. Lexical units in the corpus are additionally annotated for a wide range of semantic categories, including words indicative of opinion and attitude.

In addition to entities, the corpus is annotated for events, including major historical events, which may or may not be mentioned in the documents; minor events referred to in the documents, such as a visit by the Japanese Ambassador to the Secretary of State; and communication events. Because a large portion of the documents in the collection are so-called “memoranda of conversations”, many are near-transcriptions of meetings between Japanese and US officials. We have therefore focused on communication events down to the level of the utterance (e.g., “X asked that…”) and apply attitude-recognition procedures to each utterance attributed to a given speaker. Note that the memoranda themselves represent complex communication events, in which several layers of subjectivity may exist. For example, a memorandum may comprise a report from Secretary Welles to FDR summarizing what the Japanese Ambassador said during a meeting with Secretary Hull and how the Secretary replied.

Event and entity annotation of the FDR documents was accomplished using the General Architecture for Text Engineering (GATE) system developed at the University of Sheffield (Cunningham, et al. 2002). Our work involved considerable extension of the pattern matching rules and gazetteer lists in the ANNIE entity recognition system provided in GATE to handle our data. Automatic annotation of the full corpus was bootstrapped using machine learning based on hand-validated annotations in a 100 document (10,000 word) sub-corpus.

2.3 Attitude Analysis

Our work builds on the increasing body of research concerned with the detection of attitudes and opinions in text[2] to provide information to the historian about the orientation of language in the documents and document segments in our corpus. Because the content of all the documents in the collection is linked to the time-line representing the progression of events, it

2 For a good overview of recent work, see *Exploring Attitude and Affect in Text: Theories and Applications.* Papers from the 2004 AAAI Spring Symposium, Technical Report SS-04-07, AAAI Press, 2004.

provides a means to study the impact of events on attitudes and vice versa, as well as the overall progression of attitude change over time.

Much of the current work on attitude/opinion analysis is concerned with identifying favorability/unfavorability (polarity) toward a given topic, and/or an indication of the author's emotional state with respect to that topic. Methodologically, this work expands approaches to content analysis undertaken by social scientists in the 1960s and 70s, by largely relying on pre-defined lists of categorized words, phrases, collocations, etc. constructed either by hand or automatically. Most of this research has focused on polarity of opinion toward the document topic (Pang, *et al.* 2002; Turney, 2002) or in individual sentences (e.g., Wiebe *et al.,* 2002, Yu and Hatzivassiloglou, 2003), although there have been some recent attempts to address smaller text segments using deeper linguistic analysis (Bethard, *et al.,* 2004).

Because our documents deal with a limited domain and are (more or less) stylistically consistent, we are developing means to automatically identify attitude and opinion by exploiting the rich syntactic and semantic annotation of our corpus and its supporting ontology. Rather than whole documents or single sentences, we focus on contiguous text segments that are attributed to a given point of view or "profile", the different types of which include direct statements (letter content that is not reported speech or quoted material), reported speech, and third-hand reported speech, as noted above, in order to determine whose attitude is represented.

3. ONTOLOGY SUPPORT FOR HISTORICAL RESEARCH

Ontology support for historical documents poses particular problems due to changes in the "facts" about the world that the ontology represents. One approach to these problems is to construct a series of independent ontologies, each of which is called into the play for the appropriate documents. There has been some work within the ontology community on "evolving" or "dynamic" ontologies (e.g., Heflin and Hendler, 2000; Davies, et al., 2002; Kahng and McLeod, 2000), primarily because of the practical need to modify ontologies as they are developed to correct errors, enlarge or shrink the set of included concepts, or adjust concepts to reflect changes in the domain. Such ontologies exist as a series of versions, similar to versions of computer software (Klein and Fensel, 2001; Klein, 2004; Stojanovic, 2004), with the underlying assumption that, like computer software, the relevant or "best" ontology is the most recent one. In contrast, ontology "versions" have equal status in our application; the validity of a given version is dependent on the temporal context of the user's query.

Temporal contextualization can be treated as a special case of "multiple views" into an ontology. Multiple ontology views are most often intended to enable users to access information relevant to a given situation or viewpoint rather than to a given timeframe. Nonetheless, the same strategies used to support multiple views can be used to represent temporally contextualized information using a mapping mechanism such as that provided in the web ontology language (OWL) (Patel-Schneider, et al., 2004; Bechhofer et al., 2004). However, while OWL and similar representation mechanisms provide means to map different ontology versions, they provide no means to express the *semantics* of the differences.[3] In our application, the type of change that modifies the ontology is as important as fact of the change itself—that is, each change can be viewed as a particular type of event that is represented in the ontology and therefore accessed for reasoning and retrieval.

The work on ontologies that is most relevant to our application has been done in the geo-spatial domain, which models objects of the different branches of geography, including physical and political geography, geology, geomorphology, climatology, meteorology, etc. Within this field, ontologies of change and process relations have been developed and used to interconnect ontologies representing different "snapshots" of the geographical world as it unfolds over time, in order to enable spatio-temporal reasoning (Grenon and Smith, 2004; Kauppinen and Hyvönen, in press). These models are necessarily focused on processes that affect geographical entities and the specifics of their spatial characteristics and overlaps as they evolve over time. As such, this work does not address some of the problems of handling the FDR data, and at the same time addresses areas irrelevant to our domain. For example, Grenon and Smith's SNAP and SPAN spatial ontology allows for widely variant views of the geographical realm depending on granularity, from large-scale geographical features such as mountains and oceans, down to the level of particular vehicles, buildings, and even individuals. Kauppinen and Hyvönen's ontology time series, which is most related to our historical data, is primarily concerned with computing geospatial overlaps among geopolitical regions over a historical time period. However, in both of these projects the need to deal with entities that change over time is critical, and we have been able to adapt and extend many of the fundamental ideas and methods of these models to handle historical data.

[3] Extensions to OWL have recently been proposed to provide means to represent change semantics (Avery and Yearwood, 2003).

4. THE HISTORICAL ONTOLOGY

The historical ontology provides two perspectives on the domain it models: a *synchronic* view, representing the state of the world during a given time interval, and a *diachronic* view that traces the changes in the domain as they unfold over time. The synchronic perspective is provided by a series of *snapshot ontologies* that are linearly-ordered over an encompassing time span—in our case, the ten year period from 1931-1941—each of which provides a snapshot of the world during a temporal sub-interval. The diachronic perspective is provided through a single, encompassing *time and event ontology* that covers the entire time span covered by the series of snapshot ontologies.

The time spans associated with the snapshot ontologies in our model are not fixed in duration; rather, they are determined by the occurrence of historical events that change the knowledge represented in the ontology, such as the Japanese invasion of Manchuria (which changes Manchuria's status) or the signing of the Tripartite Pact (after which Japan becomes one of the Axis Powers). These *key events* may or may not be referenced in the documents themselves. Any event referenced in the documents is instantiated in the time and event ontology and associated with its date of occurrence. Figure 1 shows an overall view of the historical ontology.

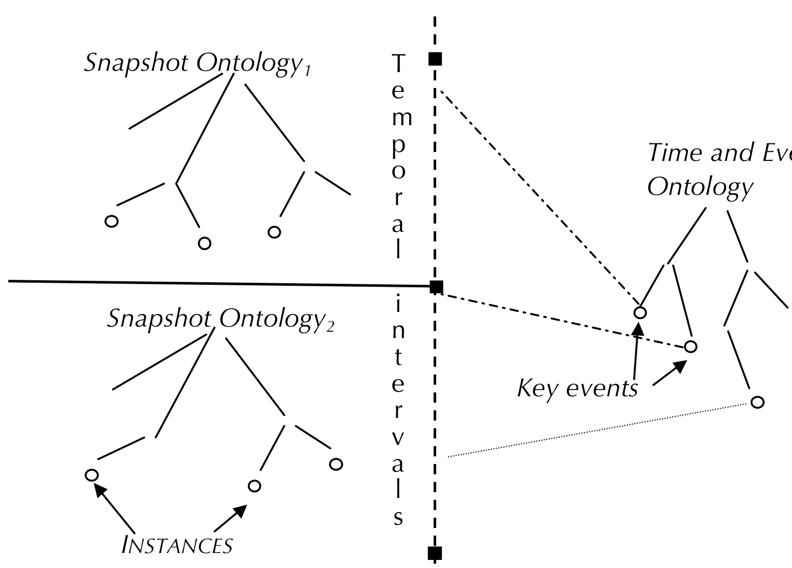

Figure 1. Overview of the historical ontology

4.1 A Model for Historical Ontologies

We define an *historical ontology* $H = < O_S, O_E, T>$, where O_S is a *temporal ontology series*, O_E is a time and event ontology, and T is the time span covered by H.

4.1.1 The temporal ontology series

The basis of the historical ontology is a *temporal ontology* $O = <S,T>$ which consists of the set of instances S representing entities that exist at any time during a time span T. Each instance $i \in S$ is associated with a temporal interval $T_i \supset T$ and is represented as a triple <*identifier*, T_i, P> where P is a set of properties associated with the instance. Instances in O represent *endurant* entities in the world, that is, entities that have continuous existence and a capacity to endure, including physical entities such as a particular person, a ship, a document, an army, etc. as well as governments, countries, and cities.

In our application, T is the ten year interval covered by our documents. The limits (i.e., start and end points) of temporal spans associated with instances in O define exclusively and exhaustively a set of *change points*. Change points are zero-length temporal intervals that mark the points at which one or more instances in O come into or go out of existence, or are modified. For example, in 1940 Vichy France comes into existence as a geo-political entity, whereas French Indochina as a French colonial protectorate ceases to exist as a unified entity. Modifications to an instance occur when property values change, as, for example, when Cordell Hull takes on the role of U.S. Secretary of State.

A *temporal ontology series* O_S is a tuple $<O,C>$, where O is a temporal ontology and $C = [t_0,\ldots t_n]$ is an ordered sequence of change points $t_i < t_i+1$, $0 \leq i < n$, so that $T = [t_0,t_n]$. Thus, O_S includes a series of n "snapshot" ontologies defined by contiguous, non-overlapping temporal intervals covering the entire span of time represented by T. Each instance included in O belongs to and persists over at least one snapshot ontology. A snapshot ontology for a given time span $T_s = [t_i,t_{i+1}]$ is created by identifying all instances that persist over T_s.[4]

Figure 2 provides a graphic representation in which the temporal limits associated with four ontology instances define five change points t_0 through t_4, which in turn define four distinct and contiguous temporal spans. A snapshot ontology representing the "state of the world" during a given span is constructed from all resources that persist over that span. For example, in

[4] The model for representing temporal ontologies and ontology series is adapted from the model for *ontology time series* defined in Kauppinen and Hyvönen, in press.

Figure 2 instance i_1 is associated with interval $[t_1,t_3]$, instance i_3 is associated with interval $[t_1,t_4]$, and instance i_4 is associated with interval $[t_2,t_3]$. This information can be used to construct ontology $O_3 = <\{i_1, i_3, i_4\}, [t_2,t_3]>$.

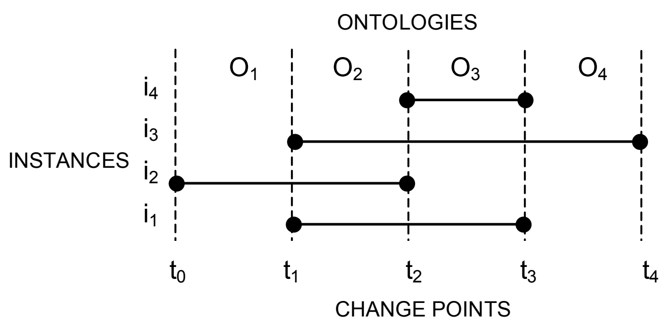

Figure 2. A temporal ontology series

4.1.2 The Time and Event Ontology

In addition to a temporal ontology series, the historical ontology includes a *time and event ontology* $O_E = <R,T>$, where R is a set of instances representing events that occur in time, such as a communication between Roosevelt and his Secretary of State, the attack on Pearl Harbor, the imposition of an oil embargo on Japan, etc., and temporal intervals. T is the entire span of time represented in the temporal ontology series described above. Event instances are associated with temporal intervals.

Historical events may be categorized as either *KeyEvents*, such as a military invasion or a change in an administrator, which cause modifications to the ontology and are therefore associated with temporal intervals that are change points; or *InformationalEvents* that do not modify the ontology, such as a communication event involving two government officials or a visit to the White House by the Japanese Ambassador. Note that some historical events represented in E are not mentioned in the documents in the collection, but rather comprise a portion of general world knowledge concerning major events prior to and during World War II.

4.1.2.1 Ontology Change events

A class of *OntologyChangeEvents* is defined that provide mappings between instances in different snapshot ontologies, including, for example, *unification*, *separation*, *name change, change of political control*, *change of administrator,* etc. *OntologyChangeEvents* are conceptually similar to the *change bridges* described in (kaupp); however, we follow (Smith and Grenon 2004) in making *OntologyChangeEvents* entities in their own right. This enables reasoning involving change events themselves, including patterns of causal relations, etc., which are of particular interest to historians and political scientists.

OntologyChangeEvents may be *qualitative* or *substantial*. Qualitative changes include the following:

– *Change in property value*: In many cases the value of one or more properties of an entity will be instantiated by different values at different times. For example, after the Japanese invasion of Manchuria, the value of the *governed-by* relation changes, but the country or area itself remains the same and is the subject of the change—that is, the property itself remains associated with the entity while transitioning through successive values.
– *Qualitative creation*: Henry Lewis Stimson takes on the role of Secretary of State.
– *Qualitative destruction*: Henry Lewis Stimson ceases to be Secretary of State.

Qualitative changes are *temporally transitive*—that is, the entity affected by the change persists across time spans. For example, Stimson is still the same person instantiated in the ontology before and after he becomes Secretary of State. Similarly, Japan is still Japan even after it becomes an Axis Power.

Substantial changes occur when entities are created or destroyed, as, for example, when a geo-political entity is divided up so as to produce two or more new geo-political entities, or when two or more geo-political entities are unified. For example, French Indochina was once one entity, governed by French Colonial rule; after the Japanese invasion in 1940, its territory is divided into Northern Indochina and Southern Indochina, each with its own government and control. Effectively, at the point of change the concept of French Indochina as a geo-political entity becomes obsolete, while two new entities representing Northern Indochina and Southern Indochina are introduced. Substantive changes are *temporally intransitive.*

4.1.2.2 Temporal intervals

Temporal intervals are associated with both historical events (KeyEvents and InformationalEvents) and OntologyChangeEvents. *Change points* are a special sub-class of temporal interval used to delimit the temporal spans associated with members of O_S, as defined above. Events in O_E are related to temporal intervals via a "TemporalLocation" property. Time in our framework is regarded as a linear continuum; temporal intervals are related to one another with the primitive relation "before", a strict total order that holds between two temporal intervals when one is earlier than the other.

In principle, temporal intervals can be defined at any level of granularity and may overlap or be discontiguous. In our current implementation, we have taken a simplified approach wherein instances of temporal intervals in the ontology represents a single date in month-day-year form. In the terminology of interval temporal logic, we can say that every temporal interval has length zero and therefore represents a temporal *point* consisting of a single day. Each event in O_E is associated with one such temporal interval; events in our ontology are thus regarded as strictly *punctual*, with no meaningful duration. Durational events—for example, the Japanese occupation of Northern Indochina or the war between Great Britain and Germany—can be inferred on the basis of events (e.g., a country is at war with another power if it has declared war on that country and no truce has been established) or information in O_S (e.g., the change points associated with the instance of Northern Indochina whose properties indicate that it is under military occupation by Japan define the interval during which this property holds).

4.2 Relations among ontologies

The historical ontology framework includes several types of relations within and among ontologies. *Intra-ontological* relations exist between constituents of a single ontology, for example, the relation of part to whole between southeast Asia and the Pacific Region. *Trans-ontological* relations exist between entities that are constituents of different ontologies, that is, between two or more snapshot ontologies or between the endurant temporal ontology and O_E. A typical example involves a participation relation between an agent in one of the endurant ontologies with an event in O_E, such as the Japanese military force's participation in the invasion of Manchuria. Finally, *Meta-ontological* relations exist between whole ontologies or between an ontology and an entity. An example of a between-ontology relation is the relation of temporal order between different snapshot ontologies in O_E.

Figure 3 gives an example of some relations among components of O_S (in white) and O_E (in gray). Instances are depicted as squares and classes are in ovals. The center of the figure is a separation event ("separation21"), which is an entity in O_E. It is associated with a temporal interval that is a change point ("t_6"), and through the "before" and "after" relations provides a bridge from the instance of French Indochina and the instances of Northern and Southern Indochina into which is it split after change point t_6. These three instances represent endurant entities, and are therefore a part of O_S. French Indochina is a constituent of the snapshot ontology associated with (indexed by) change point t_5, and potentially other snapshot ontologies associated with change points prior to t_5 as well. Northern and Southern Indochina are constituents of the snapshot ontology indexed by t_6 and may be constituents of snapshot ontologies with later change point indexes.[5] Figure 3 also shows the key event from O_E that is associated with t_6.

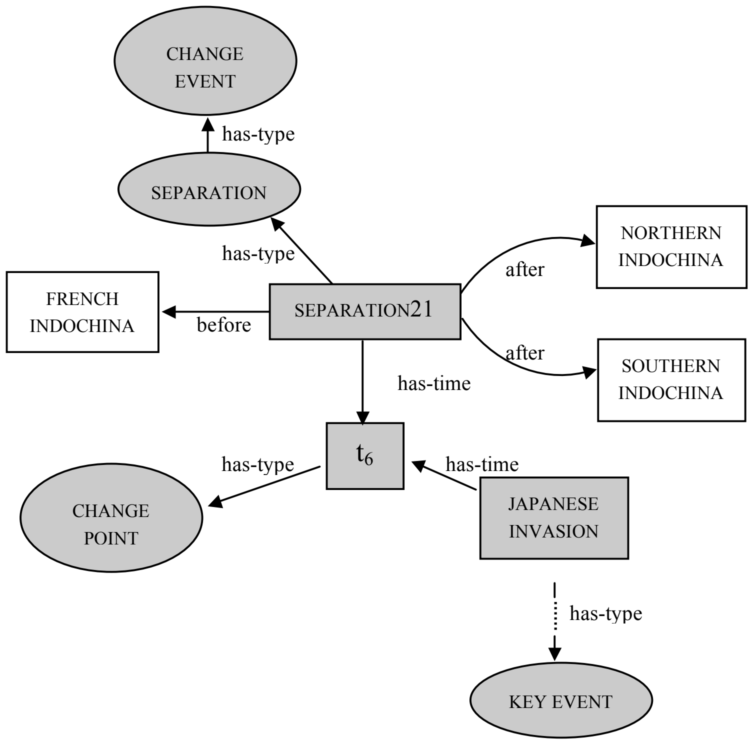

Figure 3. Ontology relations

5. THE FDR ONTOLOGY

The FDR/Pearl Harbor project is building an historical ontology based on the model presented in the previous section representing the entities and events in its document collection. The starting point for the ontology is the

[5] It should be noted that it would be possible for French Indochina to be a constituent of a snapshot ontology or ontologies with change points later than t_6 if Northern and Southern Indochina were to have re-merged.

Suggested Upper Merged Ontology (SUMO) (Niles and Pease, 2001) together with the Mid-Level Ontology (MILO) and several ontologies available from the Agent Semantic Communication Service (ASCS) Agent Semantic Communication Service (ASCS)[6]. Our current adaption includes only the information from these ontologies that is relevant to our domain, which comprises a relatively small subset—for example, of the over 1500 classes in MILO, we use only 185. As such, our strategy for building the our ontology is both top-down, in the sense that we begin with the upper level concepts defined in ontologies like SUMO, as well as bottom-up, in that we rely on the entities identified in our document collection to determine the set of classes and relations to include. Our most substantial modification to the ontologies is extension; for example, we extend the "form of government" class to include "collaborative governments", such as Vichy France or northern Indo-China between July 1940 and July 1941, and "governments-in-exile," such as the Dutch Government during the same period. We also eliminate inapplicable classes such as "Former Soviet or Eastern European Country" and a sizable number of nations/governments that did not exist until after the end of World War II. All of our ontologies were developed within GATE (Boncheva, et al., 2004) using Protégé 2000[7].

The FDR temporal ontology includes only those endurant entities that are explicitly mentioned in the document collection or appear in the associated metadata (author, recipient, location, date, etc.), which fall into the following general categories:

– *geopolitical entities*, such as countries, cities, and regions.
– *geopolitical organizations*, primarily governments, their major subdivisions (diplomatic, executive, legislative, etc.), departments, and officials.
– *military organizations*, most of which are associated with governments. This category includes the various military branches and forces, and military positions.
– *military vehicles/apparatus*, such as ships, tanks, aircraft, etc.
– *geographical objects*, including major forms such as continents, lakes, oceans, mountains, and islands. Note that geographical objects also include geopolitical entities such as countries and cities.
– *geographical artifacts*, including major artifacts with strategic importance such as roads, bridges, canals, etc.
– *documents*, such as treaties, pacts, *modus vivendi*, etc., regarded as physical objects.

[6] http://reliant.teknowledge.com/DAML/
[7] http://protege.stanford.edu

- *agreements/contracts/cooperations*, which includes pacts, treaties, etc. as entities that depend for their existence upon the countries involved. Some agreements can exist as physical document without also being a contract, as, for example, the *modus vivendum* presented by the Japanese to the U.S. several weeks prior to the bombing of Pearl Harbor.
- *people,* most of whom fill government or military positions in the U.S., Japan, Great Britain, France, Germany, and Russia, but also several others who served as formal or informal advisors to the Roosevelt Administration.
- *political organizations*, including entities such as the Nazis and the Axis and Allied Powers.

Entities in the time and event ontology include both events mentioned in the documents and major historical events that cause modifications to the domain model. An important class of events in the ontology is communication events, which are sub-classified at a relatively fine level of granularity due to their importance in our data. A major goal of the FDR/Pearl Harbor project is to enable retrieval of communication content (typically individual statements but also including whole documents), apply attitude analysis to determine its orientation along dimensions such as power/control, submission, hostility, cooperation/friendliness, etc., and map attitudinal orientation over time. For example, the historian may wish to investigate changes in attitude in responses by any representative of the Japanese government to a question posed by a U.S. government official between June 1941 and December 1941. We have adopted a portion of the ontology of communication verbs in the FrameNet database (Ruppenhofer et al 2005), which includes sub-classes such as request, statement, judgment-communication, etc., as a starting point to which we have added further sub-classifications[8]. In particular, we have made distinctions on the basis of polarity, which is not accounted for in FrameNet (e.g., "acclaim" and "condemn" are both categorized as judgement_communication verbs). Communication event classes in FrameNet were mapped to subclasses of LinguisticCommunication in SUMO where possible, or used to extend SUMO.

5.1 Open Problems and Future Work

The creation of the FDR ontology poses some interesting challenges beyond accounting for temporal contextualization. For example, in some

[8] We utilized a clustering algorithm to automatically generate sub-classes of communication events and sense-tag the verb sets with WordNet senses; see (Ide, 2006) for a description of the method.

cases, the temporal transitivity of a given change—i.e., the persistence or non-persistence of an entity as a result of the change—is not clear-cut. This problem arises in the representation of France, which is divided after its submission to Nazi Germany in 1940 into two politically distinct regions: the northern area occupied and controlled by the Germans, and the southern area that is officially ruled by the puppet Vichy government. However, it is inappropriate to eliminate the original instance of France as a geo-political entity associated with its entire land area from the ontology after June, 1940, since the concept of the whole of France as a sovereign nation remains, if only due to the existence of the Free French Forces of Charles DeGaulle based in London that claimed to be the legitimate government of France during this period. There is no straightforward way to model this situation in existing systems such as SUMO without violating constraints that are obviously appropriate in most cases, such as the requirement that a geopolitical area has only one government, or without losing the information that the geopolitical area represented by the whole of France corresponds to the two geopolitical areas represented by Vichy France and German-occupied France. This raises the general question of the degree to which the identity of a geopolitical area persists as it undergoes political and geographical changes, that is, need there be a general concept of, say, France that persists over time despite such changes? Conversely, at what point are changes substantial enough to effectively create a different or new concept?

Our current implementation supports modifications and the creation and deletion of ontology *instances*, but it does not support modifications to ontology *classes*. In principle, the model outlined above could be extended to enable modifications to ontology classes as well as instances, but in practice, such modifications would involve considerable maintenance overhead and potentially lead to conflicting or invalid instance declarations. However, to handle historical data covering significantly greater time spans, it may be necessary to provide for modification of class definitions.[9] We are currently exploring this possibility.

6. CONCLUSION

The historical ontology framework outlined in this chapter is designed to support historical research involving the documents drawn from the *FDRL*

[9] Creation and destruction of classes in the current model would be handled by associating classes with the temporal intervals during which they are valid.

that relate to Japanese-American relations between 1931 and 1941. With an ontology in the background, historians have a previously unavailable capability for "generic search", involving, for example, requests to see documents in which any representative of the Japanese government (rather than a specific person or list of people) is mentioned. In addition, the historical ontology provides for *temporal contextualization* of query terms, in order to retrieve all relevant results. For example, if a query concerns the Secretary of State, the results will include documents referring to Henry Lewis Stimson prior to 1933, and Cordell Hull thereafter. Temporal contextualization can also be explicitly provided in the query itself; for example, a query referring to members of the Axis Powers that is constrained to the period between 1936 and 1939 will yield results for Germany and Italy only, since Japan became an Axis Power only after September, 1940, the date of the signing of the Tripartite Pact.

The two-part architecture of the historical ontology separates temporally-defined entities, including events and temporal intervals themselves, from endurant entities that exist in the physical and geo-political realms. We feel that this makes the task of domain modeling somewhat more manageable, since this division corresponds conceptually to familiar perspectives on reality, that is, the world of objects and entities (in the snapshot ontologies) vs. events. Also, the inclusion of OntologyChangeEvents as a first class object in the time and event ontology enables retrieval of information represented in the ontology itself about the changes to an entity or property over time, such as the course of change in the Asian regions under Japanese control between 1935 and 1941, or all "invasion events" that occurred in 1939.

7. REFERENCES

Avery, J. and J. Yearwood. 2003. dOWL: A Dynamic Ontology Language. *ICWI 2003*: 985-988

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuiness, D., Patel-Schneider, P. and L. A. Stein. 2004. Owl Web Ontology Language Reference. On-line publication at http://www.w3.org/TR/owl-ref/

Bontcheva, K., Tablan, V., Maynard, D. and H. Cunningham. 2004. Evolving GATE to Meet New Challenges in Language Engineering. *Natural Language Engineering*. 10:3/4, 349-373.

Cunningham, H., Maynard, D., Bontcheva, K. and V. Tablan. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02).*

Davies, J., Duke, A., and A. Stonkus. 2002. Ontoshare: Using Ontologies for Knowledge Sharing. *Proceedings of the International Workshop on the Semantic Web at the Eleventh International World Wide Web Conference.*

Grenon, P. and B. Smith. 2004. SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation*, 4:1, 69–103.

Heftin, J. and J. A. Hendler. 2000. Dynamic Ontologies on the Web. *Proceedings of AAAI/IAAI 2000*, 443-449.

Ide, N. 2006. Making Senses: Bootstrapping Sense-Tagged Lists of Semantically Related Words. In Gelbukh, A. (ed.). *Computational Linguistics and Intelligent Text Processing.* Lecture notes in Computer Science 3878, Springer, 13-27.

Kahng, J. and D. McLeod. 2000. Dynamic classification ontologies. In Arbib. M. A. and J. Grethe (eds.), Computing the Brain: A Guide to Neuroinformatics, Academic Press. 241-254.

Kauppinen, T. and E. Hyvönen, Modeling and Reasoning about Changes in Ontology Time Series. 2005. In Kishore, R., Ramesh, R. and R. Sharman (eds.), *Ontologies: A Handbook of Principles, Concepts, and Applications in Information Science*. Springer, 319-338.

Klein, M. 2002. Supporting Evolving Ontologies on the Web. In Lindner, W. and J. Stuller (eds.), *Proceedings of the EDBT 2002 PhD Workshop*, 51-58.

Klein, M. and D. Fensel. 2001. Ontology versioning on the Semantic Web. *Proceedings of the International Semantic Web Working Symposium (SWWS)*, 75–91.

Noy, N. and M. Klein. 2003. Ontology evolution: Not the same as schema evolution. *Knowledge and Information Systems 5*.

Patel-Schneider, P., Hayes, P. and I. Horrocks. 2004. Owl Web Ontology Language Semantic and Abstract Syntax. On-line publication at http://www.w3.org/TR/owl-semantics/

Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D. and A. Kirilov. 2004. KIM - A Semantic Platform for Information Extraction and Retrieval. *Journal of Natural Language Engineering,* 10: 3-4. Cambridge University Press, 375-392,.

Ruppenhofer, J., Ellsworth, M., Petruck, M., and C. Johnson. 2005. FrameNet: Theory and Practice. On-line publication at http://framenet.icsi.berkeley.edu/

Sider, T. 2001. Four-Dimensionalism. An Ontology of Persistence and Time. Oxford: Clarendon Press.

Stell, J. G. and M. West. 2004. A 4-dimensionalist Mereotopology. In Varzi, A.C. and L. Vieu (eds.), Formal Ontology in Information Systems. IOS Press, 261–272.