Question Answering Based on Temporal Inference

Sanda Harabagiu and **Cosmin Adrian Bejan**
Human Language Technology Research Institute
The University of Texas at Dallas
Richardson, TX 75083-0688, USA
sanda,ady@hlt.utdallas.edu

Abstract

Answering questions that ask about temporal information involves several forms of inference. First, relations between events and temporal expressions need to be inferred, either in the question or in the answer. Second, semantic inference between events, entities and their definitions needs to be performed. Sometimes, semantic inference employs aspectual information to relate events expressed by different lexical units. Third, temporal inference is preconditioned by relations between events, or between events, their arguments and their time anchors. In this paper we present a Question Answering (QA) methodology for handling temporal inference by combining all these forms of inference.

The Problem

The AQUAINT[1] program is a multi-project effort aimed at creating capabilities of processing complex questions and finding their answers from heterogeneous collections of texts. An important component of this effort deals with the recognition and processing of temporal information for Question Answering (QA). When asking a question that refers directly or indirectly to a temporal expression, the answer is expected to validate the temporal constraints. To achieve such functionality, QA systems need to (1) incorporate relations between temporal expressions and events or entities mentioned in the question; and (2) to rely on temporal inference for justifying the answer. Answer justification was presented in (Harabagiu *et al.* 2001) as a multi-feedback methodology, but this mechanism did not include any temporal constraints. Whenever the answer to a question needs to be justified, if temporal expressions are involved, the justification must contain some form of temporal inference. For example, the *expected answer type* of question Q_1 is a DATE:

Q_1: *"When did Iraq invade Kuwait?"*

The expected answer type is an argument of the event E_1=*"invade"* which has two more arguments: *"Iraq"* and *"Kuwait"*. The answer to Q_1 is *"2 August 1990"*, extracted from the context:

A_1: *"Iraqis have been struggling under UN-sanctions ever since Hussein's annexation of Kuwait on 2 August 1990."*

In the paragraph A_1, the answer marks the DATE of the event E_2=*"annexation"*, which has *"Hussein"* and *"Kuwait"* as arguments. The event E_1 differs from the event E_2 , but they are related. The arguments *"Iraq"* and *"Hussein"* are also different but related. To justify the answer, these relations must (a) be recognized and (b) incorporated into the temporal inference.

The relation between the events E_1 and E_2 is granted by the semantic information encoded in WordNet 2. The second sense of the verb *"invade"* is defined as *"marching aggressively for the purpose of conquest"* and the first sense of *"annexing"* is defined as *"taking territory by conquest"*. *"Conquest"*, the common part of the definitions, represents the GOAL of E_1 and also the MEANS of E_2. The temporal inference makes the DATE relation transferable between E_1 and E_2 if GOAL(E_1)=MEANS(E_2). Moreover, the temporal inference is complete only if the relation between *"Iraq"* and *"Hussein"* is identified. The PERSON *"Hussein"* is deemed a metonymy and coerced into the COUNTRY *"Iraq"*. This is possible because the entire text collection is tagged with all the named entities, which are used as indexes. Whenever a PERSON and a COUNTRY or NATIONALITY belong to the same NP, the PERSON is coerced into the COUNTRY. The same relation exists between PERSONS and ORGANIZATIONS.

Processing questions that involve temporal inference relies on (1) the recognition of events/states and of entities that participate in them; (2) the relative ordering of events in the question and in the texts; (3) the temporal properties of the entities being questioned; and (4) identification of the expected answer and its relations to temporal expressions mentioned in the question or candidate answers. The first three aspects are currently addressed by the TimeML specification language (Ingria & Pustejovsky 2002). In this paper we address the fourth aspect, namely the temporal inference of the exact answer.

The remainder of the paper is organized as follows. Section 2 describes the Q/A model whereas section 3 reports on the

[1]AQUAINT is an acronym for Advanced QUestion Answering for INTelligence.

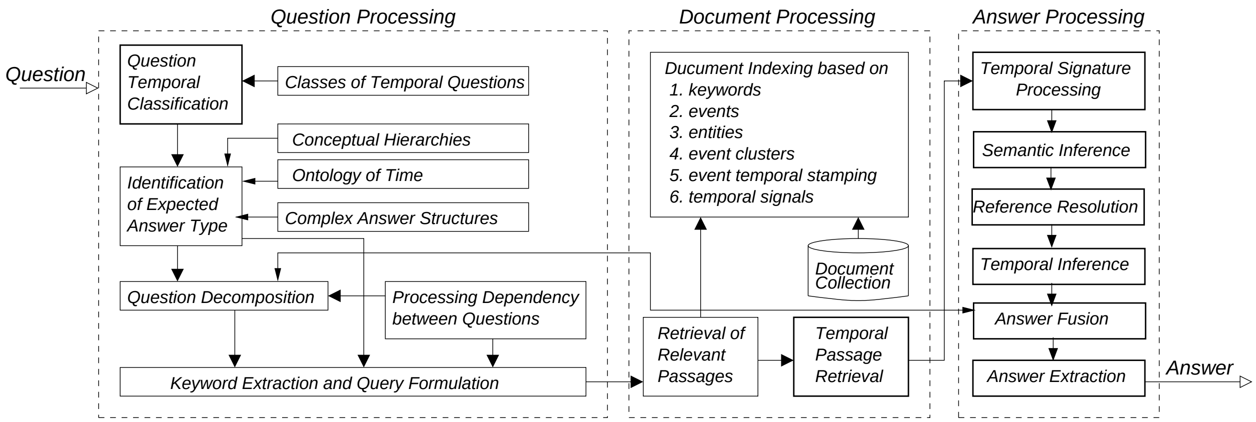

Figure 1: Architecture of Q/A system that incorporates Temporal Inference.

temporal signature and the temporal inference. Section 4 summarizes the conclusions.

The Q/A Model

Question Answering that is based on temporal inference involves three main processing stages: (1) *question processing* for interpreting the question, its temporal requirements and for formulating a query that produces candidate answers; (2) *document processing*, that is based on several forms of indexing, including indexing based on temporal information; and (3) *answer processing*, where most of the temporal inference takes place before finding and extracting answers. Figure 1 illustrates our Q/A architecture that incorporates temporal inference in all three main modules.

The first step in question processing is based on the classification of questions according to the temporal information that is recognized in the question. As a starting point for our classification, we have used the list of questions produced in the TERQAS Workshop[2]. Some of the question classes we considered are listed in Figure 2. In Figure 2 the underlined words in each question indicate the non-content words that are responsible for the question classification. Factual questions are characterized by the question stem *"when"*, expecting a date as an answer. Thus the expected answer type associated with such questions is either a textual expression that can be identified as a DATE by a Named Entity Recognizer or a list of such expressions. For the question exemplifying factual questions in Figure 2, John Paul II might have had several visits to Poland, and thus all the corresponding dates are correct answers. Questions asking about repetitive events or a time range that determined some superlative or comparative attribute may also be stemmed by the word *"when"*. Their classification into different classes is due to the presence of additional words, that indicate either repetition, e.g. the adverb *"normally"* modifying the verb *"arrive"* or superlative attributes, e.g. the adjective *"major"* modifying the noun *"growth"*. The other classes of questions that are listed in Figure 2, are characterized by (i) the presence of a date, time range expression in the question; (ii) the presence of temporal signals, e.g. *"since"*, *"after"*; or (iii) the need to decompose the question due to a temporal relation between events, indicated by a temporal signal.

As Figure 1 indicates, before question decomposition, the identification of the expected answer type is performed by using three resources: (a) conceptual hierarchies, built from the WordNet database or from other ontological resources; (b) an ontology of time encoding the concepts and relations reported in (Hobbs 2002); and (c) a set of complex answer structures that model event structures and processes. The complex answer structures help establish relations between the answer types of decomposed questions. For example, for the question Q_2:*"Where did Michael Milken work while attending graduate school?"* the complex answer structure is illustrated in Figure 3.

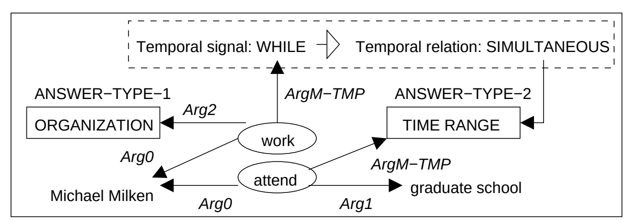

Figure 3: Complex answer structure with temporal information.

The answer type of the question is given by ANSWER-TYPE-1, which is an ORGANIZATION, since this is where people typically work. The two predicates *"work"* and *"attend"*, which are recognized in question Q_2 are represented in Figure 3 along with their arguments. The predicate-argument structures that are recognized are similar to the ones annotated in PropBank[3]. The expected arguments of each verbal predicate are numbered sequentially from Arg0 to Arg5. Generally, Arg0 would stand for *agent*, Arg1 for *direct object* or *theme* whereas Arg2 represents *indirect object*, *benefactive* or *instrument*, but mnemonics tend to be verb

[2] TERQAS was an ARDA Workshop focusing on Temporal and Event Recognition for Question Answering Systems. The final report of the Workshop is available from www.cs.brandeis.edu/∼jamesp/arda/time/readings.html

[3] The PropBank annotations, performed at University of Pennsylvania are described at *www.cis.upenn.edu/∼ace*

Question Class		Example
1 Factual (single/multiple dates)	Time−Related	When did the Pope visit Poland?
2 Time range (single event)		How long did Iraq fight with Iran?
3 Relative time range		Where can I find research information in the Israeli Palestinian issuess since 1991?
4 Repetitive event	Event−Related	When does the temporao normally arrive in Brasil?
5 Typical event		How long does it take on average to build a 500−room hotel in Las Vegas?
6 Time anchored event		What important things happened in the year 1987?
7 Events in time range		What did George Bush do after the U.N. Security Council ordered a global embargo on trade with Iraq in August 90?
8 Entity change in time period	Change−Related	What happened to world oil prices after the Iraqi "annexation" of Kuwait?
9 Entities related to events/states changes		I want to find pictures of presidents from the 1940−1949.
10 Quantity change in time period		How much did Las Vegas grow in population since 1980?
11 Entity related to events at time stamp	Entity−Related	Which two nations met in Washinton on August 14, 1990 to discuss a naval blockade against Iraq?
12 Age at time stamp		How old was Michael Milken in January 1989?
13 Comparative	Comparative	What is the difference beetwen the teenager's average weight today and in the 80's?
14 Period of comparative/superlative attribute		When was the period of major growth in Las Vegas?
15 Alternative temporal relation	Temporal−Order	Did John Sununu resign before or after George Bush's ratings began to fall?
16 Temporal relation		Where did Michael Milken work while attending graduate school?

Figure 2: Classes of Temporal Questions.

specific. For example, when retrieving the argument structure for the verb-predicate *attend* we find Arg0:*thing attending* and Arg1:*thing attended*. Additionally, the argument may include functional tags from Treebank, e.g. ArgM-DIR indicates a directional, ArgM-LOC indicates a locative, and ArgM-TMP stands for a temporal. Figure 3 illustrates two ArgM-TMP relations. The first relation is established between the predicate *"work"* and its argument *"while attending graduate school"*. This argument is headed by the temporal signal *"while"*, which indicates a temporal relation of equality with the implicit TIME RANGE in which the predicate structure headed by *"attend"* occurred. The second ArgM-TMP relation is established between the predicate *attend* and the TIME RANGE, which represents an implicit answer type for question Q_2.

To extract the correct organization as an answer to Q_2, employment relations to Michael Milken must be discovered within the TIME RANGE, representing the ANSWER-TYPE-2. Therefore two simpler questions, one for ANSWER-TYPE-1 and one for ANSWER-TYPE-2, are generated in the question decomposition module. The module that processes the dependency between the questions determines that the question corresponding to ANSWER-TYPE-2 needs to be processed before processing the question corresponding to ANSWER-TYPE-1. The dependencies between the questions also determine the keyword extractions. Typically, the keywords that are used to retrieve candidate answers for a question are also used when processing any other question it shares dependencies with.

The complex answer types also enable the retrieval based on events, entities and event clusters. Moreover, the document processing module assigns time-stamps to event clauses with the method presented in (Filatova & Hovy 2001). Time stamping of document clauses is complemented by inferring temporal relations both at sentence level and at paragraph level. Temporal ordering is produced in an ad-hoc way based on the indexing of temporal signals like *"after"*, *"since"* or *"before"*. All documents are indexed with all these forms of information, that enables the retrieval of candidate text passages. However, before starting the answer processing, passages that do not contain time stamps or do not comply with the temporal relations that are searched are filtered out. For example, when retrieving relevant passages for the question *"What African countries gained independence in the 60s?"* every passage that is not timestamped with any year from the 60s decade is filtered out.

The answer processing module starts by inferring a temporal signature for each candidate paragraph. For question answering that involves temporal inference, the temporal passage retrieval module has allowed only passages that contained at least one absolute or relative time expression. The temporal signature captures temporal relationships between time expressions and predicates that are related to the complex answer structures. It also captures the event temporal orderings of the predicates and their relations to the answer structures. For example, a candidate paragraph for Q_2:*"Where did Michael Milken work while attending graduate school?"* is:

> He was enrolled at University of Pennsylvania's Wharton School in 1967. After finishing his master's degree in bussiness administration two years later, he developed financial theories that had been proven in the world's markets and are now considered mainstream.

The temporal expressions *"1967"* and *"two years later"* are connected to the predicates *"enrolled"* and *"finish"* (*"his master's degree"*) respectively. Both of these predicates are related to the predicate *"attend"* from Q_2, since being enrolled presupposes attending a school, and finishing a graduate degree indicates that the attendance has stopped. The presence of semantic expression of type TIME RANGE is not detected, but the paragraph has a temporal signature from which the time range 1967-1969 can be inferred, which is accomplished in the next phase of the Answer Processing module: the temporal inference phase illustrated in Figure 1. The answer fusion phase enables the recognition of consistent temporal functions between paragraphs that correspond to different question decompositions. For example, such consistency between the time range 1967-1969 is inferred with the following paragraph, which contains an expression in the semantic class ANSWER TYPE1 from Figure 3:

> Starting in 1969, when he joined the firm that would became Drexel Burnham Lambert, Milken helped finance thousands of companies.

The temporal signature of the paragraph indicates (1) an inclusion relation between 1969 and the time range 1967-1969; (2) a time stamp for the event *"start"* and *"1969"* and (3) a relation between *"joining"* and *"helping"* marked by the signal *"when"*. The final phase of the answer processing module extracts the ORGANIZATION (or list of ORGANIZATIONS) that is sought by Q_2, i.e. *"Drexel Burnham Lambert"*. Answer extraction and fusion depends on temporal inference, which in turn is informed by semantic inference and reference resolution. Section 3 details the support of semantics and reference resolution in the temporal inference process.

Temporal Inference

Models of time consider either *instants* or *intervals* or both, as indicated in (Setzer 2001). Both forms of time models are captured by the TIMEX3 expressions annotated by TimeML (Hobbs & Pustejovsky 2003). There are three types of TIMEX3 expressions: (a) fully specified temporal expressions, e.g. *August 14, 1990*; (b) underspecified temporal expressions, e.g. *Monday*, *next month*, *last year*, *two days ago*; and (c) durations, e.g. *two months*, *a week*. Time expressions anchor events and states in narratives. To represent and to reason about how the world changes, (Hobbs & Pustejovsky 2003) claim that event recognition and time anchoring drive the basic inferences from text. TimeML considers "events" (and the corresponding <EVENT> tag) as a cover term for situations that *"happen"* or *"occur"*. TimeML also considers predicates describing *"states"* or *"circumstances"*. There are seven types of events considered in TimeML: (1) occurrence, e.g. *die*, *crash*, *build*, *merge*, *sell*; (2) state, e.g. *on board*, *kidnapped*, *loved*; (3) reporting, e.g. *say*, *report*, *announce*; (4) immediate-action, e.g. *attempt*, *try*, *promise*, *offer*; (5) immediate-state, e.g. *believe*, *intend*, *want*; (6) aspectual, e.g. *begin*, *finish*, *stop*, *continue*; and (7) perception, e.g. *see*, *hear*, *watch*, *feel*.

In texts, temporal objects (either time expressions or events) are related. Such relations are signaled by (a) temporal prepositions, e.g. *during*, *on*; (b) temporal connectives, e.g. *when*, *while*; and (c) temporal subordinates, e.g. *if*, *then*. To capture such relations, TimeML uses the SIGNAL tag, whose functionality was introduced by (Setzer 2001). This tag also marks polarity indicators, such as *not*, *no*, *none*, as well as indicators of temporal quantification, e.g. *twice*, *three times*. It is important to note that not all temporal relations are marked in texts by temporal signals. Moreover, the signals indicate a relation, but they are ambiguous, and thus the same signal may correspond to many different relations. For example, *"while"* may indicate temporal equality as well as causation.

To capture all temporal relations in text and to provide means for disambiguating them, TimeML uses a set of three LINK tags: (1) TLINK or Temporal Link, representing temporal relations holding between events or between an event and a time; (2) SLINK or Subordination Link, used for contexts introducing relations between two events, or an event and a signal; and (3) ALINK or Aspectual Link representing the relationship between an aspectual event and its argument event. The TLINK makes explicit the following relations: (1) SIMULTANEOUS; (2) BEFORE; (3) AFTER; (4) IMMEDIATELY BEFORE; (5) IMMEDIATELY AFTER; (6) INCLUDING; (7) HOLDS; (8) BEGINNING and (9) ENDING. The SLINKS are one of the following sorts: (1) MODAL; (2) FACTIVE; (3) COUNTER-FACTIVE; (4) EVIDENTIAL; (5) NEGATIVE EVIDENTIAL and (6) NEGATIVE. The aspectual relations encoded by the ALINK are (1) INITIATION; (2) CULMINATION; (3) TERMINATION and (4) CONTINUATION.

The recognition of temporal expressions and of temporal relations similar to those encoded in TimeML is important for textual QA. For example, the question Q_3: *"How long did Iraq fight with Iran?"* is classified to ask about a TIME RANGE, due to the presence of the question stem *"how long"*. However, the correct answer does not necessarily contain a temporal expression representing a time range/duration/interval. The answer is provided by the following paragraph:

> The Iran−Iraq War started on 22 September 1980. Initially, most countries treated this war as nothing more than border skirmishes between two neighbouring countries and were also quick to label Iran as the agressor then (not surprising after the Islamic Revolu− tion in Iran in 1979). The war finally ended when both Iraq and Iran accepted UN Resolution 598 in August 1988 .

The answer that is inferred from this paragraph is 22 September 1980 - August 1988. There are seven time expressions recognized in the paragraph: *"22 September 1980"*, *"initially"*, *"quick"*, *"then"*, *"1979"*, *"finally"* and *"August 1988"*. Of interest are only the time expressions related to events that paraphrase the question. In Q_3, the event of *"Iraq fighting with Iran"* can be paraphrased by the *"Iraq-Iran War"* expressed in the first sentence of the paragraph. The same event is referred later two more times in the paragraph, by underlined expressions *"this/the war"*. Only the first and the last reference are relevant. The first reference (*"The Iran-Iraq War started on 22 September 1980"*) indicates an aspectual relation of INITIATION between the event *"the war"* and the fully specified temporal expression *"22 September 1980"*. The third reference has an aspectual relation of TERMINATION, which is strengthened by the adverb *"finally"*. The temporal argument of the TERMINATION relation is implicit being marked by the connector *"when"* to the anchored event *"both Iraq and Iran accepted UN Resolution 598 in August 1988"*. The inference is that the war ended both (a) when the resolution was signed and (b) because the resolution was accepted. The final generic inference that is drawn enables the recognition of a TIME RANGE of an event when (1) a time expression is identified for its initiation; (2) a time expression is identified for its termination. An equally correct answer to Q_3 is given by the following paragraph:

> Iran invaded Iraq on 22 September 1980. After eight long years of fighting , UN Resolution 598 ended the war.

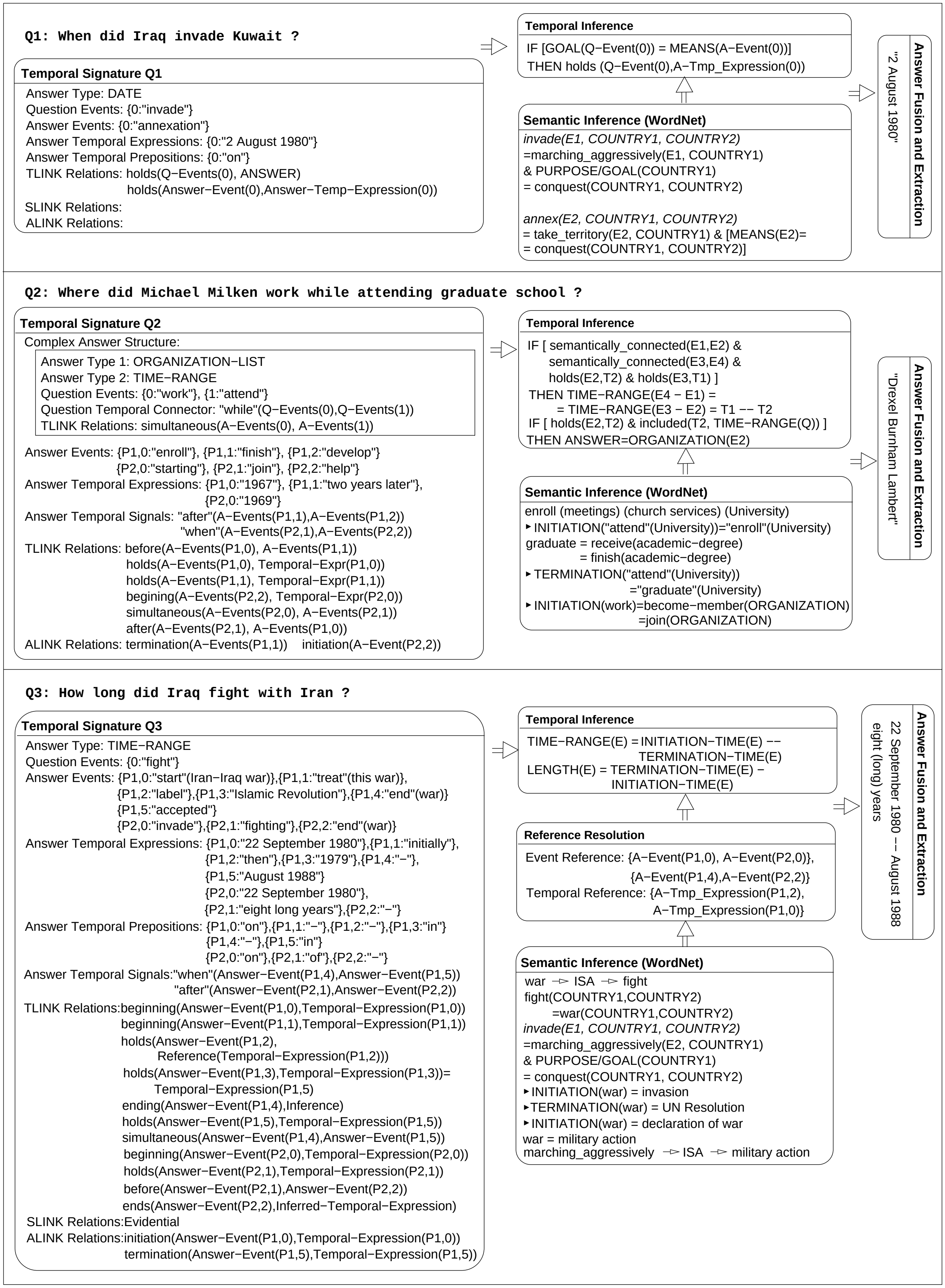

Figure 4: Temporal signatures enabling temporal inference.

Two correct answers are provided by the paragraph. First, the temporal expression *"eight long years"* can be extracted. Second, an inferred, more exact answer, *"22 September 1980 - August 1988"* can be provided. This time, the aspectual relation of INITIATION between the event *"Iran invaded Iraq"* and the time expression *"22 September 1980"* presupposes an implicit event of war between the two countries. This implicit event is referred in the second sentence of the paragraph, in which a TERMINATION aspectual relation is established between the war event, the TIME RANGE *"eight long years"* and the temporal connector *"after"*. The inference that needs to be drawn here is that an event (*"the war"*) terminated after a TIME RANGE had a duration equal to the TIME RANGE. The duration may be also anchored in time, by determining the *"eight long years"*. The initial time stamp of TIME RANGE corresponds to the event having the INITIATION aspect, whereas the ending time stamp corresponds to the event having the TERMINATION aspect.

Figure 4 illustrates the temporal signatures of questions Q_1, Q_2 and Q_3. The signatures contain the expected answer type inferred from the questions: a DATE, an ORGANIZATION-list and a TIME RANGE respectively. In the case of Q_2 the expected answer type is part of a complex answer structure, that is part of the temporal signature. The temporal signatures list all the events mentioned in the question in a sequence of pairs {index:event}. This format helps referring to an event from the list by using its index. Such reference is used for identifying TLINKS, SLINKS and ALINKS. The same format is used for listing time expressions. If more than one candidate answer is considered, the pairs are replaced by triplets {Paragraph,index:event}.

In addition to inferring temporal relations in the temporal signature, we also infer semantic relations between events based on the information available from WordNet, the lexico-semantic knowledge base. The lexico-semantic information from WordNet is used in several ways. First, information encoded in lexico-semantic relations, e.g. IS-A, IS-Part, CAUSE is used directly between predicates. Second, selectional constraints used for defining lexico-semantic predicates are utilized. Third, information from the glosses of synsets is also considered. For example, when processing Q_3, we mine from WordNet a relation between the verbs *"war"* and *"fight"* and we discover an IS-A relation between them. Thus, predicate *"fight"* with COUNTRY1 and COUNTRY2 as arguments entails the predicate *"war"* with the same arguments.

The semantic inference performed when processing Q_2 uses the selectional restrictions of the predicate *"enroll"*, which may apply to *"meetings"*, *"church services"* or a *"University"*. By selecting the third constraint , which is most related to the expression "graduate school" from Q_2, we enable further descriptions of aspectual information related to Q_2.

The semantic inference processed for Q_1 unifies the logical representations of the glosses of *"invade"* and *"annex"*. The unification of the GOAL of one event with the MEANS of the second event constitute the conditions of the temporal inference for Q_1. To supplement the information from WordNet, we have handcrafted a knowledge base for aspectual information. For example, for Q_2, we assigned the initiation aspect of attending a University to the activity of enrolling at that University.

Similarly, the termination of attending a University is marked by the event of graduating. Although the event *"graduate"* is not recognized in any of the relevant paragraphs, it is connected through its gloss (receiving a degree) to the predicate finish (academic-degree) that is identified in the relevant text. Sometimes aspectual information is not derived from handcrafted knowledge, but is inferred from the candidate paragraphs at merging time. In the case of Q_3, both events {*P1,0: "start"(Iran-Iraq war)*} and {*P2,0:"invade"*} have the same time stamp: *"22 September 1980"*. Thus, we can infer that a possibility of initiating a war is through invasion, which is an alternative to the aspectual information derived from the gloss of the sense of war viewed as a legal state. In the later case, the initiation is marked by a declaration of war whereas the termination by an official declaration, a treaty or a resolution (e.g. *UN Resolution 598*). The war, viewed through the WordNet sense #1, is a military action. The genus of the gloss *"invade"*, *"marching aggressively"* is also a form of military action. This semantic information unifies the predicate *"war"* with the predicate *"invade"*, thus enabling the inference that an invasion may initiate a war.

Temporal inference cannot be accomplished without two forms of reference: (a) event reference and (b) temporal reference. The processing of Q_3 involved both forms of reference. Since in the Semantic Inference, the initiation of the war is marked by invasion, Answer events {*P0,1*} and {*P2,0*} co-refer. Similarly, the end of the same war determine the coreference between answer events {*P1,4*} and {*P2,2*}. Such inference is based on the cross-paragraph coreference that establishes that the war refers in both cases to the Iran-Iraq war started on 22 September 1980. It is to be noted that such coreference is different from other forms of within-document nominal coreference that was employed previously in QA systems (e.g. (Vicedo & Ferrández 2000), (Harabagiu & Maiorano 2002)).

Temporal reference resolution involves the identification of the referent *"then"* to the temporal expression *"22 September 1980"*. Temporal reference resolution was produced by adding a few heuristics to COKTAIL (Harabagiu, Bunescu, & Maiorano 2001).

The temporal inference is produced through a set of rules, three of each being illustrated in Figure 4. The rules can be characterized by (1) the presence of predicates, semantic relations and temporal expressions or (2) aspects characterizing an ontology of time, in the vein of the directions described in (Hobbs 2002). The temporal inference used for processing Q_1 and Q_2 falls in the former class, whereas the temporal inference used for Q_3 falls in the latter one. Temporal inference enables the answer extraction for each of the questions listed in Figure 4.

For Q_1, the TLINK relation HOLDS attributes *"2 August 1980"* as the time stamp for the question event indexed as {*0:"invade"*}. The time stamping was possible because semantic inference that is available establishes a connection between the two events listed in the temporal signature of Q_1: *E1=Question-Event{0:"invade"}* and *E2=Answer-Event{0;"annexation"}*. As reported in (Harabagiu, Miller, & Moldovan 1999), the definition of lexical concepts can be translated into Logical Forms Transformations (LFTs) that encode davidsonian representations of actions. The question event {*0:"invade"*} is represented as *invade*(E1, Subject, Object), in which E1 is an argument that stands for the eventuality of the invasion to occur. In our representation E1 is an unique index to the event of invading. The *subject* and *object* arguments are identified in the question: Iraq and Kuwait respectively, which are generalized semantically to COUNTRY1 and COUNTRY2.

In WordNet2, the gloss of *"invade"* is *"marching aggressively for the purpose of conquest"*. This defining gloss is partitioned into (1) the genus = *"marching aggressively"*; and (2) the differentia = *"for the purpose of conquest"*. For this gloss, the genus and the differentia are connected by a PURPOSE/GOAL relation, recognized due to the cue phrase *"for the purpose of"*. Moreover, the PURPOSE/GOAL pertains only to COUNTRY1, whereas the differentia uses the same two arguments as the defined event *"invade"*. The second event from the temporal signature of Q_1 is *"annexation"*, a nominalization derived from the verb *"annex"*. The LFT of *"annex"* is indexed by E2, and it uses as arguments COUNTRY1 and COUNTRY2 as well, since the answer paragraph where *"annexation"* occurs lists Kuwait (COUNTRY2) as one of the arguments of the event, along with Hussein, coerced into Iraq (COUNTRY1). The gloss has the genus expressed by the phrasal verb *"take territory"* and the differentia *"by conquest"*, connected to the genus through a MEANS relation. Like for *"invade"*, the genus of *"annex"* has a single argument: COUNTRY1. Both *"invade"* and *"annex"* events have the same differentia. However, the differentia is reached by different relations: MEANS and PURPOSE respectively.

The analysis of the temporal signature of Q_1 indicates (a) a TLINK HOLDS relation between the answer event and the temporal expression *"2 August 1980"*; (b) the need to establish a TLINK expression between the question event and the only temporal expression from the candidate answer. If semantic inference indicates temporal consistency between the question and answer events, the same TLINK relation can be transferred between the temporal expression which represents the answer and the question event. Temporal consistency is expressed as reaching identical differentia when using the MEANS and PURPOSE relations. It is to be noted that even when (1) only one question event and only one answer event are relevant to the answer; and (2) a single temporal expression is identified as relevant to the answer; different TLINK relations may connect the question and answer events to the temporal expression, and moreover, temporal inference needs to distinguish the most relevant temporal expression from the other temporal expressions present in the candidate answer.

The temporal signature of Q_2 represents a case when (a) multiple events are identified both in the question and the answer; (b) multiple TLINK relations between answer events exists, some identical to the TLINK relation from the question events, but many different than this relation. Multiple question events require the recognition of the temporal relations between the events, which influences (i) the answer type of the question, (ii) a partial answer type or (iii) dictates an alternative question. In Q_2, the relation between the question events is SIMULTANEOUS, indicating that the partial answer type (TIME RANGE) constrains the ANSWER TYPE1 of Q_2 (ORGANIZATION-list). The temporal relation between the question events in question Q_{15} from Figure 2 generates an alternative. The relation between the events in *"How long did Iraq fight with Iran before more than 10,000 soldiers were killed?"* dictates a time period that is limited by the partial answer that represents the date when the loss of soldiers reached the 10,000 mark. In the temporal signature of Q_2, a SIMULTANEOUS relation exists between the question events *Q-Event(0)* indexed as {*0:"work"*} and *Q-Event(1)* indexed as {*1:"attend"*}. Figure 5(a) illustrates the relations between the question events and the TIME RANGE that constitutes the partial answer. Since the question asks

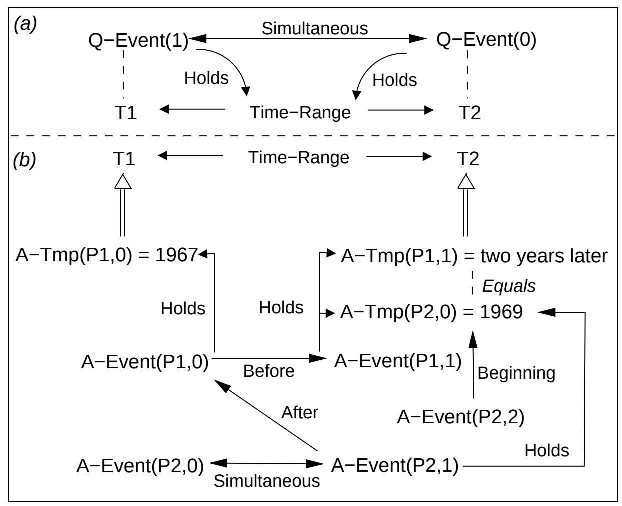

Figure 5: (a) Relations between events and time in the question; (b) Relations between events and time in the answer.

about organizations related to *Q-Event(0)* it is important to find connections to *Q-Event(1)* and *Q-Event(0)* to the answer events represented in Figure 5(b). The semantic inference derived from WordNet determines these connections: *A-Event(P1,0:"enroll")* is related to *Q-Event(1:"attend")* and also to *A-Event(P1,1:"finish")*. These relations have an aspectual component, since the INITIATION of *A-Event(P1,0:"enroll")* enables *Q-Event(1:"attend")*. The TERMINATION of *Q-Event(1:"attend")*, when applied to the argument University is marked by the event of graduation which is glossed as *"finish(academic-degree)"*, and is therefore an instantiation of *A-Event(P1,1:"finish")*. As *A-Event(P2,0)* and *A-Event(P2,1)* are simultaneous and holding to *"1969"*, which is within the TIME RANGE, they are relevant to the question. The relevance is also dictated by the semantic relation between INITIATION(*work*) and *join*(ORGANIZATION).

Finally the temporal inference assigns the TIME RANGE (T1 – T2 from Figure 5) when there are semantic connections (as those inferred from WordNet) between the question events and the answer events. The role of E1 in Figure 4 for the temporal inference of Q_2 is played by *Q-Event(0)("work")* in Figure 5, whereas E2 is *A-Event(P2,1)("join")*. The role of E3 is *A-Event(P1,0)("enroll")* and E4 is *Q-Event(1)("attend")*. When the TIME RANGE is defined, the answers are extracted based on the constraint that they have to be organizations related to E2 (*"join"*).

In order to answer a time range question like Q_3 *"How long did Iraq fight with Iran?"*, QA systems must be able to process temporal information associated with multiple answer events in order to calculate the duration of the single event mentioned in the question. In the temporal signature for Q_3, three types of TLINK relations can be identified between events mentioned in answer passages: (i) beginning relations, which identify events associated with the start of the Iran-Iraq conflict, (ii) ending relations, associated with conclusion of the fighting, and (iii) holding relations, which can be used to characterize the inherent duration of the event itself. In Q_3's temporal signature, information extracted from adverbials and prepositions associated with answer events is used to infer that *A-Event(P1,0): "start"(Iran-Iraq war)* is associated with the start of the question event and likewise, *A-Event(P1,5): "accepted"* represents an event associated with the end of the question event. Semantic inferences obtained from WordNet allow for the recognition that the *"fight"* mentioned in Q_3 represents a ontological category which includes a subtype, war, which denotes a fight between two countries. As with Q_1, mention of an *"invasion"* event between two countries can lead to the recognition that the invading country's goal is *"conquest"* of the invaded country; by recognizing that *"conquest"* represents another type of *"fighting"* event (like *"war"*), we can infer that one possible way of initiating a war is invasion. A similar process can be used to infer that the termination of a fighting event can result from an *"acceptance"* event between two countries. Once initiation and termination events are identified for the question event, the exact duration of the event can either be calculated (by comparing the times associated with each boundary event), or be extracted (by identifying answer passages that denote durations).

Conclusions

In this paper, we introduced a methodology for computing temporal inference for QA that allows for the enhanced recognition of exact answers to a variety of questions about time. We have argued that answering questions about temporal information requires several different forms of inference, including inferences that derive from relations between events, their arguments, and the time anchors available in the discourse context. We have demonstrated that the temporal annotations produced as part of TimeML can be used to generate temporal signatures which, when combined with sources of semantic inference and information about coreferring events, can be used to formulate sophisticated temporal inferences that can be used to identify exact answers to temporal questions.

References

Filatova, E., and Hovy, E. 2001. Assigning Time-Stamps to Event-Clauses. In *Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing*, 88–95.

Harabagiu, S. M., and Maiorano, S. 2002. Three Ways to Customize Reference Resolution. In *Proceedings of the 2002 International Symposium on Reference Resolution for Natural Language Processing*, 17–24.

Harabagiu, S. M.; Moldovan, D. I.; Pasca, M.; Mihalcea, R.; Surdeanu, M.; Bunescu, R. C.; Girju, R.; Rus, V.; and Morarescu, P. 2001. The Role of Lexico-Semantic Feedback in Open-Domain Textual Question-Answering. In *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics (ACL-2001)*, 274–281.

Harabagiu, S. M.; Bunescu, R.; and Maiorano, S. 2001. Text and Knowledge Mining for Coreference Resolution. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association of Computational Linguistics (NAACL-2001)*, 55–62.

Harabagiu, S.; Miller, G.; and Moldovan, D. 1999. WordNet 2 - A Morphologically and Semantically Enhanced Resource. In *Proceedings of SIGLEX-99*, 1–8.

Hobbs, J., and Pustejovsky, J. 2003. Annotating and Reasoning about Time and Events. In *Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*.

Hobbs, J. 2002. Toward an Ontology of Time for the Semantic Web. In *Proceedings of the LREC-2002 Workshop on Annotation Standards for Temporal Information in Natural Language*.

Ingria, B., and Pustejovsky, J. 2002. TimeML: A Formal Specification Language for Events and Temporal Expressions.

Setzer, A. 2001. *Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study*. Ph.D. Dissertation, University of Sheffield, Sheffield, UK.

Vicedo, J., and Ferrández, A. 2000. Importance of Pronominal Anaphora Resolution in Question Answering Systems. In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics (ACL-00)*, 555–562.