Abductive Processes for Answer Justification

Sanda M. Harabagiu
Language Computer Corporation
Dallas TX 75206 USA
sanda@languagecomputer.com

Steven J. Maiorano
University of Sheffield
Sheffield S1 4DP UK
stevemai@home.com

Abstract

An answer mined from text collections is deemed correct if an explanation or justification is provided. Due to multiple language ambiguities, abduction, or inference to the best explanation, is a possible vehicle for such justifications. To interpret abductively an answer, world knowledge, lexico-semantic knowledge as well as pragmatic knowledge need to be integrated. This paper presents various forms of abductive processes taking place when answers are mined from texts and knowledge bases.

Introduction

In (Harabagiu and Maiorano 1999) we claimed that finding answers from large collections of texts is a matter of implementing paragraph indexing and applying abductive inference. Subsequent experiments, reported in (Harabagiu et al.2000) have showed that in fact, answer justification provided by lightweight abduction has the merit of improving the accuracy of a Question Answering (QA) system by over 20%. These results were later confirmed in the TREC-9[1] evaluations (Voorhees 2001), where LCC-SMU's system was the only one that implemented abductive processes for answer justification.

To evaluate the results of a QA system, the TREC results were considered as five possible answers, of either 50 bytes of length (short answers) or 250 bytes of length (long answers). The scoring, detailed in (Voorhees 2001), uses a Mean Reciprocal Rank (MRR) by relying on only six values:$(1, .5, .33, .25, .2, 0)$, representing the score each set of five answers obtains. If the first answer is correct, it obtains a score of 1, if the second one is correct, it is scored with .5, if the third one is correct, the score becomes .33, if the fourth is correct, the score is .25 and if the fifth one is correct, the score is .2. Otherwise, it is scored with 0. No credit is given if multiple answers are correct. As shown in Figure 1, the accuracy of the five answers returned by the LCC-SMU system was substantially better than those of other systems. (Paşca and Harabagiu 2001) reports the observation that abductive processes (a) double the precision of QA since they filter out unjustifiable answers, and (b) have the same overall effect as the lexico-semantic alternations of keywords used to search the answer. Given this significant contribution of abductive inference to the accuracy of QA systems, we take issue in this paper on the formalization of several abductive processes and their general interaction with the larger concept of *context*, primarily as viewed by Graeme Hirst in (Hirst 2000).

In (Hobbs et al.1993) abduction is defined as inference to the best explanation. This is because abduction is different from deduction or induction. In deduction, from $(\forall x)(a(x) \rightarrow b(x))$ and $a(A)$, one concludes $b(A)$. In induction, from a number of instances $a(A)$ and $b(A)$, one concludes $(\forall x)(a(x) \rightarrow b(x))$. Abduction is the third possiblity, as noted in (Hobbs et al.1993). From $(\forall x)(a(x) \rightarrow b(x))$ and $b(A)$, one concludes $a(A)$. If $b(A)$ is the observable evidence, then $(\forall x)(a(x) \rightarrow b(x))$ is the general principle that can explain $b(A)$'s occurrence, whereas $a(A)$ is the inferred cause or explanation of $b(A)$. For QA, if $b(A)$ stands for a candidate answer of question $a(A)$, we need to discover the implication $(\forall x)(a(x) \rightarrow b(x))$. In state-of-the-art QA systems, there are several different abductive processes that concur to the discovery of the justification of an answer.

Abductive processes are determined by (1) lexico-semantic knowledge that enables the retrieval of candidate answers; (2) logical axioms that are generated from various sources (e.g. WordNet (Fellbaum 1998)); (3) pragmatic knowledge, allowing bridging inference; and (4) several forms of coreference information. We stress out the fact that for each answer, the lexico-semantic, logical, pragmatic and coreferential knowledge spans a relatively small search space, thus enabling efficient abductive processing. Typically, answers are mined from short paragraphs containing only a few sentences. Moreover, the lexico-semantic information is determined by the keywords employed to retrieve the paragraphs containing candidate answers. As reported in (Harabagiu et al.2001), alternations of these keywords as well as the topically related concepts of these keywords do not increase substantially the search space for abductive

[1]The TExt Retrieval Conference (TREC) is the main research forum for evaluating the performance of Information Retrieval(IR) and Question Answering (QA) Systems. TREC is sponsored by the National Institute of Standards and Technology (NIST) and the Defense Advanced Research Projects Agency (DARPA) to annually conduct IR and QA evaluations.

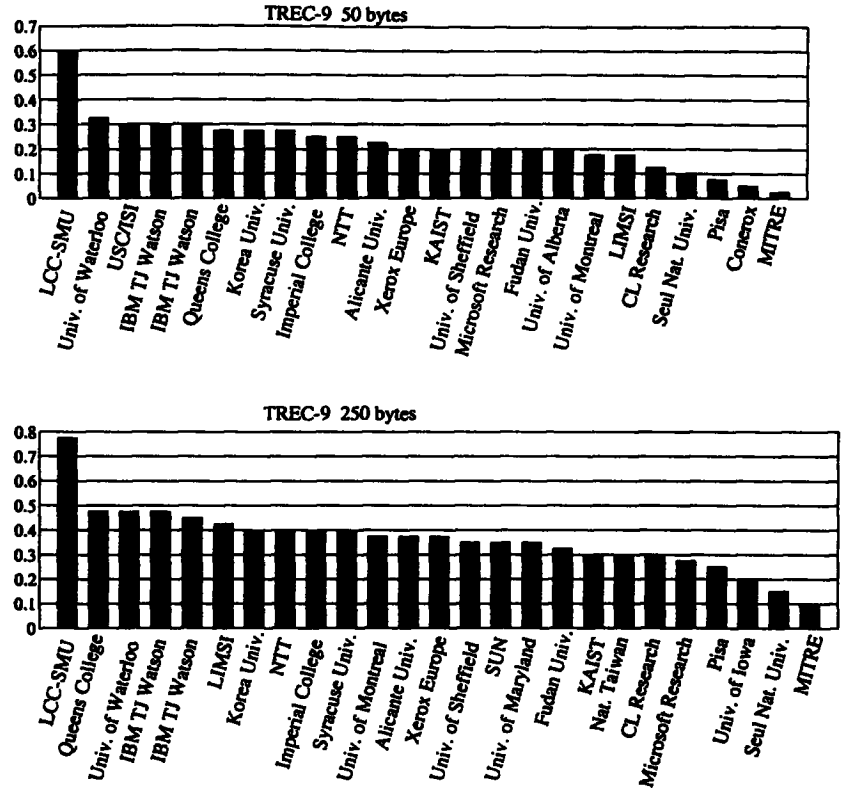

Figure 1: Results of the TREC-9 evaluations.

processing - on the contrary- they guide the abductive processes.

This paper is organized as follows. Section 2 describes the answer mining techniques that operate on large on-line text collections. Section 3 presents the techniques used for mining answers from lexical knowledge bases. Section 4 details several abductive processes and their attributes. Section 5 discusses the advantages and differences of using abductive justification as opposed to quantitative ad-hoc information provided by redundancy. Section 6 presents results of our evaluations whereas Section 7 summarizes the conclusions.

Answer Mining from On-line Text Collection

There are questions whose answer can be mined from large text collections because the information they request is explicit in some text snippet. Moreover, for some questions and some text collections, it is easier to mine a question in a vast set of texts rather than a knowledge base. In this section we describe several abductive processes that account for the mining of answers from text collections:

Appositions

Appositive constructions are widely used in Information Extraction tasks or coreference resolution. Similarly, they provide a simple and reliable way for the abduction of answers from texts. For example, the question *"What is Moulin Rouge?"* evaluated in TREC-10 is answered by the apposition following the named entity: *Moulin Rouge, the famous Parisian nightclub.* If appositions are not recognized and the mining of answers is solely based on keywords, many false positives may be encountered. For example, when searching on GOOGLE for `''Moulin Rouge''` a vast majority of the URLs contain information about the 2001 movie entitled "Moulin Rouge", starring Nicole Kidman and Ewen McGregor. Due to this, the page-ranking algorithm used by GOOGLE categorizes the named entity as the title of a movie. Linguistically, we are dealing in this case with a *metonymy*, since movies get titles because of the main place of themes they depict, in this case the Parisian cabaret. However, to be able to answer the question, a long chain of inference is needed. One of the reviews of the movie says: *" "Resurrects a moment in the history of cabaret when spectacle ravished the senses in an atmosphere of bohemian abandon."*. The elliptic subject of the review refers to the movie. Because of this reference, the first clause introduces the theme of the movie: "a moment in the history of cabaret". This enables the coercion of the name of the movie into the type "cabaret", leading to the same answer that was mined through an apposition from the TREC collection. As opposed to a simple recognition of an apposition, this example shows the special inference chains imposed by on-line documents.

Copulative expressions

Copulative expressions are used for defining features of a wide range of entities. This explains why, for so-called definition questions, recognized by one of the patters:

(Q-P1): *What {is|are} <phrase_to_define>?*
(Q-P2): *What is the definition of <phrase_to_define>?*
(Q-P3): *Who {is|was|are|were} < person_name(s)>?*

the mining of the answer is done by matching some of the following patterns,as reported in (Paşca and Harabagiu 2001):

(A-P1): [*<phrase_to_define>* {*is|are|became*}]
(A-P2): [*<phrase_to_define>*, {*a|the|an* }]
(A-P3): [*<phrase_to_define>* –]

A special observation needs to be made with respect to the abduction of answers for definition questions. Another apposition, from the same TREC collection of 3 GBytes of text data, corresponding to the question asking about Moulin Rouge is *Moulin Rouge, a Paris landmark.* At this point, additional pragmatic knowledge needs to be used, in order to discard this second candidate answer due to its level of generality. One needs to know that the Notre Dame and the Eiffel Tower are also Paris landmarks, but however Moulin Rouge is a cabaret, the Notre Dame is a cathedral whereas the Eiffel Tower is a monument. Therefore the definition of the Moulin Rouge as a Parisian landmark is too general, since abductions classifying it as a cathedral or monument would be legal, but untruthful.

Unfortunately, pragmatic knowledge is not always readily available, and thus often we need to rely on classification information available either from categorizations provided by popular search engines like GOOGLE, or from ad-hoc categorizations determined by hyperlinls on various Web pages. For example, when searching GOOGLE for `''landmark + Paris''`, the resulting pointers comprise Web pages discussing the Arc de Triomphe, the Eiffel Tower but also the Hotel Intercontinental, boasted as a Paris landmark, or the Maiile Way mustard shop or even the products of Iguana, a company manufacturing heirloom glass ornaments representing Parisian landmarks such as the Eiffel Tower, the Notre Dame or the Arc de Triomphe. Furthermore, lexical knowledge bases such as WordNet do not classify a cabaret as a landmark, but rather as a place of entertainment.

Possessives

Possessive constructions are useful textual cues for answer mining. For example, when searching for the answer to the TREC-10 question *"Who lived in the Neuschwanstein castle?"*, the possessive *"Mad King Ludwig's Neuschwanstein Castle* indicates a possible abduction, due to the multiple interpretations of the semantic of possessive constructs. In the case of a house, even if luxurious as a castle, the "possessor" is either the architect, the owner or the person that lives in it. Since the question specifically asked for the person who lived in the castle, the other two possibilities of interpretation are filtered out unless there is evidence that King Ludwig built the castle but did not live in it.

Possessive constructions are sometimes combined with appositions, like in the case of the answer to the TREC-10 question *"What currency does Argentina use?"*, which is *"The austral, Argentina's currency"*. However, mining the answer to a very similar TREC-10 question *"What currency does Luxemburg use?"* relies far more on knowledge of other currencies than on pure syntactic information. In this case, the answer is provided by the complex nominal *"Luxemburg franc"*, that is recognized correctly because the noun *"franc"* is categorized as CURRENCY by a Named Entity tagger similar to those developed for the Message Understanding Conference (MUC) evaluations. If such a resource is not available and the answer mining relies only on semantic information from WordNet, the abduction cannot be performed, since in WordNet 1.7, the noun *"franc"* is classified as a *"monetary unit"*, a concept that has no relation to the *"currency"* concept.

Meronyms

Sometimes questions contain expressions like *"made of"* or *"belongs to"* which are indicative of meronymic relations to one of the entities expressed in the question. For example, in the case of the TREC-10 question *"What is the statue of liberty made of?"*, even if Statue of Liberty is not capitalized properly, it is recognized as the New York landmark due to its entry in WordNet. The same entry is sub-categorized as a statue, and the name of its location *"New York"* can be retrieved from its defining gloss. However, WordNet does not encode any meronymic relations of the type IS-MADE-OF. Instead, the meronymic relation is implied by the prepositional attachment *"the 152-foot copper statue in New York"*, where the nominal expression corefers to the Statue of Liberty in the same TREC document. Since *"copper"* is classified as a substance in WordNet, it allows the abduction of the metonymy between copper and the Statue of Liberty, thus justifying the answer.

A second form of meronymy is illustrated by the TREC-10 question *"What county is Modesto, California in?"*. The processing of this question presupposes the recognition of *Modesto* as a city name and of *California* as a state name, which is a function easily performed by most named entity recognizers. Usually the names of US cities are immediately followed by the name or the acronym of the state they belong to - to avoid ambiguity between different cities with the same name, but located in different states, e.g. Arlignton, VA and Arlington, TX. However, the question asks about the name of a county, not for the name of the state. If gazeteers of all US cities, their counties and states are not available, an IS-PART relation is inferred from the interpretation of prepositional attachments determined by the preposition *"in"*.. For example, the correct answer to the above question is indicated by the following attachment: *"10 miles west of Modesto in Stanislaus County.*

A third form of meronymy results from recipes or cooking directions. It is commonsense knowledge that any food is made of a series of ingredients, and the general format of any recipe will list the ingredients along with their measurements. This insight allows the mining of the answer to the TREC-10 question *"What fruit is Melba sauce made from?"*. The expected answer is a kind of fruit. Both peaches and raspberries are fruits. However, we cannot abductively prove that *"peach"* is the correct answer, because the text snippet *"it is used as sauce for Peach Melba"* , indi-

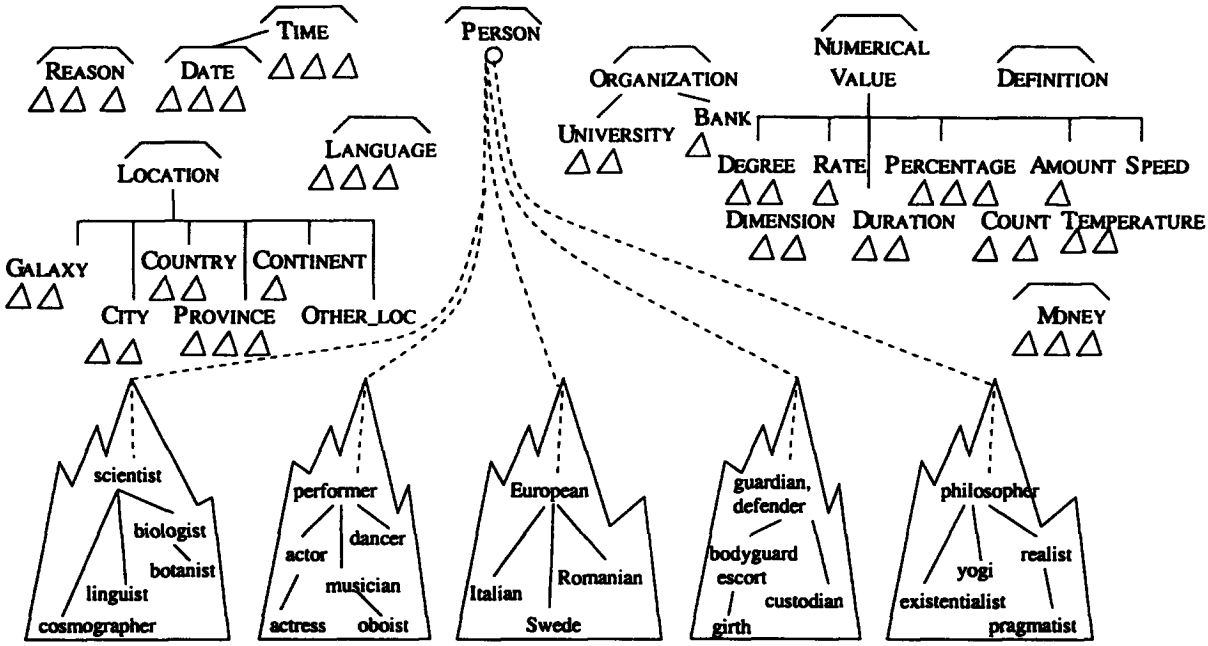

Figure 2: Answer Type Taxonomy

cates that the sauce, which corefers with Melba sauce in the same text, is just in ingredient of the Peach Melba. However, the text snippet *"2 cups of pureed raspberries"* extracted from the recipe of Melba sauce provides evidence it is the correct answer.

Answer Mining from Lexical Knowledge Bases

Lexical knowledge bases organized hierarchically represent an important source for mining answers to natural language questions. A special case is represented by the WordNet database, which encodes a vast majority of the concepts from the English language, lexicalized as nouns, verbs, adjectives and adverbs. What is special about WordNet is the fact that each concept is represented by a set of synonym words, also called a *synset*. In this way, words that have multiple meanings belong to multiple synsets. However, word sense disambiguation is not hindering the process of mining answers from lexico-semantic knowledge bases because question processing enables a fairly accurate categorization of the expected answer type.

When open-domain natural language questions inquire only about entities or events or some of their attributes or roles, as was the case with the test questions in TREC-8, TREC-9 and the main task in TREC-10, an off-line taxonomy of answer types can be built by relying on the synsets encoded in WordNet (Fellbaum 1998). (Paşca and Harabagiu 2001) describes the semi-automatic procedure for acquiring a taxonomy of answer types, similar to the one illustrated in Figure 2.

An answer type hierarchy enables the mining of the answer of a wide range of questions. For example, when searching for the answer to the question *"Who defeated Napoleon at Waterloo?"* it is important to know that the expected answer represents some national armies, let by some individuals. In WordNet 1.7, the glossed definition of the synset {*Waterloo, battle of Waterloo*} is *(the battle on 18 June 1815 in which Napoleon met his final defeat; Prussian and British forces under Blucher and the Duke of Wellington routed the French forces under Napoleon)*. The exact answer, extracted from this gloss is *"Prussian and British forces under Blucher and the Duke of Wellington"*, representing a noun phrase, whose head, Prussian and British forces, is coerced from the names of the two nationalities, that are encoded as concepts in the answer type hierarchy. However, since the expected answer type is PERSON, several entities from the gloss are candidates for the answer: *"Napoleon"*, *" Prussian and British forces"*, *"Blucher"*, *"the Duke of Wellington"* and *"the French forces"*. If *"Napoleon"*, *"Prussian"*, *"British"* and *"French"* are encoded in WordNet, it is to be noted that *"Blucher"* and *"the Duke of Wellington"* are not. Nevertheless, they are recognized as names of persons due to the usage of a Named Entity recognizer.

Since the verb *"defeat"* is not reflexive, both *"Napoleon"* and *"the French forces"* are ruled out, the second due to the prepositional attachment *"the French forces under Napoleon"*. Because the only remaining candidates are syntactically dependent through the coordinated prepositional attachment *"Prussian and British forces under Blucher and the Duke of Wellington"* they represent the exact answer. If a shorter exact answer is wanted, a philosophical debate over who is more important- the general or the army needs to be resolved. From the WordNet gloss, *"Prussian and British forces"* is recognized as the genus of the definition of Waterloo - and thus stands for an answer mined from WordNet. Mining this answer from WordNet is important, as it

Answer 1. Score: 280.00 The family gained prominence by financing such massive projects as European railroads and the British military campaign that led to the final defeat of Napoleon Bonaparte at Waterloo in June 1815
Answer 2. Score: 244.69 In 1815 , Napoleon Bonaparte met his Waterloo as British and Prussian troops defeated the French forces in Belgium
Answer 3. Score: 244.69 In 1815 , Napoleon Bonaparte met his Waterloo as British and Prussian troops defeated the French forces in Belgium
Answer 4. Score: 146.14 In 1815 , Napoleon Bonaparte arrived at the island of Saint Helena off the west African coast where he was exiled after his loss to the British at Waterloo
Answer 5. Score: 132.69 Subtitled " World Society 1815 - 1830 , " the book concentrates on the period that opens with Napoleon 's defeat at Waterloo and closes with the election of Andrew Jackson as president of the U.S. Science and technology are as important to his story as the rise and fall of empires

Table 1: Answers returned by LCC's QAS for the question "Who defeated Napoleon at Waterloo?"

Answer 1. Score: 888.00 Not surprisingly , Leningrad 's Communist Party leadership is in the forefront of those opposed to changing it back again
Answer 2. Score: 856.00 The vehicle plant named after Lenin 's Young Communist League and known as AZKL , shut its car production line yesterday after supplies of engines from
Answer 3. Score: 824.00 " I 'm against changing the name of the city , " Yuri P. Belov , a secretary of the Leningrad regional Communist Party committee , said in an interview
Answer 4. Score: 820.69 Expressing the interests of the working class and all working people and relying on the great legacy of Marx , Engels and Lenin , the Soviet Communist Party is creatively developing socialist
Answer 5. Score: 820.69 In both Moscow and Leningrad , the local Communist Party has taken control of the newspapers it originally shared with

Table 2: Answers returned by LCC's QAS for the question "Who was Lenin?"

can determine a better scoring of QA systems that mine answers only from text collections. For example, if this answer were known, it would have improved the precision of LCC's Question Answering Server (QAS^{TM}), that produced the answers listed in Table 1. In this case, answers 4 and 5 would have been filtered out and answers 2 and 3 would have taken precedence over answer 1.

Definition questions

Mining answers from knowledge bases has even more importance when processing definition questions. The percentage of questions that ask for definitions of concepts, e.g. *"What are capers?"* or *"What is an antigen?"* represented 25% of the questions from the main task of TREC-10, an increase from a mere 9% in TREC-9 and 1% in TREC-8 respectively. The definition questions normally require an increase in the sophistication of the QA system. This is determined by the fact that questions asking about the definition of an entity use few other words that could guide the semantics of the answer mining. Moreover, for a definition, there will be no expected answer type. A simple question such as *"Who was Lenin?"* can be answered by consulting the WordNet knowledge base rather than a large set of texts like the TREC collection. Lenin is encoded in WordNet, in a synset that lists all his name aliases: {*Lenin, Vladimir Lenin, Nikolai Lenin, Vladimir Ilyich Ulyanov*}. Moreover, the gloss of this synset lists the definition: *(Russian founder of the Bolsheviks and leader of the Russian Revolution and first head of the USSR (1870-1924))*. This is a correct answer, and furthermore, a shorter answer can be obtained by simply accessing the hypernym of the Lenin entry, revealing that he was a communist. When mining the answer for the same question from large collections of texts, as illustrated in Table 1 no correct complete answers can be found, as the information required by the answer is not explicit in any text, and thus only related information is found.

Q912: What is epilepsy?
Q1273: What is an annuity?
Q1022: What is Wimbledon?
Q1152: What is Mardi Gras?
Q1160: What is dianetics?
Q1280: What is Muscular Distrophy?

Table 3: Examples of definition questions.

Several additional definition questions that were evaluated in TREC-10 are listed in Table 3. For the some those questions the concept that needs to be defined is not encoded in WordNet. For example, both *diametics* from Q1160 and *Muscular Distrophy* from Q1280 are not encoded in WordNet. But whenever the concept is encoded in WordNet, both its gloss and its hypernym help determine the answer. Table 4 illustrates several examples of definition questions that are resolved by the substitution of the concept with its hypernym. The usage of WordNet hypernyms builds on the con-

jecture that any concept encoded in a dictionary like WordNet is defined by a *genus* and a *differentia*. Thus when asking about the definition of a concept, retrieving the genus is sufficient evidence of the explanation of the definition, especially when the genus is identical with the hypernym.

Concept that needs to be defined	*Hypernym from* WordNet	*Candidate answer*
What is a shaman?	{priest, non-Christian priest}	Mathews is the priest or shaman
What is a nematode?	{worm}	nematodes, tiny worms in soil.
What is anise?	{herb, herbaceous plant}	aloe, anise, rhubarb and other herbs

Table 4: WordNet information employed for minings answers of definition questions.

In WordNet only one third of the glosses use a hypernym of the concept being defined as the genus of the gloss. Therefore, the genus, processes as the head of the first Noun Phrase from the gloss, can also be used mine for answers. For example, the processing of question Q1273 from Table 3 relies on the substitution of *annuity* with *income* the genus of its WordNet gloss, rather than its hypernym, the concept *regular payment*. The availability of the genus or hypernym helps also the processing of definition questions in which the concept is a named entity, as it is in the case of questions Q1022 and Q1152 from Table 3. In this way *Wimbledon* is replaced with *a suburb of London* and *Mardi Gras* with *holiday*.

Pertainyms

When a concept A is related by some unspecified semantics to another concept B, it is said that A pertains to B. WordNet encodes such relations, which prove to be very useful for mining answers for natural language questions. For example, in TREC-10, when asking *"What is the only artery that carries blood from the heart to the lungs'?"*, the correct answer is mined from WordNet because the adjective *"pulmonary"* pertains to lungs, therefore the Pulmonary Artery is the expected answer. Knowing the answer makes the retrieval of a text snippet containing it much easier. The textual answer from the TREC collection is: *"a connection from the heart to the lungs through the pulmonary artery"*.

Bridging Inference

Questions like *What is the esophagus used for?"* or *"Why is the sun yellow?"* are difficult to process because the answer relies on expert knowledge, from medicine in the former example, and from physics in the latter one. Nevertheless, if a lexical dictionary that explains the definitions of concepts is available, some supporting knowledge can be mined. For example, by inspecting WordNet (Fellbaum 1998), in the case of *esophagus* we can find that it is *"the passage between the pharynx and the stomach"*. Moreover, WordNet encodes several relations, like *meronymy*, showing that the esophagus is part of the *digestive tube* or *gastrointestinal tract*. The glossed definition of the *digestive tube* shows that one of its function is the *digestion*.

The information mined from WordNet guides several processes of bridging inference between the question and the expected answer. First the definition of the concept defined by the WordNet synonym set {*esophagus, gorge, gullet*} indicates its usage as a *passage* between two other body parts: the pharynx and the stomach. Thus the query " *esophagus AND pharynx AND stomach"* retrieves all paragraphs containing relevant connections between the three concepts, including other possible functions of the esophagus. When the query does not retrieve relevant paragraphs, new queries combining esophagus and its holonyms (i.e. gastrointestinal tract) or functions of the holonyms (i.e. digestion) retrieve the paragraphs that may contain the answer. To extract the correct answer, the question and the answer need to be semantically unified.

The difficulty stands in resolving the level of precision required by the unification. Currently, LCC's QAS^{TM} considers an acceptable unification when (a) a textual relation can be established between the elements of the query matched in the answer (e.g. esophagus and gastrointestinal tract); and (b) the textual relation is either a syntactic dependency generated by a parser, a reference relation or it is induced by matching against a predefined pattern. For example, the pattern "X, particularly Y" accounts for such a relation, granting the validity of the answer *"the upper gastrointestinal tract, particularly the esophagus"*. However, we are aware that such patterns generate multiple false positive results, degrading the performance of the question-answering system.

Abductive Processes

In (Hobbs et al.1993) abductive inference is used as a vehicle for solving local pragmatics problems. To this end, a first-order logic implementing the Davidsonian reification of eventualities (Davidson 1967). The ontological assumptions of the Davidsonian notation was reported in (Hobbs et al.1993). This notation has two advantages: (1) it is as close as possible to English; and (2) it localizes syntactic ambiguities, as it was previously shown in (Bear and Hobbs 1988). Moreover, the same notation is used in a larger project first outlined in (Harabagiu et al.1999) that extends the WordNet lexical database (Fellbaum 1998). The extension involves the disambiguation of the defining gloss of each concept as well as the transformation of each gloss into a logical form. The Logic From Transformation (LFT) codification is based on syntactic-based relations such as: (a) syntactic subjects; (b) syntactic objects; (c) prepositional attachments and (d) adjectival/adverbial adjuncts.

The Davidsonian treatment of actions considers that all predicates lexicalized by verbs or nominalizations have three arguments *action/state/event-predicate(e, x, y)*, where:

- *e* represents the *eventuality* of the action, state or event to have taken place;
- *x* represents the agent of the action, represented by the syntactic *subject* of the action, event or state;
- *y* represents the syntactic direct *object* of the action, event or state.

In addition, we express only the eventuality of the actions, events or states, considering that the subjects, objects or other themes expressed by prepositional attachments exist

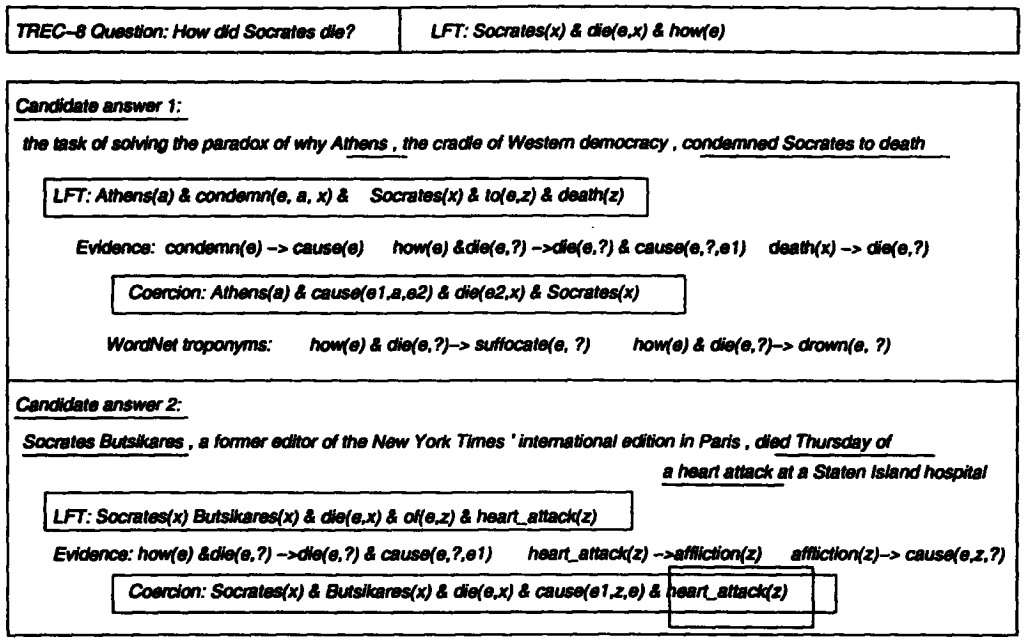

Figure 3: Abductive coercion of candidate answers.

by virtue of actions they participate in. For example, the LFT of the gloss *(a person who inspires fear)* defining the WordNet *synset*[2] {*terror, scourge, threat*} is: [*person(x) & inspire(e, x, y) & fear(y)*]. We note that verbal predicates can sometimes have more than three arguments, e.g. in the case of bitransitive verbs. For example, the gloss of the synset {*impart, leave, give, pass on*} is *(give people knowledge)*, with the corresponding LFT:[*give(e, x, y, z) & knowledge(y) & people(z)*], having the subject unspecified.

Since most thematic roles are represented as prepositional attachments, predicates lexicalized by prepositions are used to account for such relations. The preposition predicates have always two predicates: the first argument corresponds to the head of the prepositional attachment whereas the second argument stands for the object that is being attached. The predicative treatment of prepositional attachments was first reported in (Bear and Hobbs 1988). In the work of Bear and Hobbs, the prepositional predicates are used as means for localizing the ambiguities generated by possible attachments. In contrast, when LFTs are generated for WordNet glosses, all syntactic ambiguities are resolved by a high-performance probabilistic parser trained on the Penn Treebank (Marcus et al.1993). However, the usage of preposition predicates in the LFT is very important, as it suggests the logic relations determined by the combination of syntax and semantics. Paraphrase matching can thus be employed to unify relations bearing the same meaning, and thus enabling justifications. The paraphrases are correctly matched also because of the lexical and semantic disambiguation that is produced concurrently with the parse. The probabilistic parser also correctly identifies base NPs, allowing complex nominals to be recognized and transformed in a sequence of predicates. The predicate *nn* indicates a complex nominal, whereas its arguments correspond to the components of the complex nominal. For example, the complex nominal *baseball team* is transformed in [*nn(a, b) & baseball(a) & team(b)*].

The rules that generate LFTs were reported in (Moldovan and Rus). The transformation of WordNet glosses into LFTs is important because they enable two of the most important criteria needed to select the best abduction. These criteria were reported in (Hobbs et al.1993) and are inspired by Thagard's principles of *simplicity* and *consilience* (cf. (Thagard 1978)). Roughly, *simplicity* requires that the chain of explanations should be as small as possible, whereas *consilience* requires that the evidence triggering the explanation be as large as possible. In other words, for the abductive processes needed in QA, we expect to have as much evidence as possible to infer from a candidate answer the validity of the question. We also expect to make as few inferences as possible. All the glosses from WordNet represent the search space for evidence supporting an explanation. If at least two different justifications are found, the consilience criterion requires to select the one having stronger evidence. However, the quantification of evidence is not a trivial matter.

For the example illustrated in figure 3, the evidence used to justify the first answer of the TREC-8 question comprises axioms about the definition of the *condemn* concept, the definition of the *how* question class and the relation between the nominalization *death* and its verbal root *die*. The LFT of the first answer, when unified with these axioms produces a coercion that cannot be unified with the LFT of the question because the causality relation is reversed: the answer explains why Socrates dies and not how he did die. The second candidate answer however contains the right direc-

[2]A set of synonyms is called a synset.

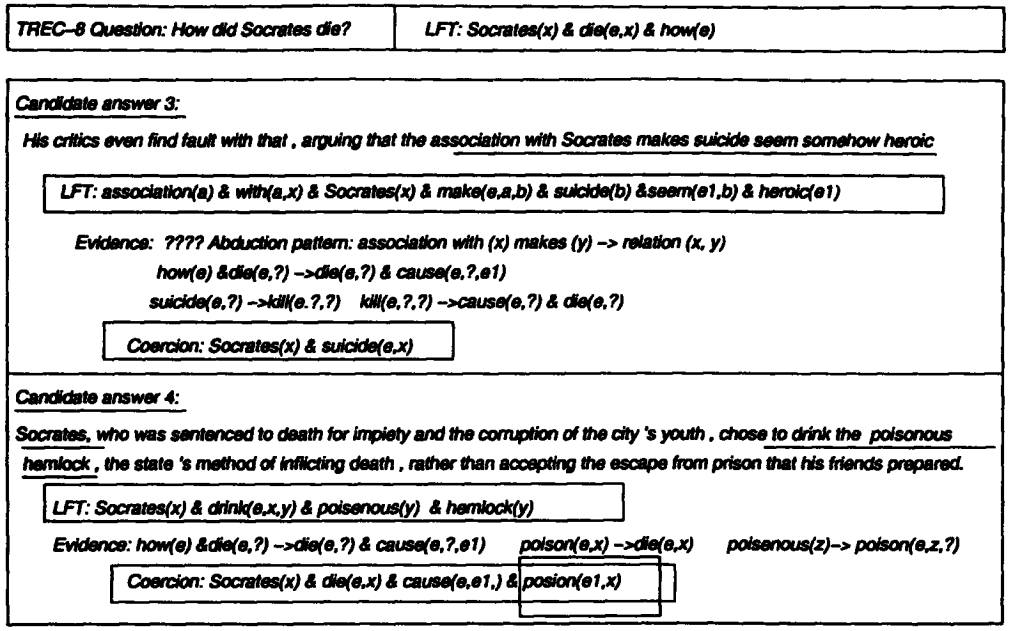

Figure 4: Abductive justification of candidate answers.

tion of the causality imposed by the semantic class of the question and it is considered a correct, although surprising answer. This is due to the ambiguity of the question - it is not clear whether we are asked about the more famous Socrates, i.e. the philosopher, or about somebody whose first name is Socrates. The coercion of the first candidate answer could not use any of the WordNet evidence provided by the hyponyms of the verb *die*.

The simplicity criterion is satisfied because the coercion of the answer always uses the minimal path of backward inference. Hobbs considers that in abduction, the balance between simplicity and consilience is guaranteed by the high degree of redundancy typical in natural language processing. However, we note that redundancy is not always enough. An example is illustrated in Figure 4 by the third candidate answer to the same TREC question. Without an abductive pattern recognizing a relation between concept X and concept Y in the expression *"association with X makes Y"* it is practically impossible to justify this answer, which is correct. The fourth candidate answer is also correct, and furthermore it is a more specific answer. However, only by using the axiomatic knowledge derived from WordNet through the LFTs, it is impossible to show that the third answer is more general than the fourth one - therefore it is hard to quantify the simplicity criterion - due to the generative power of natural language, which also has to be taken into consideration when implementing abductive processes.

When implementing abductive processes, we considered the similarity between the simplicity criterion and the notion of *context simpliciter* discussed in (Hirst 2000). Hirst notes that context in natural language interpretation cannot be oversimplified and considered only as a source of information - its main functionality when employed in knowledge representation. With some very interesting examples, Hirst shows that in natural language context constrains the interpretation of texts. At a local level, it is known that context is a psychological construct that generates the meaning of words. Miller and Charles argued in (Miller and Charles 1991) that word sense disambiguation (WSD) could be inferred easily if we could know all the contexts in which words occur. In a very simplified way, local context can be imagined as a window of words surrounding the targeted concept, but work in WSD shows that modeling context in this way imposes sophisticated ways of deriving the features of these contexts. To us it is clear that the problem of answer justification is by far more complex and difficult than the problem of WSD, and therefore we believe that simplistic representations of context are not beneficial for QA. Hirst notes in (Hirst 2000) that *"having discretion in constructing context does not mean having complete freedom"*.

Arguing that context is a *spurious concept* that may come dangerously to being used in incoherent manners, Hirst notes that *"knowledge used in interpreting natural language is broader while reasoning is shallow"*. In other words, the inference path is usually short while the evidential knowledge is rich. We used this observation when building the abductive processes.

To comply with the simplicity and consilience criteria we need (1) to have sufficient textual evidence for justification - which we accomplish by implementing an information retrieval loop that bring forward all paragraphs containing the keywords of the question as well as at least one concept of the expected answer type[3]; (2) to consider that questions can be unified with answers, and thus provide immediate explanations; and (3) to justify the answer through abduc-

[3]More details of the feedback loops implemented for QA are reported in (Harabagiu et al.2001)

Question: Why are so many youngsters sent to jail?

Answer: Jobs and education diminish.

Context 1: (Reverend Jesse Jackson in an article on Martin Luther King's "I have a dream" speech)

The 'giant sucking sound' is not merely American jobs going to NAFTA and GATT cheap labor zones. The giant sucking sound is that as jobs and education diminish, our youth are being sucked into the jail industrial complex.

Context2: (Ross Perot first presidential campaign)

If the United States approves NAFTA, the giant sucking sound that we hear will be the sound of thousands of jobs and factories disappearing to Mexico.

Figure 5: The role of context in abductive processes.

tions that use axioms generated from WordNet along with information derived by the reference resolution or by world knowledge; (4) to make available the recognition of paraphrases and to generate some normalized dependencies between concepts. However we find that we also need to take into account the contextual information. To account for our belief, we present an example inspired by the discussion of context from (Hirst 2000). The example is illustrated in Figure 5.

When the question from Figure 5 is posed the answer is extracted from a paragraph contained in an article of Reverend Jesse Jackson on Martin Luther King's "I have a dream" speech. This paragraph contains also an anaphoric expression *"the giant sucking sound"* which is used twice. The anaphor is resolved to the metaphoric explanation that this sound is caused by youngsters being jailed due to the diminishing jobs and education. In the process of interpreting the anaphor, the expression *: being sucked into the jail industrial complex"* is recognized as a paraphrase of *"sent to jail"* even if is contains the morphologically relation to the *"sucking sound"*. So what is this *"sucking sound"*? As Hirst first points out in (Hirst 2000)[4], at the time of writing the article containing the answer, Reverend Jesse Jackson alludes to a widely reported remark of Ross Perot. This remark bring forward the context of Perot's speech in his first presidential campaign, when he predicts that NAFTA will enable the disappearing of jobs and factories to Mexico. This example shows that both anaphora resolution, that sometimes require access to embedded context, and the recognition of causation relations are at the crux of answer justification. The recognition of the causation underlying the text snippet *"as jobs and education diminish, our youth are being sucked into the jail industrial complex"* is based on cause-effect patterns triggered by cue phrases such as *as*. This example also leads to the conclusion that abductive processes that were proven successful in the past TREC evaluations need to be extended to deal with more forms of pragmatic knowledge as well as non-literal expressions (e.g. metaphors).

[4]This example was taken from (Hirst 2000)

Redundancy vs. Abduction

Definition questions impose a special kind of QA because often the question contains only one concept - the one that needs to be defined. Moreover, in a text collection, the definition may not be matched by one of the predefined patterns - and thus required additional information to be recognized. Such information may be obtained by using the very large index of a search engine, e.g. *google.com*. When retrieving documents for the concept to be defined, the most frequent words encountered in the summaries produced by GOOGLE for each document could be considered. Due to the redundancy of these words, we may assume that they port important information. However, this is not always true. For the question *"What is an atom?"* we found that none of the redundant word were used in the definition of *atom* in the WordNet database. Nevertheless this technique is useful for concepts that are not defined in WordNet. When trying to answer the question *"Who is Sanda Harabagiu?"*, GOOGLE returns documents showing that she is a faculty member and an author of papers on QA. Therefore for such questions, redundancy may be considered a reliable source of justifications.

Conclusions

In this paper we have presented several forms of abductive processes that are implemented in a QA system. We also draw attention to more complex abductive explanations that require more than lexico-semantic information or axiomatic knowledge that can be scanned in a left-to-right back-chaining of the LFT of candidate answer, as it was reported in (Harabagiu et al.2000). We claim in this paper that the role of abduction in QA will break ground in the interpretation of the intentions of questioners. We discussed the difference between redundancy and abduction in QA and concluded that some forms of information redundancy may be safely considered as abductive processes.

Acknowledgements: This research was partly funded by the ARDA's AQUAINT program.

References

S. Abney, M. Collins, and A. Singhal. Answer extraction. In *Proceedings of the 6th Applied Natural Language Processing Conference (ANLP-2000)*, pages 296–301, Seattle, Washington, 2000.

J. Bear and J. Hobbs. Localizing expressions of ambiguity. In *Proceedings of the 2nd Applied Natural Language Processing Conference (ANLP-1998)*, pages 235–242, 1988.

P. Cohen, R. Schrag, E. Jones, A. Pease, A. Lin, B. Starr, D. Easter, D. Gunning, and M. Burke. The DARPA High Performance Knowledge Bases Project. *Artificial Intelligence Magazine*, 19(4):25–49, 1998.

M. Collins. A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association of Computational Linguistics (ACL-96)*, pages 184–191, 1996.

D. Davidson. The logical form of action sentences. In N. Rescher, ed. *The Logic of Decision and Action*, University of Pittsburg, PA, pages 81–95, 1967.

Christiane Fellbaum (Editor). WordNet - An Electronic Lexical Database. MIT Press, 1998.

A. Graesser and S. Gordon. *Question-Answering and the Organization of World Knowledge*, chapter Question-Answering and the Organization of World Knowledge. Lawrence Erlbaum, 1991.

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

S. Harabagiu and S. Maiorano. Finding answers in large collections of texts: Paragraph indexing + abductive inference. In *AAAI Fall Symposium on Question Answering Systems*, pages 63–71, November 1999.

S. Harabagiu, M. Paşca and S. Maiorano. Experiments with Open-Domain Textual Question Answering. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, August 2000, Saarbruken Germany, pages 292-298.

S. Harabagiu, D. Moldovan, M. Paşca, R. Mihalcea, M. Surdeanu, R. Bunescu, R. Gîrju, V. Rus and P. Morărescu. The role of lexico-semantic feedback in open-domain textual question-answering. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL-2001)*, Toulouse, France, 2001, pages 274–281.

G. Hirst. Context as a spurious concept. In *Proceedings, Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, February 2000, pages 273–287.

E. Hovy, L. Gerber, U. Hermjakob, C-Y. Lin, D. Ravichandran. Toward semantic-based answer pinpointing. In *Proceedings of the Human Language Technology Conference (HLT-2001)*, San Diego, California, 2001.

J. Hobbs, M. Stickel, D. Appelt, and P. Martin. Interpretation as abduction. *Artificial Intelligence*, 63, pages 69–142, 1993.

B. Katz. From sentence processing to information access on the world wide web. In *AAAI Spring Symposium on Natural Language Processing for the World Wide Web*, Stanford, California, 1997.

W. Lehnert. *The Process of Question Answering: a computer simulation of cognition*. Lawrence Erlbaum, 1978.

Mitch Marcus, Beatrice Santorini and Mark Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330, 1993.

G.A. Miller. WordNet: a lexical database. *Communications of the ACM*, 38(11):39–41, 1995.

G.A. Miller and W.G. Charles. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28, 1991.

D. Moldovan and V. Rus. Transformation of WordNet Glosses into Logic Forms. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL-2001)*, Toulouse, France, 2001, pages 394–401.

M. Paşca and S. Harabagiu. High Performance Question/Answering. In *Proceedings of the 24th Annual International ACL SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2001)*, September 2001, New Orleans LA, pages 366-374.

G. Salton, editor. *The SMART Retrieval System*. Prentice-Hall, New Jersey, 1969.

P.R. Thagard. The best explanation: criteria for theory choice. *Journal of Philosophy*, pages 79–92, 1978.

E.Voorhees. Overview of the TREC-9 Question Answering Track. In *Proceedings of the 9th Text REtrieval Conference (TREC-9)*, 2001, pages 71–79.