International Journal of Medical Informatics (2006) **75**, 513—529

ELSEVIER

International Journal of Medical Informatics

www.intl.elsevierhealth.com/journals/ijmi

Semantic representation of consumer questions and physician answers

Laura A. Slaughter[a,*], Dagobert Soergel[b], Thomas C. Rindflesch[c]

[a] *Center for Shared Decision Making and Nursing Research, Rikshospitalet-Radiumhospitalet RH, Forskningsveien 2b, 4. etg., 0027 Oslo, Norway*
[b] *College of Information Studies, University of Maryland, College Park, MD, USA*
[c] *National Library of Medicine, Bethesda, MD, USA*

Received 29 October 2004; received in revised form 29 June 2005; accepted 17 July 2005

KEYWORDS
Semantic processing;
Public health;
Unified Medical Language System;
Information retrieval;
Natural language processing

Summary The aim of this study was to identify the underlying semantics of health consumers' questions and physicians' answers in order to analyze the semantic patterns within these texts. We manually identified semantic relationships within question—answer pairs from Ask-the-Doctor Web sites. Identification of the semantic relationship instances within the texts was based on the relationship classes and structure of the Unified Medical Language System (UMLS) Semantic Network. We calculated the frequency of occurrence of each semantic relationship class, and conceptual graphs were generated, joining concepts together through the semantic relationships identified. We then analyzed whether representations of physician's answers exactly matched the form of the question representations. Lastly, we examined characteristics of the answer conceptual graphs. We identified 97 semantic relationship instances in the questions and 334 instances in the answers. The most frequently identified semantic relationship in both questions and answers was *brings_about* (causal). We found that the semantic relationship propositions identified in answers that most frequently contain a concept also expressed in the question were: *brings_about*, *isa*, *co_occurs_with*, *diagnoses*, and *treats*. Using extracted semantic relationships from real-life questions and answers can produce a valuable analysis of the characteristics of these texts. This can lead to clues for creating semantic-based retrieval techniques that guide users to further information. For example, we determined that both consumers and physicians often express causative relationships and these play a key role in leading to further related concepts.

1. Introduction

Recent research in medical information processing has focused on health care consumers. These users

* Corresponding author. Tel.: +47 23 07 54 63; fax: +47 23 07 50 10.
E-mail address: laura.slaughter@rikshospitalet.no (L.A. Slaughter).

1386-5056/$ — see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2005.07.025

often experience frustration while seeking online information [1—3], due to their lack of understanding of medical concepts and unfamiliarity with effective search strategies. We are exploring the use of semantic relationships as a way of addressing these issues. Semantic information can guide the lay health consumer by suggesting concepts not overtly expressed in an initial query. For example, imagine that a user submits a full question to a search system in the health care domain to find out whether exercise helps prevent osteoporosis. The semantic relationship *prevents* in the proposition representing the question, namely ''exercise *prevents* osteoporosis'', can support this effort; *prevents* might be used with *osteoporosis* to determine additional ways of avoiding this disorder.

We present an analysis of semantic relationships that were manually extracted from questions asked by health consumers as well as answers provided by physicians. Our work concentrates on samples from Ask-the-Doctor Web sites. The Semantic Network from the Unified Medical Language System (UMLS) [4,5] served as a source for semantic relationship types and this inventory was modified as we gained experience with relationship types identified in the health consumer texts.

The objective of this study was to characterize the semantics of consumer health texts in order to develop better search functions within systems that provide health information to consumers. Our first task towards accomplishing this was to identify propositions within these texts and calculate the frequency of occurrence of semantic relationships. For example, we record the proposition ''X <*follows*> bypass surgery'' from the question, ''What *happens after (follows)* bypass surgery?'' Our second task was to identify the ways that questions are connected to answers, since this provides a useful start for constructing query strategies involving semantic information. For instance, we record when the physician answers the question directly with ''loss of appetite *follows* bypass surgery.'' We also look for cases in which the physicians' answer contains an implied relationship such as the *isa* relation between ''bypass surgery'' in the above question and ''operation'' in the answer, ''If you're undergoing a scheduled operation for debilitating angina...''. Our final task was to study the patterns of semantic relationships within answers. We record when the physician expands on the concept ''debilitating angina'', for instance, with an expression such as ''angioplasty *treats* debilitating angina.''

2. Background

2.1. Semantic relationships

A semantic relationship associates two (or more) concepts expressed in text and conveys a meaning connecting those concepts. A large variety of such relationships have been identified in several disciplines, including linguistics, philosophy, computer science, and information science. Some researchers have organized hierarchies of semantic relationships into meaningful but not formal structures [6,7]. Others examine specific relationships in depth, for instance, subsumption [8,9], temporality [10], and meronymy [11]. In addition, ontologies (general purpose and medical) contain semantic relationships that are elements of the overall system. WordNet, for example, contains these primary relationships between concepts: hypernymy (superordination), antonymy, entailment (something inferred), and meronymy (part—whole) [12].

A number of projects have involved the study of semantic relationships specifically within the domain of medicine. Work on the GALEN Common Reference Model examined part—whole relationships [13] and other aspects of ''tangled'' taxonomies [14]. Other ontology projects, such as the foundational model of anatomy (FMA), are central to the delineation of relationships for use in specific types of applications, in this case representation of anatomical structures [15]. Smith and Rosse [16], for example, address how the formal treatment of taxonomy and partonomy in the FMA can support alignment of ontologies.

The UMLS Semantic Network [5] contains 54 semantic relationships, which serve as the basis of our analysis of health consumer texts. This system was chosen as a starting point for our study because it contains a comprehensive hierarchy of relationships linking to a broad range of medical concepts. Each relationship type in the Semantic Network has a definition and is also connected to two semantic types to form a binary proposition (this provides context about what the relationship might mean as well).

2.2. Semantic relationships in information systems

Several researchers have proposed the integration of semantic relations in information retrieval systems as a way of addressing inadequate domain knowledge in users. For example, Chakravarthy and Haase [17] describe an application that explicitly shows the semantic relationships within an expressive query language used for retrieval. The HIBROWSE [18] project illustrated the idea of semantic relationships in ''view-based searching'' using ''knowledge structure hierarchies.'' More recently, Khoo et al. [19,20] explored techniques for using causal relationships for retrieval and proposed a way that causal semantic relationships extracted from Web documents can be ''chained or connected to give a conceptual map of the information available on a particular topic.'' Miles-Board et al. [21] illustrated how ''ontological hypertext'' can be used to provide ''principled and intelligent navigation of knowledge'' through exposing underlying domain knowledge to users.

Several systems within the medical domain have employed the use of semantic relationships in order to aid health-care professionals' search. Mendonca et al. [22,23] used conceptual map representations (terms and their relationships) from patients' records to add contextual information to searches of medical literature for clinicians. Brandt et al. [24] built a retrieval system for an anesthesiology information source that displayed semantic relationships within the retrieved results that linked to the searched keyword. The TAMBIS/STARCH projects [25] have demonstrated the use of ''ontology-driven'' interfaces for query formulation assistance when searching biological information sources. Detwiler et al. [15] created a query engine for the foundational model of anatomy (FMA) that enables users to explore structural relationships. The major goal of the FMA query engine is to support educational computer-based anatomy programs.

Although semantic information used in information retrieval systems can improve performance, there is no direct evidence that explicit semantic representations help lay users understand medical texts and formulate effective questions about health concerns. However, research in text comprehension and cognitive models of health consumers suggests that such representations may provide an effective method for lessening user frustration.

Zeng et al. [1] state that ''consumers employ a different mental model from clinicians.'' When learners with low domain knowledge read texts, the solution to improve coherence is to ''make the relations between items in the text or between general knowledge and the text fully explicit'' [26]. Jonassen et al. [27] suggests that technology-based external representation tools (such as Semantic Networks) can have a positive effect on students' problem-solving performance. In this study we provide an analysis of some of the semantic representations underlying dialogue between lay health consumers and professional information providers. We see our work as underpinning the construction of tools that exploit semantic relationships to help lay users navigate medical knowledge.

3. Methods

3.1. Overview

Our goal was to characterize the semantics of consumers' health questions and physicians' answers found on Ask-the-Doctor Web sites. We had three main research questions: (1) what semantic relationships exist in health consumers' questions and in the answers provided by physicians, (2) how are concepts in the answers connected to those in the questions, and (3) how are the propositions within answers interconnected to form overall semantic structure. Our methods are described below in three sections that correspond to the research questions: identifying propositions, identifying the ways that questions are connected to answers, and identifying patterns of semantic relationships in answers.

3.2. Materials: characteristics of question and answer texts

We analyzed 12 question—answer pairs from seven Ask-the-Doctor Web sites, such as *Ask the Diabetes Team*. All selected sites are designated trustworthy using American Medical Association guidelines [28]. In order to avoid institutional bias and to maximize the types of semantic relationships identified, we selected sites that cover a variety of medical topics. An example of question and answer is given in Fig. 1. The average length of each question—answer pair in our sample was 275 words. The average question message sent to a physician contains four sentences, three assertions providing background and one actual question sentence. Each individual question sentence was labelled as a ''subquestion.'' For example, in Fig. 1, the question message contains only one sub-question. The maximum number of sub-questions asked in a message was three. There were a total of 20 sub-question sentences in the 12 question messages sampled. The average number of words per answer was 199 and the average number of sentences per answer was 10.

3.3. Identifying propositions: the coding process

We completed several iterations of manual coding before arriving at the final set of propositions from

Question: [Background statement 1] I was recently diagnosed with GERD and began taking Prilosec after dietary changes failed to lessen painful heartburn symptoms that accompanied every meal. **[Background statement 2]** While I was on Prilosec, the heartburn disappeared. **[Background statement 3]** However, I often had the uncomfortable sensation that my stomach was bloated and yet very empty at the same time. **[Background statement 4]** Because of insurance coverage changes, I switched to Prevacid, and the discomfort has disappeared. **[Background statement 5]** My doctor says that Prilosec and Prevacid are interchangeable. **[Sub-question]** Do you know of any differences between the two drugs?

Answer: People with gastroesophageal reflux disease (GERD) often see a doctor after trying over-the-counter medications such as antacids. The doctor may recommend other more effective anti-acid medications, such as ranitidine (Zantac). However, a small percentage of patients, usually ones with more severe reflux changes seen on endoscopy in the esophagus, will not respond fully to these medications. These patients usually receive a more potent class of drugs called proton-pump inhibitors (PPIs).

PPIs include omeprazole (Prilosec) and lansoprazole (Prevacid), as well as the newly released rabeprazole (Aciphex). Laboratory studies done on PPIs have found only minor differences in their activity, and clinical studies done with patients show no differences among the drugs with respect to symptom relief or healing of esophagitis.

I am not surprised that your symptoms of GERD disappeared with both Prilosec and Prevacid. However, while the scientific data indicate all PPIs have similar side effects. I have often been told by various patients that one drug or the other seems to suit them better. The bloating you felt while on Prilosec could have been a medication side effect. It could also have a coincidental symptom of GERD, unrelated to the drug.

If you are doing well with Prevacid, there is no reason to change your medication. If bloating or other side effects happen to return, check with your doctor to make sure no other problem exists.

http://www.ivillagehealth.com/print/0,,172136,00.html

Fig. 1 An example of a question—answer pair from an Ask-the-Doctor Web site. The question message contains one sub-question and five background statements.

the question—answer texts. The first author was the main coder throughout the process of iterative coding. The second and third authors were often consulted while coding instances of semantic relationships.

As the first step in the process, we created an initial conceptual graph [29] to help sort out areas of inconsistent coding and uncertainty in class assignment. Using the graphs as a guide, the relationships between the concepts were reviewed. At this stage, the classes of relationships found in the UMLS Semantic Network were revised in order to capture the perceived meaning underlying the textual data. When we observed instances that did not quite fit, we consulted other sources and adapted the inventory to the relationship classes identified in the health consumer texts. The final list of semantic relationship classes is shown in Table 1.

The propositions were then represented as frame structures. Each class of relationship became a frame with multiple possible slots (now n-ary not binary relationships). Slot names for the frames were constructed, reviewed, and revised during coding iterations. In the final iteration, the propositions were checked for errors. We also regenerated the conceptual graphs, based on the frame instances, for further analysis of the characteristics of the inter-connections between the many concepts in each text.

The first author was the primary coder of semantic relationships within the question and answer texts for this project. To assess coding consistency, the first author recoded three of the questions and answers several months after the last iteration was completed without re-examining the previously identified semantic relationships. Intra-coder reliability was calculated using percent agreement.

Two additional coders, a physician and a medical librarian, coded one question—answer pair. The main purpose of this was to assess differences between coding styles. Agreement is difficult to achieve between coders because at least three (or more) pieces must match: the relationship identified must be the same, and the concepts selected as arguments must also be the same. We obtained mean percent agreement for the inter-rater coding. The main use of this data was, however, to refine our coding methods.

3.3.1. Guidelines for coding propositions

Selecting concepts from the texts requires some amount of subjective speculation on the part of

Table 1 Frequencies of semantic relationship types found in the questions (Qs) and answers (As)

Relationship	Frequency (%)		Relationship	Frequency (%)	
	Q	A		Q	A
0 associated_with	—	—	3 temporally_related_to	—	—
1 topologically_related_to	—	—	3.1 ***co-occurs_with***	***6.2***	***2.1***
1.1 part_of	—	—	3.2 ***precedes***	***6.2***	***3.0***
1.1.1 consists_of		1.2	3.3 age_of	2.1	—
1.1.2 contained_in	—	3.9	3.4 cyclic_frequency_of	—	0.6
1.1.3 ingredient_of	—	1.8	3.5 delays (also 2.1.2)	1.0	0.6
1.1.4 component_of	—	0.3	3.6 ***duration_of***	***6.2***	***0.9***
1.2 connected_to	—	—	3.7 time_position_of	2.1	1.5
1.2.1 branch_of	—	—	4 conceptually_related_to	—	—
1.2.2 interconnected_with	—	—	4.1 analyzes	—	0.9
1.2.3 tributary_of	—	—	4.1.1 assesses_effect_of	—	2.4
1.3 location_of	4.1	3.3	4.1.2 ***diagnoses***	***4.1***	***6.0***
1.3.1 adjacent_to	1.0	0.3	4.1.3 measures	—	0.9
1.3.2 surrounds	—	0.9	4.1.4 evaluation_of	—	—
1.3.3 traverses	—	0.3	4.1.5 degree_of	1.0	—
2 functionally_related_to	—	—	4.1.6 measurement_of	—	—
2.1 affects	—	0.6	4.1.7 ***compared_to***	***2.1***	***5.4***
2.1.1 absorbs	—	0.6	4.2 property_of	2.1	4.2
2.1.2 delays (also 3.5)	1.0	0.6	4.3 requires	1.0	1.8
2.1.3 complicates	—	0.3	4.4 derivative_of	—	—
2.1.4 disrupts	—	0.3	4.5 developmental_form_of	—	—
2.1.5 facilitates	—	0.3	4.6 method_of	—	—
2.1.6 increases	—	1.2	4.7 issue_in	—	0.6
2.1.7 decreases	—	1.5	5 ***isa***	***1***	***6.3***
2.1.8 interacts_with	1.0	3.6			
2.1.9 manages	—	—			
2.1.10 ***prevents***	***5.2***	***2.7***			
2.1.11 ***treats***	***11.3***	***9.0***			
2.2 ***brings_about***	***27.8***	***15.6***			
2.3 performs	1.0	0.3			
2.3.1 carries_out	—	—			
2.3.2 exhibits	—	0.3			
2.3.3 practices	—	—			
2.4 occurs_in	4.1	1.2			
2.5 process_of	1.0	0.6			
2.6 uses	—	—			
Additional relations					
definition_of	—	***5.1***			
has_family_relationship	3.1	0.9			
same_concept_syn_term	—	2.4			
relation_x (unknown relationship)	4.1	—			

For questions (Q) $N = 97$ and for answers (A) $N = 334$. The relationships occurring with a frequency greater than 5% are shown in bold italic type.

the coder. Within the data, the string chosen to participate in a relationship is usually the same as what has been specified in the text (or normalized string). In general, complex concepts were avoided, especially when they consisted of full phrases that included verbs and prepositions. Noun compounds, such as ''saliva protein,'' were not broken down if they formed one distinct concept.

We recorded an anaphor by replacing it with its referent and writing the anaphor in brackets after it. For example, the concept ''flu'' in the second sentence, ''I thought that I had the flu. Later, I found out it was bronchitis'' appears in the data as ''flu [it].''

The UMLS Semantic Network relationship classes are hierarchical and we coded for the most specific relationship in the hierarchy whenever possible. Therefore, many of the general categories of relationship types, such as *temporally_related_to* were never coded within the texts since we always made an attempt to identify the most specific relationship possible.

Inferred relationships recover from the text something that was not overtly asserted. Some of the identified relationships were inferred within the question texts or within the answer texts. For example, ''He broke his arm and went to the hospital'' does not explicitly state a relationship that going to the emergency room was the result of breaking an arm.

3.3.2. Illustration of coding process of semantic relationships in health consumer texts

In this section, we illustrate the process of coding using one sentence taken from a question statement:

''I have had migraines frequently for the last twenty years and during the last ten years I have had two TIA's (Transient Ischemic Attack).''

The concepts identified in this sentence are ''migraines'' and ''TIA's (transient ischemic attack).'' They are discussed with temporal concepts: ''frequently,'' ''last twenty years,'' and ''last ten years.'' ''Migraines occurring in the time period of last 20 years'' and ''TIAs occurring in the last 10 years'' might be represented using the UMLS relation *occurs_in*.

The *occurs_in* (inverse: *has_occurrence*) relation of the UMLS Semantic Network incorporates aspects of time within the definition. ''Takes place in or happens under given conditions, circumstances, or time periods, or in a given location or population. This includes appears in, transpires, comes about, is present in, and exists in.'' However, the main difficulty is that the *occurs_in* relation appears subsumed under *functionally_related_to* while other time relationships are placed under *temporally_related_to*.

Although we did not code the semantic types of the concepts identified, we did look at the semantic types paired with each UMLS semantic relationship as a method of improving our understanding of the definitions provided by the Semantic Network. A brief review of the semantic types in the Semantic Network that can be related through *occurs_in* illustrates that the relation is associated with populations (semantic type 'group'), other entities and processes (e.g. semantic types 'organism function', 'disease or syndrome', and 'injury or poisoning') in addition to time (semantic type 'temporal concept'). However, as seen in this example, the semantic type of migraines and TIAs, 'disease or syndrome', does not connect to semantic type 'temporal concept' using the *occurs_in* relationship. We then reviewed some of the other coded instances from our question—answer texts to find that *occurs_in* was mainly used to represent ''occurring in a population'' subsumed under the relation *functionally_related_to*.

To untangle the multiple uses of *occurs_in*, temporal relations were required to be subsumed under *temporally_related_to* and we therefore modified the definition of *occurs_in* to: ''Takes place in or happens in a given population. This includes appears in, comes about in, is present in, and exists in a population.'' We created a new relationship, subsumed under *temporally_related_to*, called *duration_of* (inverse: *has_duration*) with the definition ''Related to the length of time an activity continues.''

We then identified two instances of the relationship *duration_of* that served the purpose of relating the concepts ''migraine'' and ''TIA (transient ischemic attack)'' to the time periods in which they were experienced. The frame slots PhenomenonInTime, FrequencyWithinDuration, and OverallDuration were added to hold the values identified in the text:

''have had migraines frequently for the last twenty years''

duration_of-1
PhenomenonInTime [migraines]
FrequencyWithinDuration [frequently]
OverallDuration [last twenty years]

''during the last ten years I have had two TIA's (Transient Ischemic Attack)''

duration_of-2
PhenomenonInTime [TIA's (Transient Ischemic Attack)]
FrequencyWithinDuration [two times]
OverallDuration [last ten years]

3.3.3. Calculating the frequency of semantic relationships

We computed frequencies for each semantic relationship type in order to assess how often each semantic relationship type appears in consumer health questions and answer texts. We computed these separately for propositions coded in the questions and propositions coded within the answer messages.

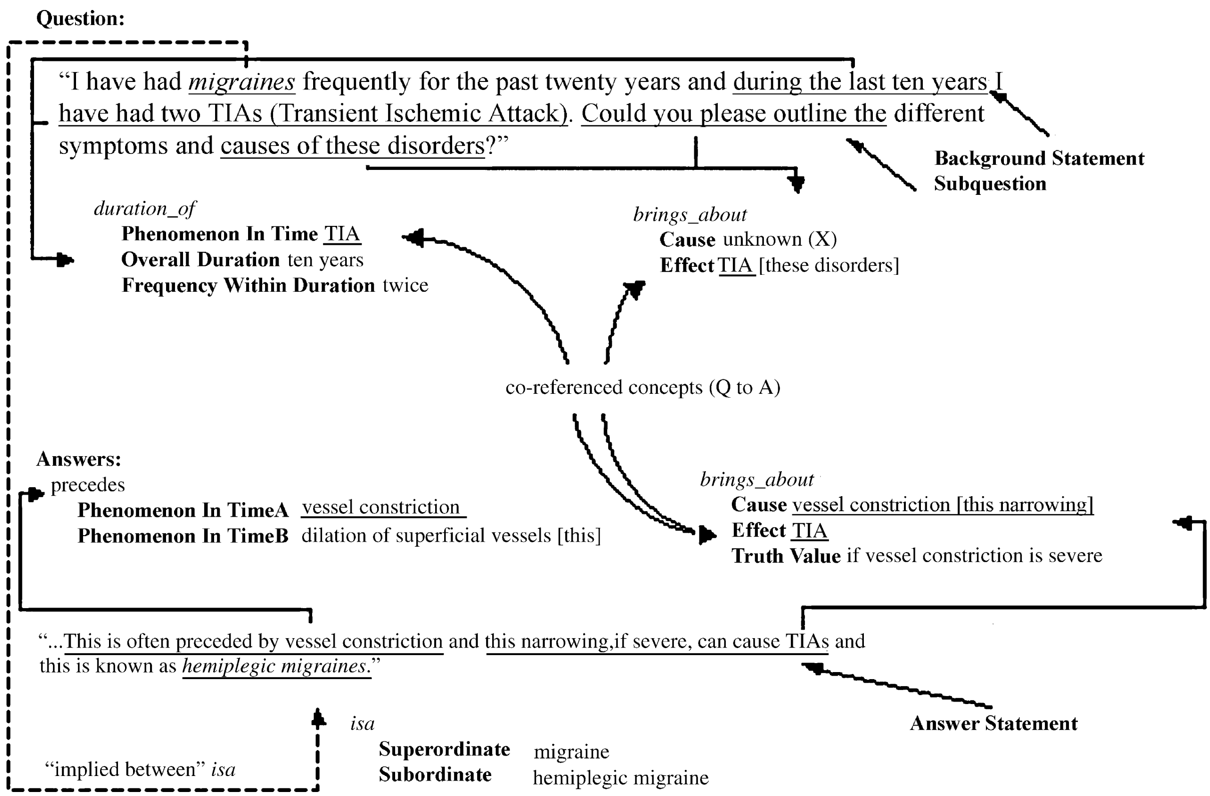

Fig. 2 An example of coded semantic relationships in question and answer texts.

3.4. Identifying the ways that questions and answers are connected

The 12 question—answer pairs we analyzed contained 20 sub-questions. Each of these sub-questions has at least one semantic relationship instance that represents the question itself, that is, what information is lacking but desired by the questioner. In the semantic relationship frames, the unknown could be one or more of the following three: an unknown concept in a slot, an unknown relationship between two or more concepts, or a request for verification (V) of the truth of the statement.

We assessed how the sub-questions are connected to answers through their semantic relationships. Fig. 2 illustrates how the frames were connected for the analysis. In this example, we chose two frame instances, *duration_of* and *brings_about* from a question statement and two frame instances, *precedes* and *brings_about* from its answer. The slot value of PhenomenonInTime ''transcient ischemic attack'' in *duration_of* contains the same concept as the *Effect* slot ''transcient ischemic attack'' of *brings_about* from the answer. We refer to these as co-referenced concepts. This forms a link between the question and the answer. In addition to these question-to-answer connections, an implied link from question to answer was identified using subjective assessment of the unstated relationships between question concepts and answer concepts. ''Hemiplegic migraine *isa* migraine'' is the implied relationship that was not explicitly stated in the text itself. Co-referenced concepts within the answers, for example ''vessel constriction'' are also of concern for our analysis; these link semantic relationship instances to form large graph structures representing answer texts.

3.4.1. Analysis of the question to answer link

We first computed the frequency of semantic relationship types containing co-referenced concepts. This means that we calculated the frequency that each semantic relationship type occurred for the subset of question messages' propositions that contain concepts physicians repeated in the answers. We also did the same for the subset of answer propositions containing concept arguments from the questions. From this analysis, we were able to tell what semantic relationship types are most often used in leading the questioner to concepts that might be new to them. For example, when a question states ''Wolfram's syndrome is *diagnosed_by* genetic tests ...'' and the answer discusses, ''these genetic tests (the co-referenced concept) *analyzes* nuclear and microsomal disorders,'' the questioner did not express ''nuclear disorders and microsomal disorders,'' so these might be unknown concepts to that person.

We then calculated the frequency that the implied relationships occurred between the questions and the answers. These were the relationships between a question concept and an answer concept that are not explicitly written. They require external knowledge to understand. For example, ''hemiplegic migraine *isa* migraine'' was not explicitly stated by the physician when he wrote, ''The headache that accompanies migraine is due to dilation of superficial vessels. This is often preceded by vessel constriction and this narrowing, if severe, can cause TIAs and this is known as hemiplegic migraine.'' The questioner had only asked about migraines in the question ''I have had migraines frequently for the past twenty years...'' and therefore the reader must deduce that the physician is discussing a more specific type of migraine, called hemiplegic migraine.

3.5. Patterns of semantic relationships within answers

We calculated the frequency of what we term ''direct answers'' within the set of semantic relationships in the answers. A direct answer is provided by physicians when they have repeated the semantic relationship of the questioner in a completed form; they fill-in the missing concept, relationship, or answer ''yes, true'' or ''no, false'' to verifications requested. The only other method for receiving an answer to questions is through inferences to further concepts using the semantic relationship connections in the answers.

Our next step was to examine the graphs of these intra-answer connections. The main reason for creating graphs of the propositions was to examine the discourse structure of physicians' answers. Fig. 3 illustrates how several sentences from an answer form a sub-graph (extracted from a larger complete answer graph). Due to the small sample size in this study, it is not possible to do calculations on the characteristics and patterns within the graphs. However, we were able to make several observations concerning the 20 answers by looking at graph size and the patterns that formed the discourse structure.

4. Results

The results are organized according to the three main research questions. In the first section, we discuss the types of semantic relationships expressed in the question—answer texts and their frequency of occurrence. We then give the results concerning how answer concepts are connected to the questions asked by health consumers. The last section covers the discourse structure of the propositions within answers. These results, taken together, characterize the semantic structure of consumer health questions and physicians' answers.

4.1. Results of coding semantic relationships

4.1.1. New relationships added to the UMLS SN

The changes to the Semantic Network included revisions to the definitions and the hierarchical structure as well as the addition of new semantic relationships. The final set of semantic relationships is listed in Table 1. This section briefly reviews the modifications made to the SN inventory of relationships. Each change was essential to making the inventory functional; however, future work is required to reach the ideal. A more detailed review of the changes made to the SN during coding can be found in [30].

The SN relationships subsumed under *physically_related_to* and *spatially_related_to* were moved under a new heading, *topologically_related_to* with three main subordinate relationships: *part_of*, *connected_to*, and *location_of*. The relationships subsumed under the Semantic Network *affects* relationship were not modified substantially; we added *facilitates*, *increases*, *decreases*, and *absorbs* as the children of *affects*. There are four causal relationships in the UMLS Semantic Network: *brings_about*, *causes*, *produces*, and *results_in*. Because these relationships are similar, *causes*, *produces*, and *results_in* were removed, and all the causal relationships were coded as *brings_about*. The UMLS definition for *occurs_in* was modified to remove references to time and location. Additional time relationships were added under *temporally_related_to*: *age_of*, *delays*, *duration_of*, *time_position_of*, and *cyclic_frequency_of*. A significant number of relationships under the parent *conceptually_related_to* in the Semantic Network addressed evaluation and measurement. In coding the texts, it became apparent that comparison relationships were necessary. *Compared_to* was introduced and placed under *analyzes*.

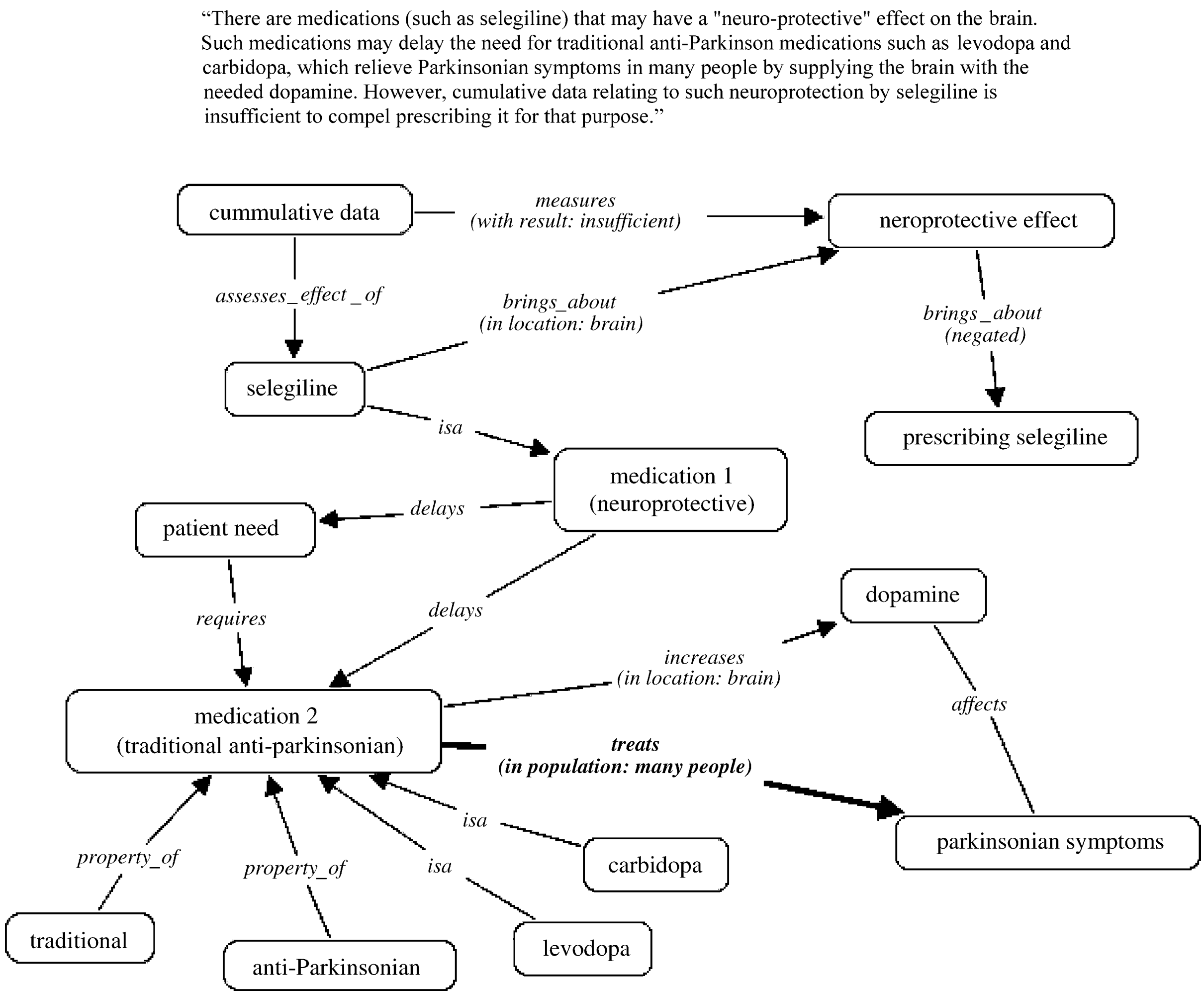

Fig. 3 An example of a segment of answer text represented as a conceptual graph. The bolded relationship, *treats*, was a direct answer to the question statement ''what *treats* Parkinsonian symptoms?''.

4.1.2. Coding agreement

In our assessment of stability of coding, intra-rater reliability percent agreement was 88%. For the inter-rater coding, Coder 1 identified a total of eight relationships in the question text and 38 in the answer texts. Coder 2 identified a total of 12 relationships in the question text and 119 in the answer texts. Of these, 24 instances matched between the two coders with both semantic relationship as well as the arguments identified being exactly the same. The mean percentage of exact agreement between the two coders was 35%. Differences between the relationships identified, such as the use of *frequency_of* and *property_of* led to re-examination of those relationships (including definitions and often structural arrangement) within the inventory we used for coding.

We show the results of the coding performed by our two experts in Table 2. The table illustrates the differences between coders' coding styles and selection of semantic relationships. From the text ''Most parents are quite familiar with middle ear infections, which occur behind the eardrum'', Coder 2 and the authors both identified middle ear infections having the location ''behind the eardrum.'' The difficult part of the sentence was to be able to represent ''Most parents are quite familiar with middle ear infections.'' Coder 1 chose to make ''familiarity'' or ''commonality'' a property of ''middle ear infection'' while Coder 2 inserted a new relationship into the coding scheme, *frequency_of*. The authors felt that the relationship expressed, *familiarity*, is a cognitive state relationship (such as *to know* and *to think*) not represented in the UMLS.

4.1.3. Semantic relationship frequencies

In total, we identified 97 relationship instances in the questions and 334 instances in the answers. The frequency of relationship occurrence is shown

Table 2 Semantic relationship coding: inter-rater comparison

Semantic relationships		
Coder 1	Coder 2	Authors
	location_of **Object** middle ear infections **LocationSite** behind eardrum	*location_of* **Object** middle ear infections **LocationSite** behind eardrum
property_of **Phenonemon** middle ear infection **Property_value** common	*frequency_of* **Event** middle ear infection **Frequency** common	
	has_family_relationship **FamilyMemberA** child **FamilyMemberB** parent	

There are two main types of ear infection: otitis media (infection of the middle ear) and otitis externa (infection of the outer ear). *Most parents are quite familiar with middle ear infections, which occur behind the eardrum*. The patient, usually a young child, is feverish, irritable and has ear pain. Children too young to speak often simply tug at their ears to indicate their discomfort. Diagnosis, which involves viewing the ear canal and eardrum with an instrument called an otoscope, is quite simple. Treatment frequently involves a course of oral antibiotics. *Note*: The semantic relationship instances were identified within the italicized sentence.

in Table 1. The relationships most often occurring in the questions were: *prevents*, *treats*, *brings_about*, *co_occurs_with*, *precedes*, and *duration_of*. The most frequently identified relationships in the answers were: *treats*, *brings_about*, *diagnoses*, *compared_to*, and *isa*. An additional relationship, *definition_of*, was created in order to track segments of text that provide definitions of health concepts, and these occurred with high frequency in the answers. We used *relationship_x* when the relationship asserted between two (or more) concepts was the focus of the question, as in ''Do you know of any differences between Prilosec and Prevacid?'' Four percent of all the semantic relationship instances identified in questions were *relationship_x*.

4.2. Question-to-answer co-referenced patterns

A total of 497 co-references occurred between the 97 question propositions and 334 answer propositions identified. Question relationships that most frequently contained co-referenced arguments with answer relationship instances were: *brings_about* (127/497, 25.6%), *duration_of* (45/497, 9.1%), *co_occurs_with* (44/497, 8.9%), *relationship_x* (42/497, 8.5%), and *treats* (41/497, 8.2%). Answer relationships that most frequently contained arguments that co-referenced concepts in the question relationship instances were: *brings_about* (121/497, 24.3%), *isa* (62/497, 12.5%), *co_occurs_with* (48/497, 9.7%), *diagnoses* (42/497, 8.5%), and *treats* (36/497, 7.2%).

In 70% (350/497) of the co-references from above, the answer relationship linked a question concept with a concept not stated in the sub-question. We calculated that 21% (74/350) of the 70% were not stated in any part of the question text and are potentially new concepts to the questioner. Some example concepts introduced by the physicians are listed in Table 3 (in the last column).

The idea of an ''implied'' relationship is that concepts in questions and answers are linked together through external knowledge, meaning that a relationship was never stated in the text. The search for ''implied'' relationships was very subjective; however, it provided an account of some of the implicit relationships existing between question and answer concepts. We identified 78 ''implied'' semantic relationships between the questions and answers; these are shown in Table 4. The most prevalent ''implied'' relationship identified was *isa*. Most often (70%), a broader concept is stated in the question and a more specific concept is presented in the answer. In the other 30%, the opposite occurred.

4.3. Related concepts in answers

4.3.1. How answers relate to questions—direct answers

At least one direct answer was provided to 6 of the 20 sub-questions (6 out of the 12 question—answer messages had a direct answer). For the 14 sub-questions that do not have at least one direct answer, the questioner receives an answer either

Table 3 Examples of linked question-to-answer propositions leading to concepts that are potentially new to the questioner

Example	Concept in question proposition	Question relationship	Co-referenced concept	Answer relationship	Related concept in answer proposition (further search term)
1	severe spinal stenosis→	precedes	dizziness→	co_occurs_with	disequilibrium
2	Wolfram's syndrome→	diagnoses (inverse)	genetic tests→	analyzes	microsomal disorders
3	mosquito bites→	brings_about	allergic reaction→	brings_about	lightheadedness
4	buzzing sound→	co_occurs_with	dizziness→	brings_about (inverse)	acoustic neuroma

Table 4 Frequency of implied relationships

Relationship type	%
1 topologically_related_to	
1.1 ***part_of***	***6.4***
1.3 location_of	3.9
2 functionally_related_to	
2.1 affects	4.3
2.1.11 ***treats***	***13.8***
2.2 ***brings_about***	***16.7***
3 temporally_related_to	
3.1 co-occurs_with	3.9
4 conceptually_related_to	
4.1 analyses	5.2
4.2 property_of	1.0
4.3 requires	3.1
5 ***isa***	***37.4***
Additional relationships	
synonym_specified	4.2

The relationships occurring with a frequency greater than 5% are shown in bold italic type. $N = 78$.

through links to co-referenced answer frames or through implied relationships (or through both). Of all the answer relationship instances summarized in Table 2, those providing direct answers make up only 3% (10/334) of all the relationship instances in the answers. Not many answers contain a semantic representation of the exact format specified by the questioner. We found that only 16% (53/334) contain a co-referenced concept with a sub-question and 7% (23/334) contain a co-referenced concept with a background sentence. The 74% (247/334) remaining answer relationship instances are not directly connected to a question concept but may be linked with other answer relationships.

4.3.2. Discourse patterns

In the previous section, we found that the answers we analyzed contained many semantic relationship propositions that are not connected at all to a question concept (74% of them). We next analyzed the structure of answers by connecting the propositions as graphs. We identified semantic relationships that involved anaphor participants, and so it was possible to connect statements from sentence to sentence. Hence, we assumed and observed that most of the concepts discussed in these paragraphs were interconnected through the semantic relationships. In all answers, there is one main graph along with one or more smaller graphs. Overall, we see one large connected graph because the answers physicians wrote consisted of well-formed and conceptually related paragraphs. These smaller graphs exist because there is no semantic relationship to connect them with any of the concepts within the main (larger) answer graph. However, these smaller graphs should be connected to the main graph because, in no case, did a physician talk about two disparate topics. On further inspection we found that there were two possible reasons why they were not connected; either a semantic relationship was missed in the analysis or making these connections required external knowledge not explicitly stated in the answer. In Table 5, a summary of the number of graphs constructed from each answer is shown. We provide both the number of concept nodes and the semantic relationships in order to give an idea of the size of the graphs within each answer. Relationships that we did not include as part of the semantic relationship inventory, *definition_of*, *has_family_relationship*, *same_concept_synonymous_term*, and *relationship_x* were not included in these graphs.

There are some concepts in answers that participate in numerous semantic relationship frames. In the 12 answer graphs, we observed that one or two concepts have a very large number of semantic relationships connected with them, in comparison with the other concepts. For example, in Fig. 3 above, the concept ''traditional anti-Parkinsonian medication,'' is found in eight propositions. Concepts with relatively high number of semantic relationships connected to them, compared with the other concepts in the graph, can be thought of as the focal concepts within the answer. We counted 18 of these focal concepts in the 12 answer graphs, 7 were concepts also expressed in the question statement, 10 participated in a semantic relationship with a question concept (were neighbors of a question concept) and 1 was a distance of two semantic relationships from a question concept.

In the above cases, we found that sub-patterns in the shape of a star formed around the focal concepts. So, for example, in Fig. 3 a star forms around the node ''traditional anti-Parkinsonian medication.'' This focal concept is connected to many concepts and these connected concepts are not connected to anything else. Larger star graphs than the example above are found in the other 12 answers, an example is shown in Fig. 4. Star patterns are often found in explanations that involve short descriptions of concepts' properties.

Most graphs, but not all, contained cycles. Cycles are closed paths that form a loop. For example, as in the closed loop cycle ''levodopa *increases* dopamine,'' ''dopamine *affects* Parkinson symptoms,'' ''anti-Parkinsonian medication *treats* Parkinson symptoms,'' and ''levodopa *isa* anti-Parkinsonian medication.'' In our 12 examples, many cycles consisting of semantic relationships between 3 or 4 concept nodes were identified within the larger graph structures making it possible to see explanations in the discourse structure.

Table 5 Size and number of conceptual graphs generated from answer frame instances

Answer number	Graph	$\|C\|$	$\|\|SR\|\|$
1	a	16	20
	b	9	8
	c	3	2
	d	2	1
	e	2	1
2	a	19	20
	b	2	1
3	a	14	16
	b	8	7
	c	2	1
	d	2	1
4	a	19	25
	b	2	1
5	a	10	14
	b	5	4
	c	4	3
6	a	14	16
7	a	13	12
	b	4	4
	c	3	2
	d	3	2
8	a	24	28
	b	3	2
	c	2	1
9	a	11	10
	b	2	1
	c	2	1
10	a	16	15
11	a	14	16
	b	6	5
	c	3	2
	d	2	1
	e	2	1
12	a	33	35
	b	3	2
Total		279	281

$|C|$ denotes the number of concept nodes in the graph and $||SR||$ denotes the number of semantic relationships.
Note: Only 281 of the 334 semantic relationships were represented in the conceptual graphs because *definition_of*, *has_fam_relation*, *relationship_x*, and *synonym* relationships were not included. Each graph constructed from the answer relationships was assigned a lower case letter according to graph size (the ''main'' largest graph is labeled ''a'').

"A biopsy is the only way to tell for sure what any lesion is. That will give a definitive answer and direct the need for any further therapy. If it's a benign nevus, then no further therapy is required. If it is a melanoma or other form of cancer, then the entire lesion and a margin of surrounding normal-appearing tissue will have to be removed. Examination of the skin, including the skin of the genitals, should be a monthly habit for all people, female and male. For women, I suggest doing a thorough check of your skin every month at the same time you do your breast-self exam. Enlist a large mirror, a friend or your partner to look at your back. You should consult a doctor about any lesions that seem to be new, are getting larger or darker, are bigger than a pencil eraser, are itchy or scaly, bleed or don't heal."

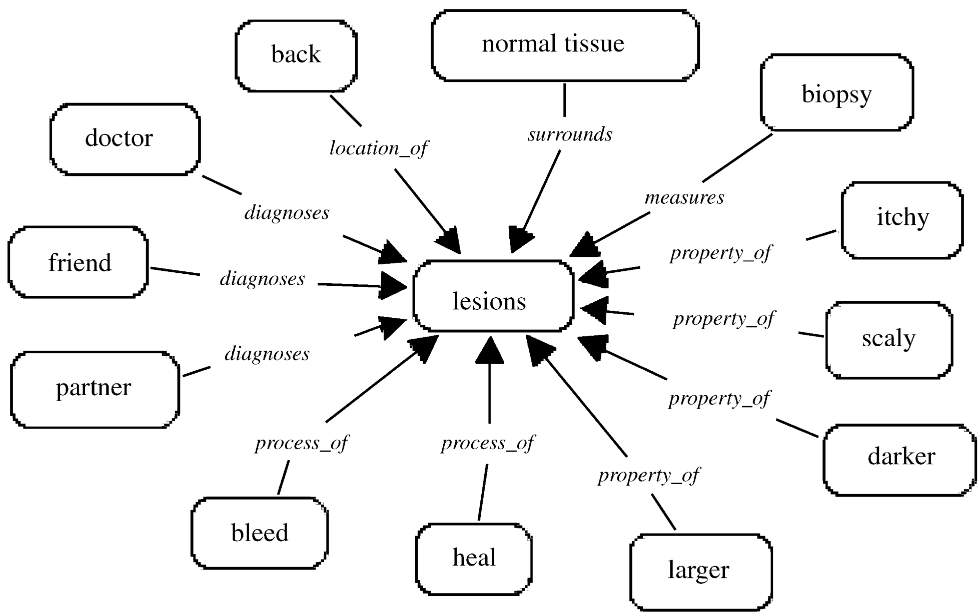

Fig. 4 An example of a star-shaped subgraph within an answer shows that there are many concepts discussed that relate to the concept *lesions*.

5. Discussion

Our first task in this study was to identify propositions in question—answer texts and calculate the frequency of occurrence of semantic relationships. A number of useful observations are based on the frequency of semantic relationship occurrence within the texts. For one, we show which semantic relationships from the UMLS Semantic Network are found within these texts. Also, we learned what other relationships may be needed if these texts are to be represented in a retrieval system that makes use of semantic metadata.

Semantic relationships expressing comparison, combined with frequently expressed *treats*, are essential to providing explanations of different treatment options and a frequent purpose of a consumer question is to compare treatments [31]. We found that these physicians often make comparisons in answers and this led to a modification in the UMLS Semantic Network inventory. The *compared_to*, *equivalent*, *similar_to*, and *different_from* (see Table 1, # 4.1.7) relationships were not part of the UMLS Semantic Network but we, along with other researchers, have identified a need for the addition [32]. Another observation of the study is that some semantic relationships occur predominantly in questions and others occur predominantly in answers. Health consumers, for example, often use temporal semantic relationships in explaining their information need. The work also provides evidence that identification of causal relationships is fundamental to question answering. The *brings_about* (causal) relationship is the most frequently expressed relationship in both questions and in the answers.

The results from looking at the question-to-answer links provide a picture of how answers are related to the concepts asked in the questions. Co-referenced concepts often link *brings_about*, *duration_of*, *co_occurs_with*, and *treats* from questions to *brings_about*, *isa*, *co_occurs_with*, *diagnoses*, and *treats* in the answers. These are the most common ''entry points'' into answer documents that lead to additional information that the questioner might not be aware of. Also, in the implied relationship analysis results, we are not surprised to see a large number of *isa* relationships that are implied links from question concept to answer concept. This shows that physicians assume that they can introduce hierarchically related concepts without further explanation and that laypersons will be able to understand, for instance, that ''leukemia *isa* cancer,'' through their own external knowledge. Lastly, in our analysis of whether physicians provide what we call direct answers, meaning that the physician repeats the question proposition exactly as stated, we have shown this does not always occur. Instead, we see that physicians expand on the question concepts in their answers to provide explanations, and these results have been briefly described in an earlier paper [31].

From the graph analysis of within-answer semantic relationships, we learned about semantic relationship patterns within answers. Graphs give a visual representation of the discourse structure of answers. If, for example, a graph contains few concepts and a large number of semantic relationships, we know that the physician has made many connections between very few ideas. However, our graphs contain an almost equal number of concepts and semantic relationships, meaning that in most cases, the physician writes answers containing a variety of concepts and connects them with an almost equal number of semantic relationships. Of interest are the star sub-patterns that form around the focal concepts of answers and closed-loop cycles. One method of arriving at information that might prove helpful for lay users would be to identify potential focal points and then locate many different semantic relationships with that ''focal concept'' participant (e.g. what are its properties using *property_of*, what *treats* it, what *prevents* it, who *diagnoses* this). Searching for cycles might be a way to locate segments of text providing ''closed'' explanations used for reasoning and helping users understand inter-related concepts.

5.1. Implications

One reason for completing the research reported in this paper is to begin to explore how semantic-based retrieval techniques and external representations of semantic relationships can be used within health consumer information and decision systems. We have made two important assumptions with this work. While conducting this study we assumed that the questions posed in these ''Ask-the-Doctor'' texts are representative of lay information needs and that our analysis of the semantic relationships can be applied to future information retrieval/educational systems for health consumers. A second assumption is that physicians believe these answers are useful and appropriate for laypersons. This work can form the basis for an interpretive layer to mediate between lay (illness model) and professional (disease model) as proposed by Soergel et al. [33]. Building techniques that help systems bridge consumer-level language to professional-level language will require semantic interpretation components.

A finding that is important for semantic-based information retrieval research concerns the *brings_about* (causal) relationship; it is the most frequently expressed relationship in both questions and in the answers. Our results might help clarify why Khoo et al. [20] found that partial relation matching in which one member (i.e. term) of the relation is a wildcard is especially helpful for retrieval using causal relationships. They determined that the most successful type of causal relation matching occurred with the use of a wildcard (''*''). So, for example, ''smoking *cause* *'' or ''* *cause* cancer'' was more effective than the complete concept-relation-concept triple ''smoking *cause* cancer.'' In our work, we saw that *brings_about* relationships found in answers often contain concepts that are new to the questioner, and hence could only be recovered with a wildcard search. *Brings_about* is the semantic relationship with the highest number of co-referenced arguments linking the questions and answers. We further observed that long causal chains exist in the answer texts and these might be part of a retrieval strategy that provides results for assisting users in understanding multiple causative factors leading to an illness or certain symptoms.

Causal relationships play a key role in these texts, but extraction of other relationships can also be expected to improve retrieval performance. Our results provide a picture of the combinations of semantic relationships that might be useful in partial matching techniques for lay health information. We found that co-referenced concepts often link *brings_about*, *duration_of*, *co_occurs_with*, and *treats* from questions to *brings_about*, *isa*, *co_occurs_with*, *diagnoses*, and *treats* in the answers. For retrieval applications, if a questioner asks ''smoking *brings_about* lung cancer'', then we expect that it will be useful to search paths ''lung cancer *isa* (*)'', ''(*) *treats* lung cancer'',

''(*) *co_occurs_with* lung cancer'', and/or ''(*) *diagnoses* lung cancer'' in addition to ''lung cancer *brings_about* (*).''

5.2. Limitations

The major limitation of this study involves issues related to coding propositions within texts. The semantic relationship instances coded from the text went through numerous iterations before arriving at a set that, although not perfect, reflects a usable representation. The first author manually identified semantic relationships using a set of coding rules that are documented in [30] and was consistent at applying the instructions. We feel that 88% intra-rater reliability is acceptable and did not warrant another iteration of coding.

Identification of semantic relationships, like all indexing, is subjective and represents the reaction of human beings to the information they are processing. The results indicated differences in the granularity of coding for the two coders who were asked to look at one question and answer pair. Coder 2, probably due to training as an indexer, was more exhaustive than Coder 1. This resulted in revisions to our own coding rules since we realized that we needed to define the level of representation of texts. We also found that the coders differed on *process_of* and *property_of* codings, resulting in modifications to our coding scheme definitions and a closer look at these relationships. In all, the two coders contributed significantly to improving the quality of the coding procedures.

We felt that 35% mean percent agreement for the inter-rater reliability was low but expected. Having coders identify concepts and relationships together and obtaining exact agreement is quite difficult. For this to happen 24 times in the texts they coded once with minimal training was actually quite surprising. We felt that with additional training and practice, we could achieve a much higher agreement rating.

How two or more concepts are related to one another can lead to philosophical questions that can be difficult to resolve. The purpose of this work was not to construct an ontology of medical semantic relationships for consumer health question answering. We made the assumption that the UMLS Semantic Network relationships would be a valid set to build from and that it would be adequate for representing the texts. Our intent was to create a useable knowledge base that will allow us to characterize these texts in order to lay a foundation for the construction of consumer health retrieval systems.

5.3. Future work

The perspective taken from the onset was that the analysis results and products (both the semantic relationship classes and instances) would be applied to future experimentation with systems such as health-consumer concept exploration and question formulation interfaces. The next stage of research might focus on how semantic information presented in the user interface affects laypersons' cognitive structures and abilities to express their information needs. Further work will determine whether and how views of a knowledge structure help health consumers to understand the context of the search, aids in understanding, and facilitates construction of correct mental models.

This work also outlines a methodology for analyzing semantic characteristics of texts. We completed a small sample of question—answer pairs for this study due to the fact that manual extraction is a difficult task and we did not know whether any useful results would emerge. Future work will concentrate on automating extraction of semantic relationship instances from a larger set of question—answer pairs. Use of natural language processing systems, such as SemRep [34] and AQUA [35] that use the UMLS to recover semantic propositions from biomedical texts can facilitate this process. Extracting semantic information from many question—answer pairs will make it possible to repeat our analysis and statistically determine additional patterns that result in rules for finding relevant texts, for example, through search of specific paths of connections within conceptual graph representations.

6. Conclusions

The study identified some important findings concerning the semantic representation of consumers' questions and physicians' answers in a quasi-formal electronic environment. We determined that both health consumers and physicians express causative relationships and these play a key role in leading to further related concepts in the answer texts. Regarding the answering strategy of physicians, a direct correspondence between the semantic representations of the questions and those from the answers exists only about 30% of the time. These results might be exploited to help determine the typical characteristics of answers to be used in retrieval and might be used to present related concepts to users who are browsing medical information.

Acknowledgements

The authors would like to thank Keith Cogdill, Stephen B. Johnson, Kent L. Norman, and Marilyn D. White for their help with this study. This work was supported in part by the Beta Phi Mu Doctoral Dissertation Fellowship, the Eugene Garfield Doctoral Dissertation Fellowship, and National Library of Medicine (NLM) Training Grant LM07079-11.

References

[1] Q. Zeng, S. Kogan, N. Ash, R. Greenes, A. Boxwala, Characteristics of consumer terminology for health information retrieval, Meth. Inform. Med. 41 (2002) 289—298.
[2] Q.T. Zeng, S. Kogan, R. Plovnick, J. Crowell, R.A. Greenes, E.M. Lacroix, Positive attitude and failed queries: an exploration of the conundrums of consumer health information retrieval, Int. J. Med. Inform. 73 (1) (2004) 45—55.
[3] J.L. Bader, M.F. Theofanos, Searching for cancer information on the internet: analyzing natural language search queries, J. Med. Internet Res. 5 (4) (2003) e31.
[4] D. Lindberg, B. Humphrey, A. McCray, The Unified Medical Language System, Meth. Inform. Med. 32 (1993) 281—289.
[5] A. McCray, W. Hole, The scope and structure of the first version of the UMLS Semantic Network, Proc. Annu. Symp. Comput. Appl. Med. Care (1990) 126—130.
[6] D. Soergel, Some remarks on information languages, their analysis and comparison, Inform. Storage Retrieval 3 (1967) 219—291.
[7] R. Green, Relationships in the organization of knowledge: an overview, in: C.A. Bean, R. Green (Eds.), Relationships in the Organization of Knowledge, Kluwer Academic Publishers, 2001, pp. 3—18.
[8] N. Guarino, C. Welty, Identity and subsumption, in: R. Green, C. Bean, S. Hyon Myaeng (Eds.), The Semantics of Relationships: An Interdisciplinary Perspective, Kluwer Academic Publishers, 2002, pp. 111—126.
[9] R.J. Brachman, What IS-A is and isn't: an analysis of taxonomic links in semantics networks, IEEE Comput. 16 (10) (1983) 30—36.
[10] J.F. Allen, Time and time again: the many ways to represent time, Int. J. Intell. Syst. 4 (1991) 341—356.
[11] J. Odell, Six different kinds of composition, J. Object-Oriented Programm. (1994) 8.
[12] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. Miller, Five papers on WordNet, Int. J. Lexicogr. 3 (4.) (1990), special issue.
[13] J. Rogers, A. Rector, GALEN's model of parts and wholes: experience and comparison, in: J.M. Overhage (Ed.), American Medical Informatics Association Symposium, Hanley & Belfus Inc., Philadelphia, 2000, pp. 714—718.
[14] A.L. Rector, C. Wroe, J. Rogers, A. Roberts, Untangling taxonomies and relationships: personal and practical problems in loosely coupled development of large ontologies, in: Proceedings of the K-CAP'01, ACM, Victoria, BC, 2001, pp. 139—146.
[15] L.T. Detwiler, E. Chung, A. Li, J.L. Mejino, A.V. Agoncillo, J.F. Brinkley, C. Rosse, L.G. Shapiro, A relation-centric query engine for the foundational model of anatomy, in: Proceedings of the MedInfo'2004, San Francisco, CA, 2004, pp. 341—345.
[16] B. Smith, C. Rosse, The role of foundational relations in the alignment of biomedical ontologies, in: Proceedings of the MedInfo'2004, San Francisco, CA, 2004, pp. 444—448.
[17] A. Chakravarthy, K. Haase, NetSerf: using semantic knowledge to find internet information archives, in: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, 1995, pp. 4—11.
[18] A.S. Pollitt, M.P. Smith, M. Treglown, P.A.J. Braekevelt, View-based searching systems—progress towards effective disintermediation, in: Proceedings of the 20th International Online Information Meeting, London, December, Learned Information Limited, 1996, pp. 433—446.
[19] C. Khoo, S.H. Myaeng, Identifying semantic relations to text for information retrieval and information extraction, in: R. Green, C.A. Bean, S.H. Myaeng (Eds.), The Semantics of Relationships: An Interdisciplinary Perspective, Kluwer Academic Publishers, Dordrecht, 2002, pp. 161—180.
[20] C. Khoo, S.H. Myaeng, R. Oddy, Using cause-effect relations in text to improve information retrieval precision, Inform. Process. Manage. 37 (1) (2001) 119—145.
[21] T. Miles-Board, S. Kampa, L. Carr, W. Hall, Hypertext in the semantic web, in: Proceedings of the HT'01, Aarhus, Denmark, 2001.
[22] E. Mendonca, S.B. Johnson, Y.H. Seol, J.J. Cimino, Analyzing the semantics of patient data to rank records of literature retrieval, in: Proceedings of the ACL-02 Workshop on Natural Processing in the Biomedical Domain, June 11, Association for Computational Linguistics, Philadelphia, PA, 2002.
[23] E.A. Mendonca, J.J. Cimino, Automated knowledge extraction from MEDLINE citations, Proc. AMIA Symp. (2000) 575—579.
[24] C.A. Brandt, L.A. Fernandes, S.M. Powsner, K.J. Ruskin, P.L. Miller, Using coded semantic relationships to retrieve and structure clinical reference information, in: Proceedings of the 1997 American Medical Informatics Association Annual Fall Symposium, Nashville, TN, October 25—29, 1997.
[25] S. Bechhofer, R. Stevens, G. Ng, A. Jacoby, C.A. Goble, Guiding the user: an ontology driven interface, in: N.W. Paton, T. Griffiths (Eds.), UIDIS: Workshop on User Interfaces to Data Intensive Systems, Edinburgh, 1999, pp. 158—161.
[26] W. Kintsch, Comprehension: A Paradigm for Cognition, Cambridge University Press, New York, 1998.
[27] D. Jonassen, K. Beissner, M. Yacci, Structural Knowledge: Techniques for Representing, Conveying, and Acquiring Structural Knowledge, Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
[28] Medical Library Association, MLANET: A User's Guide to Finding and Evaluating Health Info on the Web. http://www.mlanet.org/resources/userguide.html.
[29] J.F. Sowa, Conceptual graphs summary, in: P. Eklund, T. Nagle, J. Nagle, L. Gerholz (Eds.), Conceptual Structures: Current Research and Practice, Ellis Horwood, 1992, pp. 3—52.
[30] L. Slaughter, Semantic relationships in health consumer questions and physicians' answers: a basis for representing medical knowledge and for concept exploration interfaces, Doctoral Dissertation, University of Maryland at College Park, 2002, Dissertation Abstracts International.

[31] L.A. Slaughter, D. Soergel, How physicians' answers relate to health consumers' questions, in: Proceedings of the American Society for Information Science and Technology (ASIST), 2003, pp. 28—39.
[32] H. Yu, C. Friedman, A. Rhzetsky, P. Kra, Representing genomic knowledge in the UMLS Semantic Network, Proc. AMIA Symp. (1999) 181—185.
[33] D. Soergel, T. Tse, L. Slaughter, Helping healthcare consumers understand: an ''interpretive layer'' for finding and making sense of medical information, Medinfo (2004) 931—935.
[34] T.C. Rindflesch, M. Fiszman, The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text, J. Biomed. Inform. 36 (6) (2003) 462—477.
[35] S.B. Johnson, A. Aguirre, P. Peng, J. Cimino, Interpreting natural language queries using the UMLS, Proc. Annu. Symp. Comput. Appl. Med. Care (1993) 294—298.