A Linguistic Analysis of Question Taxonomies

Jeffrey Pomerantz
School of Information and Library Science, University of North Carolina at Chapel Hill, Manning Hall, CB #3360, Chapel Hill, NC 27599-3360. E-mail: pomerantz@unc.edu

Recent work in automatic question answering has called for question taxonomies as a critical component of the process of machine understanding of questions. There is a long tradition of classifying questions in library reference services, and digital reference services have a strong need for automation to support scalability. Digital reference and question answering systems have the potential to arrive at a highly fruitful symbiosis. To move towards this goal, an extensive review was conducted of bodies of literature from several fields that deal with questions, to identify question taxonomies that exist in these bodies of literature. In the course of this review, five question taxonomies were identified, at four levels of linguistic analysis.

Introduction

The growth in the past decade of both the infrastructure and the number of users of the Internet has enabled a corresponding growth in the number of users of digital reference services on the Internet. This increase in the use of digital reference services has led to increases in the number of questions received by these services, thus putting a strain on the human intermediaries in these services. Both the ability of a reference service to scale up to handle an increasingly large number of questions, and the quality of the answers provided, is directly affected by the extent of automation employed by that service: The more processes that are automated, the more of the human intermediaries' time and effort can be dedicated to tasks that cannot yet be automated. There is, now more than ever, an increased and immediate need for automation in digital reference services.

A good deal of research and development has been conducted on the automation of one of the most important tasks in providing digital reference service: the task of answer formulation. Unfortunately, this research and development has not been within the arena of digital reference. Rather, it has been within the arena of question answering (QA). Question answering research grew out of the development of story understanding systems; perhaps the first attempt to automate the process of formulating answers to questions was Lehnert's (1978) system named QUALM. QUALM attempted to replicate the process by which humans understand and answer questions. These questions concerned short stories of a few simple sentences, made up mostly of facts (for example: "John went to a restaurant. The hostess seated John.") (p. 19).

More recently, the Text REtrieval Conference (TREC) Question Answering track has been the leading forum for systems development in automated QA (Voorhees, 1999, 2000, 2001, 2002). The task set to systems participating in the 1999 TREC QA track was to answer a set of questions, each having an answer that could be located in full in at least one document in a corpus. In the 2002 TREC QA track, the tasks were rather more sophisticated: an answer did not necessarily exist in the document corpus at all, and if an answer did exist it might require combining information located in more than one document.

As the TREC QA track has developed, it has become more realistic—that is, more like the task of answering real questions that might be asked by real people. In this way, the TREC QA track has become more like the task of providing digital reference service. In providing reference service—digital or desk—there is likewise no guarantee that an answer exists, and if it does, it may require integration of information from multiple sources. While the TREC QA track is as yet a long way from the complexity that is possible in reference service, it is on its way. The roadmap document prepared by the National Institute of Standards and Technology (NIST) (Burger et al., 2001) proposes a path for the future development of the TREC QA track that will make it more realistic still, by requiring that systems implement a number of sophisticated capabilities including taking the context of previously answered questions into account, formulating answers by generating complete sentences including justification of the answer, and providing "expert-level" answers to "expert-level" questions. Thus, as the TREC QA track and the systems that participate in it develop, it becomes conceivable that such a QA system may be useful for answering questions in the environment of a digital reference service.

Received December 12, 2003; revised March 22, 2004; accepted May 17, 2004

 ● Published online 16 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20162

Indeed, digital reference and QA systems have the potential to arrive at a highly fruitful symbiosis: Digital reference services could provide a useful test bed for future QA systems implementing increasingly sophisticated functionality, and QA systems could automate some answering of reference questions, thus enabling a digital reference service to scale up to handle an increasingly large number of questions.

The NIST Roadmap document (Burger et al., 2001) provides a vision of the direction in which QA research and development should take over a 5-year span. Part of this vision involves a need for question taxonomies as the basis for any future theory of QA, and the identification of "criteria along which question taxonomies should be formed" (p. 7). To answer a question, it is argued in the Roadmap document, the question must first be understood. A component of the understanding process, it is further argued, is the classification of the question. This question classification may not be a conscious process, but it is a necessary cognitive process for the answerer to determine the appropriate content and form of the answer. For example, to respond to the question "Do you have the time?" with the answer "Yes," indicates a misclassification of the question both in terms of the appropriate content and form of the answer: The appropriate content of an answer to this question is a statement of the time, rather than yes or no, and the appropriate form is to perform an action, specifically to look at one's watch and to speak the time, rather than to provide a factual response.

To develop TREC-like QA systems for use in digital reference services, it will be necessary to identify taxonomies that may be utilized to classify questions received by digital reference services. One of the fundamental tasks in library work is classification, and taxonomies already exist that have been used to classify questions in reference services. The question posed herein is: What criteria exist along which taxonomies of digital reference questions have been formed? To answer that question, existing classification schemes for questions are identified, as well as the level of linguistic analysis along which these schemes classify questions.

Fields of Literature Reviewed

An extensive review was conducted of bodies of literature from several fields that deal with questions. This review was conducted to identify question taxonomies that exist in these bodies of literature. The fields that deal with questions that were reviewed for this study were question answering, desk and digital reference, and linguistics. Each of these fields deals with questions for different reasons and treats questions as quite different sorts of entities.

Question Answering

The field of QA treats questions as something to which one and only one correct answer exists. The answer may not be found by any given QA system, and the answer may not even exist in the document corpus available to the system—but an answer exists, and if the corpus is correct and the system is good then it will be found. This perspective on questions is summed up quite succinctly in the track report from the first TREC QA track in 1999:

> The goal in the QA task is to retrieve small snippets of text that contain the actual answer to a question rather than the document lists traditionally returned by text retrieval systems. The assumption is that users would usually prefer to be given the answer rather than find the answer themselves in a document. (Voorhees, 1999, p. 77)

The assumption that users would prefer to be given an answer rather than find an answer themselves may be debatable, but more important is that this assumption rests on another assumption: that an answer can be contained in a small snippet of text. As discussed above, more recent TREC QA tracks require the integration of information from multiple sources in formulating answers, but the assumption is still that answers may be formulated in small snippets–though the 2002 QA track required a list of small snippets, rather than a single small snippet (Voorhees, 2002).

Desk and Digital Reference

Library reference, both desk and digital, on the other hand, has a very different approach to what a question is and what is the best way to answer one. Indeed, the approach to questions taken by reference librarians is nearly diametrically opposed to that taken by QA systems. First, while QA systems accept the initial query put to the system as the question to be answered, it is widely acknowledged in library reference that a patron's initial question is not actually the question he or she means to ask (see, for example, Eichman, 1978; Lynch, 1978; Taylor, 1968). Belkin, Oddy, and Brooks' (1982) well-known ASK hypothesis states that an individual's information need arises from an "anomaly" or gap in that individual's "state of knowledge," but that the individual is generally unable to articulate what that gap is, or what would be required to fill it. Taylor (1968) outlines four levels of question formation: The first of these, the visceral need, is analogous to the "anomalous state of knowledge." The second level, the conscious need, is the point at which the individual is able to articulate his or her information need, but as Taylor states, "it will probably be an ambiguous and rambling statement" (p. 182). This inability to clearly articulate a question is unsurprising. Miyake and Norman (1979) suggest that to even formulate a question, an individual must possess a framework of knowledge about the subject area in which the question occurs. The role of the reference librarian has therefore been referred to as "mind-reading" (Lynch, 1978)—to provide an answer when the patron may not even be able to articulate the question.

Another way in which library reference differs from QA systems is in the view taken of answers. While QA treats questions as something to which one and only one correct answer exists, library reference acknowledges that what is an acceptable answer varies depending on the patron's current

situation. This position is fundamental to the theory of sense-making, made popular by Dervin and colleagues (Dervin, 1977, 1983; Dervin & Dewdney, 1986): While different individuals may formulate similar or even identical questions, the content and form of an answer that will be useful to these individuals may differ depending on the different situations in which these individuals find themselves. There is thus limited utility in attempting to find *the* answer to a question; instead, a *useful* answer is one that helps an individual move through the situation that gave rise to the question, and what is useful may differ for every individual and situation.

Linguistics

Linguistics takes yet a different approach to the nature of questions. A common framework is the breakdown of linguistic analysis into seven distinct levels. These levels are described succinctly by Liddy (1998, pp. 14–15):

7. Pragmatic: Understanding the purposeful use of language in situations, particularly those aspects of language which require world knowledge;
6. Discourse: Interpreting structure and meaning conveyed by texts larger than a sentence;
5. Semantic: Determining the possible meanings of a sentence, including disambiguation of words in context;
4. Syntactic: Analysis of words in a sentence in order to uncover the grammatical structure of the sentence;
3. Lexical: Word level analysis including lexical meaning and part of speech analysis;
2. Morphological: Componential analysis of words, including prefixes, suffixes and roots;
1. Phonological: Interpretation of speech sounds within and across words.

The lower three levels of analysis deal with units the size of the word and smaller; the upper four levels of analysis deal with units of the size of the sentence and larger. In their discussion of a formal logic of questions, Prior and Prior (1955) state that a question is a sentence, just as a proposition is—but that a proposition "affirms or denies," while a question poses a condition (p. 43). Graesser and Black (1985), in their book about questions and the psychological mechanisms that generate them, refer to questions as "sentences that call for a reply" (p. 3). It is only possible to analyze questions according to the upper four levels of analysis, as a question can only be understood at the level of the sentence or larger; outside of a question, individual words lose their interrogative power. Even that most quintessential word of questioning, "why," cannot be understood outside of an interrogative sentence: The question "Why?" implies a context which disambiguates of the question. For example, the question "Why?" is meaningless until one knows that, in context, the question may be understood as, for example, "Why is the sky blue?" or "Why does the Bush administration support a Constitutional amendment banning same-sex marriages?"

The difference between a declarative and an interrogative sentence is addressed by Speech Act Theory, which outlines three layers of a speech act: (a) *locutionary acts* are acts of uttering a series of sounds that have meaning, (b) *illocutionary acts* have force and convey the speaker's meaning, and (c) *perlocutionary acts* produce effects in the listener (Austin, 1999; Searle, 1995). Thus, a question has the illocutionary force of conveying the speaker's information need, and the perlocutionary force of requiring a response from the listener.

A question also assumes that there is, in fact, a listener from whom a response will be forthcoming. Indeed, discourse analysis presupposes the existence of feedback and posits a listener who is an active participant in a conversation (Clark & Wilkes-Gibbs, 1986). A question, then, is like any speech act, in that it cannot stand alone as an utterance, but requires a reaction or response from the listener, followed by a response to the listener's response (Roberts & Bavelas, 1996). At the higher levels of linguistic analysis, then, a question does not stand alone, but rather is one step in a larger conversation. Indeed, a question cannot be understood only from the initial speech act; a question can only be understood by participating in or observing a process of communicative collaboration.

Thus, we have three different perspectives on questions—from QA, reference, and linguistics. Question answering takes a prima facie approach to questions, assuming that the way in which they are stated is perfectly clear, needs no disambiguation, and perfectly conveys an information need. Linguistics treats questions as individual acts within a larger conversation, and that these acts convey the meaning that the speaker has an information need, and compels the listener to respond. The approach to questions taken by reference is an application of the approach from linguistics: A question cannot stand on its own, but through conversation may be disambiguated and ultimately responded to.

Survey of Existing Question Taxonomies

These three fields were reviewed to identify question taxonomies because all three contain a body of work in which types of questions are identified, and questions are classified according to one or more schemes. The subsequent section is a review of these classification schemes.

In the course of this review, five question taxonomies were identified in these bodies of literature. These five question taxonomies are as follows:

1. Wh- words
2. Subjects of questions
3. The functions of expected answers to questions
4. The forms of expected answers to questions
5. Types of sources from which answers may be drawn

Wh- words

"The Five Ws" is a simple and common classification of questions in English. It is learned in school at an early age as the standard way to construct a question in English (at least it was when the author was in school at an early age). The Five Ws are also a tradition of journalistic writing style, going hand-in-hand with the "inverted pyramid," i.e., in the traditional news story, the most important elements must be included in the opening paragraph. These most important elements are *Who*, *What*, *When*, *Where*, *Why*, and *How*. The fact that the Five Ws are actually five Ws and an H is merely a reflection of the perversity of spelling in English.

Perhaps because the Five Ws is such a common and intuitive way of thinking about questions (at least for English speakers), it is popular in the literature on questions and question answering. Indeed, Robinson and Rackstraw (1972a, 1972b) devote two entire volumes to an investigation of wh- words, the forms of questions based on these words, and the forms of answers to these questions. Robinson and Rackstraw define wh- words as "the total set of lexically marked interrogative words" (1972a, p. 2). By "lexical set" Robinson and Rackstraw mean a distinct group of words that can be used in "similar linguistic environments" (p. 39). Thus, for Robinson and Rackstraw, the wh- words are a set of words that can all be used to form an interrogative sentence—in other words, a question. As a corollary, any sentence containing a wh- word is a question, according to Robinson and Rackstraw.

Despite a certain elegance in the simplicity of the Five Ws, Robinson and Rackstraw (1972a) add one class to this set. Robinson and Rackstraw draw a distinction between open and closed questions and, by extension, between questions phrased using the words *What* and *Which*. According to Robinson and Rackstraw, the word "what" indicates a question about an infinite or undefined set (e.g., "What on earth are you talking about?"), while the word "which" indicates that the question asks for the identification of a particular entity out of a finite set of entities (e.g., "Which person in the lineup did you see rob the bank?").

The taxonomy of wh-questions, according to Robinson and Rackstraw (1972a), is as follows:

- Who
- Which
- What
- When
- Where
- Why
- How

There are, however, two problems with limiting the domain of questions to only sentences containing a wh- word: (a) questions are not necessarily phrased using a wh- word, and (b) statements phrased using a wh- word are not necessarily questions. That is, a question can be phrased as a statement, but be interpreted and accepted as a question. For example, the statement "I'm looking for the name of the Secretary of State under Clinton," may reasonably be interpreted as a question asking, "Who was the Secretary of State under Clinton?" Additionally, a statement phrased as a question may be rhetorical (e.g., an interjection such as "What?" to indicate surprise) or a gripe phrased as a question (e.g., "Why me?") (Graesser & Black, 1985, p. 27).

Several TREC QA systems use wh- words as the primary criterion for the analysis and logical representation of questions (Hovy, Hermjakob, and Lin, 2001; Kwok, Grunfeld, Dinstl, & Chan, 2001; Moldovan et al., 1999). Some of these systems subdivide wh-types—for example, into such classes as WHY-FAMOUS-PERSON and ABBREVIATION-EXPANSION. The purpose of this subdivision of wh-types is to enable QA systems to recognize "the semantic type of the desired answer" (Hovy et al., 2001, p. 167). These "semantic types" combine the subject of a question and the information about that subject about which the question is asking. To use the WHY-FAMOUS-PERSON question "What is Jane Goodall famous for?" as an example—the wh-type is Why, the subject is Jane Goodall, and the question is asking about the cause or reason for her fame.

Such QA systems treat this subdivision as hierarchical. The first level of classes are wh- words, and the second level are semantic types. In fact, it may be more accurate to treat this classification of questions as a simple faceted scheme: One facet is wh- words, one facet is subject, and one facet is the function of the desired answer or the gap (Dervin, 1983) that the answer should fill. Question answering systems conflate these facets because it may be simpler to write algorithms that way, and additionally, in system-building, being theoretically rigorous is not the most important criteria—rather, what is most important is building a system that works.

Without perhaps meaning to, however, these TREC QA systems have moved beyond the simple taxonomy of wh-words, and have made some strides towards our next two question taxonomies: subjects of questions and the functions of answers.

Subjects of Questions

Many classification schemes exist that organize entities according to their subjects. These frequently take the form of thesauri, both general (e.g., the Library of Congress Subject Headings, Library of Congress, 1992; the Dewey Decimal Classification, Mitchell, Beall, Matthews, & New, 1996; and the Sears List of Subject Headings, Mooney, 1991) and subject-specific (e.g., the Art & Architecture Thesaurus, http://www.getty.edu/research/conducting_research/vocabularies/aat/; the ASIS Thesaurus of Information Science, http://www.asis.org/Publications/Thesaurus/tnhome.htm; the ERIC Thesaurus, http://www.ericfacility.net/extra/pub/thessearch.cfm, and the Medical Subject Headings, MeSH, http://www.nlm.nih.gov/mesh/).

Organization by subject has been a common means for classifying documents at least since Melvil Dewey first conceived of his subject scheme in 1873, and probably for a

long time before that. Once the leap was made in libraries to thinking about other types of artifacts as intellectual entities (such as art and architecture), then it was a smaller step to thinking about questions (non-print and indeed, immaterial entities) as intellectual entities that could be classified. And indeed, organization by subject was the first approach taken to the classification of questions. Perhaps the earliest example of a classification scheme for questions dates back three quarters of a century. Conner (1927) used the 10 main classes of the Dewey Decimal Classification (DDC) to classify questions recorded by the Reference Department of the Carnegie Library of Pittsburgh. What is perhaps most interesting about Conner's classification is that she applies the same scheme that is used to classify materials in the library to also classify reference questions. Conner analyzes the percentages of questions received by the reference department in her library to make recommendations as to the content of courses in library school, so that future librarians would be prepared to answer questions on the most commonly asked subjects. The assumption implicit in Connor's classification is that the same scheme used to classify materials from which answers may be derived is appropriate for questions, and further, that there is a one-to-one, or at least one-to-many, correspondence between the question and the source(s) in the library's collection that contain an answer.

To be fair, reference services to this day tend to treat reference questions exclusively in terms of their subject. Many reference evaluation tools require that the subjects of questions be recorded (King, 1982; Murfin & Gugelchuk, 1987), the intention being to identify subjects on which it is difficult for reference librarians to answer questions, or to which reference librarians frequently cannot give accurate answers (Crews, 1988; Rothstein, 1964). Some of the more recent work in developing standards for evaluating digital reference services suggests that the subject of a question is an important component of data to be collected about the types of questions received by a service (McClure, Lankes, Gross, & Choltco-Devlin, 2002; White, 2001).

There are many subject classification schemes in existence, but subject-based schemes for classifying questions fall into two categories: those based on pre-existing document classification schemes, and "home-grown" schemes. The former type of question classification scheme is based on the same set of assumptions made by Conner (1927), i.e., that there is a correspondence between the question and the source(s) that contain an answer. The pre-existing schemes used in these cases are the same schemes that the reference service utilizes to organize other resources. The AskERIC service, for example, organizes their archive of previously answered questions (PAQs) (http://askeric.org/Virtual/Qa/archives/) according to the ERIC thesaurus. The Virtual Reference Desk Learning Center (http://vrd.askvrd.org/) organizes their PAQ archive based on the subject elements and subelements of the Gateway to Educational Materials[SM] (GEM) metadata set. On the other hand, "home-grown" schemes are developed by a specific reference service, presumably to suit the types of questions received by that service specifically. Some examples of home-grown subject schemes include the organization of the Internet Public Library (IPL)'s Frequently Asked Reference Questions (http://www.ipl.org/div/farq/), the list of possible subjects listed on Ask Joan of Art's question submission Web form (http://nmaa-ryder.si.edu/study/reference-question.html—see the dropdown list under "Select the subject of your question"), and the categories according to which Ask Dr. Math's archive is organized (http://mathforum.org/dr.math/).

The Functions of Expected Answers

While the previous two taxonomies—wh- words and subjects—are purely classifications of questions, this taxonomy also considers the answer to the question, in classifying the question. Or more accurately, this taxonomy considers the answer that is expected, before it is actually formulated—as reference librarians know, an answer often turns out to be more complex than is apparent a priori, from consideration of the question on its face.

As discussed above, a discourse analytic approach to questions presupposes a response and a listener from whom that response will be forthcoming. The response is compelled by the illocutionary force of the question, which conveys the speaker's information need. The future existence of an answer is therefore assumed when a question is asked. In this spirit, Hermjakob (2001) suggests that to determine how best to answer a question, "it is important to classify questions with respect to their answer types" (p. 18). The example that Hermjakob provides is the question, "How tall is Mt. Everest?" This question cannot be answered unless it is first understood that the desired answer will consist of a measurement quantity. The speaker's information need leads to the speaker formulating a question, and the question compels an answer whose function is to fulfill that information need. It is this information need/answer function pairing that is the basis for the present taxonomy. Graesser, McMahen, and Johnson (1994) state that this taxonomy defines question classes "on the basis of the content of the information sought" by the question (p. 520). The term *content*, however, is ambiguous in that it can mean any type of content, both subject matter and the information concerning the subject about which the question is asking. Rather, the function of an expected answer is similar to the "semantic types" discussed above, utilized by some TREC QA systems, minus the subject: The function of an expected answer is the function the answer fulfills in meeting the questioner's information need.

This taxonomy was originally developed by Lehnert (1978) for use in her story-understanding system QUALM. Subsequently, this taxonomy was adopted by Graesser and colleagues for several studies analyzing questions asked by individuals in a variety of real-world settings: while reading texts, while learning a new computer system, and while watching television news (Graesser, Lang, & Horgan, 1988). Over time, Graesser and colleagues (Graesser, Person, & Huber, 1992; Graesser, McMahen, and Johnson, 1994)

TABLE 1. The taxonomy of functions of expected answers.

Class	Abstract specification
Short Answer	
Verification	Is a fact true? Did an event occur?
Disjunctive	Is X or Y the case? Is X, Y, or Z the case?
Concept completion	Who? What? When? Where? What is the reference of a noun argument slot?
Feature specification	What qualitative attributes does entity X have?
Quantification	What is the value of a quantitative variable? How many?
Long Answer	
Definition	What does X mean?
Example	What is an example label or instance of the category?
Comparison	How is X similar to Y? How is X different from Y?
Interpretation	What concept or claim can be inferred from a static or active pattern of data?
Causal antecedent	What state or event causally led to an event or state?
Causal consequence	What are the consequences of an event or state?
Goal orientation	What are the motives or goals behind an agent's action?
Instrumental/ procedural	What instrument or plan allows an agent to accomplish a goal?
Enablement	What object or resource allows an agent to perform an action?
Expectational	Why did some expected event not occur?
Judgmental	What value does the answerer place on an idea or advice?
Assertion	The speaker makes a statement indicating that he lacks knowledge or does not understand an idea.
Request/directive	The speaker wants the listener to perform an action.

developed a theoretical model of question asking behavior. This taxonomy reached its most fully developed form in Graesser, McMahen, and Johnson's 1994 work. In this developed form, this taxonomy is divided into classes that require short versus long answers. This taxonomy, as presented by Graesser, McMahen, and Johnson, is given in Table 1.

Lehnert (1978) and Graesser and colleagues utilize this taxonomy as a theoretical framework: Lehnert for the construction of a story-understanding system, and Graesser and colleagues for studying questioning behavior. It is left to other researchers to utilize this taxonomy to classify actual questions from real questioners. Curiously, most of the use of this taxonomy has been in medical environments. Keyes (1996) utilized this taxonomy to classify a set of queries associated with a database of documents on cystic fibrosis. Stavri (1996) utilized this taxonomy to classify physicians' questions during the process of making preliminary diagnoses. Smith (2002) utilized this taxonomy to classify artificial questions simulating those that are asked by clinicians about patients. These three studies demonstrate the usefulness of this taxonomy for classifying questions "from the field," as it were, from real questioners with real information needs. White (1998), however, was the first to utilize this taxonomy to analyze questions asked at reference desks, both by the patron and by the reference librarian.

The taxonomy of functions of expected answers, as discussed above, is not truly a classification of questions, but is rather a hybrid, a classification of question–answer pairs. The following two taxonomies, on the other hand, are purely classifications of answers.

The Forms of Expected Answers

The need for standards for measurement and evaluation of reference services has been recognized in the library profession for some time. In the mid-1970s the Library Administration and Management Association (LAMA), a division of the American Library Association (ALA), created standard definitions for two types of reference transactions for inclusion in their Library General Information Survey (LIBGIS) (U.S. Department of Education, 1981). These two types of reference transactions are as follows:

- Reference transactions: "involves the knowledge, use, recommendation, interpretation, or instruction in the use of one or more information sources by a member of the reference/information staff."
- Directional transactions: "provides assistance in finding and using library services, collections and facilities" (White, 1981, pp. 33–34).

The LIBGIS definitions were the first standardization of types of reference transactions, and for the first time provided a classification (simple as it is) of the types of services provided at a reference desk. Also for the first time, the LIBGIS enabled reference services at different libraries, holding different collections and serving different communities of patrons, to share reference statistics (Kaske & Aluri, 1980). Additionally, researchers conducting evaluation studies of reference services could utilize the LIBGIS classes (Herner, Vellucci, & Leyman, 1972; Phenix, 1983).

The classes "reference" and "directional" are, however, extremely broad. As a result, some researchers and libraries divided these classes into a variety of subclasses (White, 1981; Phenix, 1983). Rothstein (1964), predicting the classification to come, discusses grouping questions into the following types: directional, ready reference, search (or research), and readers' advisory (p. 458). Seng (1978) discusses three question types: first, direction, and then two that are subclasses of the LIBGIS reference class, but which Seng defines in a unique way: information (a question that "is concerned with information resources and/or their use"), and general (a reference question "answered through the use of information resources") (p. 22). Brown (1985) drops the directional class entirely, and divides questions into informational (any question that can be answered using ready reference sources such as the card catalog or telephone directory) and reference (any question that requires non-ready reference sources to answer it) (p. 294). Fogarty (1995) discusses the following four

types: directional, instructional, ready reference, and extended reference (p. 20).

These variations on the LIBGIS theme demonstrate that even given a standard, different services will modify and extend that standard to accommodate their specific situation and requirements. Equally interesting is the amount of "convergent evolution" that has occurred surrounding this taxonomy. Several researchers and libraries explicitly modified the LIBGIS classes. Equally many, however, independently developed question classification schemes that resembled the LIBGIS scheme, and either did not know of the existence of the LIBGIS scheme, or simply did not mention it (Bunge, 1990; Carter & Janes, 2000; Dewdney & Mitchell, 1996; Lynch, 1978; Stalker & Murfin, 1996). Looking across all of these variations on a theme, a taxonomy of the forms of the expected answer to a question can be constructed as detailed in Table 2.

The fact that this and the next taxonomy are so prevalent in the literature on reference is evidence that classification is a necessary component of the process of understanding a question. As Burger and others (2001) argue, the cognitive process of classifying questions may be unconscious, but it is necessary for understanding a question. The effort on the part of LIBGIS and its successors to formalize the taxonomy of forms of expected answers is an indication that part of the process of understanding a question for a reference librarian is to classify that question according to the task that the librarian must perform to formulate an appropriate answer. Reference librarians are trained to interview patrons to "get to the bottom" of their information need: to identify the patron's real question (on the assumption that the patron's initial question is often not their real question), and the patron's expectations about the form in which an answer will be presented. It is therefore natural that as a result reference librarians should come to view the process of question answering as being tied to the form in which the answer is formulated, and the amount of work that will be required to formulate that answer. This taxonomy takes both of those things into account: To classify a question according to the form of an expected answer to that question (before it is actually formulated), one must consider both the presentation of the answer and the amount of work that will be necessary to formulate an answer that may be presented in that form.

TABLE 2. The taxonomy of forms of expected answers.

Class	The form of the expected answer
Directional	Questions asking about the location of a specific information source.
Holdings	Questions about whether a specific information source or document is owned by the library.
Ready reference	Questions asking for simple, factual answers; the answer should be readily ascertainable from available information sources.
Exact reproduction	Questions asking for pictorial and textual materials, taken directly from an information source and unchanged.
Description	Questions asking for a description of something, briefer in length than the original thing (basically, an abstract).
Readers advisory	Questions asking for assistance in the choice of books or the gathering of data.
Bibliographic instruction	Questions asking for assistance in use of information source(s).
Research	Questions asking for involved answers; the answer should require some effort and wide use of information sources to formulate.
Citation list	Questions asking for a list of information sources on a particular subject.
Analysis	Questions asking for some form of data analysis, whatever that data might be—scientific, social, financial, etc. Questions of this type might ask for trends, pro or con arguments, cause and effect, compare and contrast, etc.
Critique	Questions asking for an evaluative discussion of a particular subject. (e.g., a movie review, Cliffs notes-like analyses of a book, etc.)

Types of Sources From Which Answers May Be Drawn

The final taxonomy that will be reviewed is another classification of answers rather than of questions. This taxonomy is of types, or genres of information sources in which answers to questions may be located. This taxonomy reinforces the point made above that part of the process of understanding a question, for a reference librarian, is to classify that question according to the task that must be performed to formulate an answer. The reference librarian's task that is the basis for this taxonomy is the consultation of one or more information sources in performing the action of formulating an answer.

This taxonomy has not, to the author's knowledge, actually been used to classify questions, but it is included here because it is prevalent in the literature on library reference. Taylor (1968) states "when an inquirer approaches the reference desk, he has some picture in mind as to what he expects his answer to look like, i.e., format, data, size, etc." (p. 187). Taylor goes on to state that the job of the reference librarian, through the vehicle of the reference interview, is to alter the patron's a priori notion of what the answer should "look like." In many cases, the patron's notion gives way to the librarian's notion of what the answer should look like. It could be argued that this is as it should be, since the librarian is the expert in answering reference questions and knows what sorts of answers can be provided using the library's collection. Others, such as Dervin (1977) and White (1989), argue that because the process of query negotiation is a process of communication, the librarian and patron should arrive at a common ground in terms of formulating an answer.

In either case, it is ultimately the reference librarian that guides the patron to an information source or sources, and to an answer within a source, if an answer exists. There are, of course, many types of questions for which the patron may hope that no answer exists (patent searches, for example), though even for these questions the reference librarian would still guide the patron to information sources. To accomplish this, reference librarians' training includes acquiring an understanding of the variety of different genres of

information sources that exist, and the sorts of information that can be located in each one. Indeed, this classification of answer sources is so ingrained in reference librarianship that several textbooks on reference librarianship devote entire chapters to each of these genres (Bopp & Smith, 2001; Katz, 1997a; Slavens, 1985), and some reference evaluation instruments ask the librarian to specify the genre of the source in which the answer was located (King, 1982). Richardson (1995) suggests that "the essential reference task" is the classification of a question according to these genres. This statement illustrates how deeply ingrained these genres are in the practice of reference librarianship. Richardson describes 12 reference formats (p. 156):

1. Abstracts
2. Atlases
3. Bibliographies
4. Biographical sources
5. Dictionaries
6. Directories
7. Encyclopedias
8. Government publications
9. Handbooks/manuals
10. Indexes
11. Statistical sources
12. Yearbooks

Of course, not all questions may be answered using only one, or even one type of information source. An answer to a question such as "What are the consequences for libraries of the USA PATRIOT Act," for example, may be compiled from information scattered across more than one type of source: perhaps some combination of government publications, publications from nonprofit organizations, and legal documents. This taxonomy, therefore, must be treated as violating the rule of mutual exclusivity that is often considered a fundamental characteristic of classification schemes (Ranganathan, 1957).

Discussion of Existing Question Taxonomies

In this article my first goal was to identify existing classification schemes for questions. The preceding sections have presented these classification schemes. For these taxonomies to be utilized by TREC-like QA systems for digital reference services, however, it is necessary to determine which taxonomies will be the most useful for that purpose.

As discussed above, several TREC QA systems classify questions utilizing a combination of the taxonomy of wh-words, a classification of subjects, and the taxonomy of functions of expected answers. Thus, it is these taxonomies on which the most work has been done in developing actual QA systems. On the one hand, these taxonomies are the most useful for developing QA systems for digital reference services, since at this point in time the QA algorithms utilizing these taxonomies are the most sophisticated.

The taxonomy of functions of expected answers has been used for classifying questions in a variety of real-world environments, including desk reference. The taxonomy of forms of expected answers and the taxonomy of types of sources from which answers may be drawn have received even more use in library reference, having been derived directly from reference librarians' training. On the other hand, these taxonomies are the most useful for developing QA systems for digital reference services, since they are the most closely aligned with practice in reference librarianship.

The taxonomy of functions of expected answers is the only one of the five taxonomies presented above that has been utilized for question classification both by existing QA systems and library reference research. If only one taxonomy had to be selected, the taxonomy of functions of expected answers would therefore seem the logical choice to use in QA systems for digital reference services. It is not necessary, however, to select only one taxonomy for this purpose. Existing QA systems currently utilize a combination of three taxonomies; there is no reason why more taxonomies could not be utilized if it would improve their performance. While it would be difficult for a human to hold five taxonomies in his or her head and to classify a question according to all five simultaneously, a computer algorithm has no such limitation. The author suggests, therefore, that all five of the taxonomies presented above should be utilized in QA systems for digital reference services.

The reason that Burger and others (2001) call for question taxonomies is that classification, they suggest, is one step in the interpretation of questions. Burger and others state that "before a question can be answered, it must be first understood" (p. 7), and part of this understanding, this interpretation of questions, is classification of the question. Burger and others would undoubtedly agree that, since understanding is a complex task, more than one taxonomy would be useful to enable the full range of interpretations of questions.

The notion of faceted classification schemes was mentioned briefly above. I propose that the five taxonomies presented here should be treated as a faceted classification scheme. Faceted classification schemes allow entities to be classified according to several aspects of their content. Ranganathan (1965), the originator of the idea of faceted classification, proposed five "Fundamental Categories," or dimensions, along which entities could be classified: Personality, Matter, Energy, Space, and Time (PMEST). Ranganathan uses the diesel engine as an example: A diesel engine has Personality in that it "is a piece of iron made functional and endowed with a personality of its own" (p. 212). A diesel engine is an object of composed Matter, it is a manifestation of Energy, and it exists in Space and Time. Other classificationists have expanded on Ranganathan's ideas, and Vickery (1966) suggests that PMEST are not the only possible categories; facets may be any categories that are logical for the entities being classified.

The five taxonomies presented above form the facets along which it is logical for questions to be classified. But these facets are formed, as Ranganathan would have it, along dimensions. Burger and others (2001) do not use the word dimensions, but rather suggest that it is important to

determine "criteria along which question taxonomies should be formed" (p. 7). The subsequent section will explore these criteria: the dimensions along which the taxonomies presented above are formed.

Linguistic Analysis of Question Taxonomies

The units of linguistic analysis at the "bottom" three levels—the phonological, morphological, and lexical levels—are units the size of the word and smaller. It does not make sense to discuss questions at these three levels of analysis, as a question has the illocutionary force of conveying the speaker's information need only at the level of the sentence or larger. Questions can, however, be discussed at the top four levels of analysis. Since a question is a special type of sentence, it makes sense to treat questions according to units of analysis the size of the sentence and larger. Indeed, it will now be argued that the five taxonomies discussed above fall into the top four levels of linguistic analysis, as outlined in Table 3.

Several QA systems that compete in the TREC QA track utilize natural language techniques, for example: Lasso (Moldovan et al., 1999), Webclopedia (Hovy et al., 2001), and the Center for Natural Language Processing's unnamed system (Chen et al., 2001; Diekema et al., 2000, 2002). It should come as no surprise, therefore, that the five taxonomies discussed above—three of which are used by some existing QA systems (wh- words, subjects of questions, and the functions of expected answers)—correspond to the top four levels of linguistic analysis. Let us start from the bottom of these four levels and work our way up.

Syntactic Level

The syntactic level of linguistic analysis deals with language as a mathematical system of symbols that are manipulated according to a set of rules (Chomsky, 1957). Those rules compose the grammar of a language; a grammar is the set of rules according to which a sentence as a string of symbols is well-formed. Chomsky (1977) began the formalization of the use of Wh- words to indicate inquiry, but the use of Wh- words was developed into a formal logic by Prior and Prior (1955).

Prior and Prior (1955) decompose questions into a subject and a request. They use the word "subject" in an unconventional sense, to mean not the grammatical subject of the question, but the possible states of the world that are presupposed by the question. The request, therefore, identifies how many of these states are desired in the answer. For example, Whether-questions presuppose a finite set of alternatives, as in the question "Is John going home?" for which the set of possible states of the world is that John either is or is not going home. Thus, while the grammatical subject of this question is John, the subject according to Prior and Prior's formal logic of questions is John's home-going. The request, therefore, identifies that the desired answer is one that specifies which of John's alternative home-going states is true: that John either is or is not going home (Belnap & Steel, 1976, pp. 19–22). Which-questions, on the other hand, presuppose an unbounded set of alternatives, as in the question "Which person is going home?" (Belnap & Steel, 1976, pp. 22–23).

Prior and Prior's formal logic thus treats questions as a way of referring to alternative states of the world utilizing the rules of grammar: Certain presuppositions lie behind the question *qua* sentence, and the particular Wh- word used in asking the question specifies how the response is expected to address those presuppositions. The taxonomy of Wh- words falls into the syntactic level of linguistic analysis because it classifies questions according to the particular word that is used to form a sentence into a question. The Wh- word itself is the grammatical symbol that causes the "movement" from a sentence to a question, and the specific word used is important for the question to be well-formed grammatically.

TABLE 3. Levels of linguistic analysis of question taxonomies.

Level of linguistic analysis	Question taxonomy
Pragmatic	Types of sources from which answers may be drawn
	The forms of expected answers to questions
Discourse	The functions of expected answers to questions
Semantic	Subjects of questions
Syntactic	Wh- words

Semantic Level

The semantic level of linguistic analysis deals with the meaning of speech acts. According to Carnap (1957), a speech act has both intension and extension. Its intension is the proposition expressed by the speech act (e.g., the proposition that "Scott is human" [p. 27]), and its extension is the truth-value of the speech act (e.g., the truth of the proposition that Scott is human). The meaning of a speech act derives from both its intension and extension: To comprehend a speech act, one must comprehend the proposition expressed by it and its truth-value. For example, to comprehend a speech act about an "actual thing" such as Plato, one must comprehend that the speech act refers to Plato, and that Plato is an actual entity.

This act of reference is central to the use of language. Certain words are names for empirical entities, such as "dog" and "Boston," while other words are names for abstract entities, such as "three" and "good" (Ryle, 1949). There are a host of issues involved in the naming of entities and in the mechanism by which a name, once given, comes to be used (Evans, 1985; Kripke, 1980). Once a name has been given, however, it is what Kripke calls a "rigid designator": The name refers to the specified entity in all possible worlds, regardless of whatever changes the entity might undergo. Thus, once a name has been given and is being used, all users of that name are referring to the same entity.

Taxonomies of subjects of questions fall into the semantic level of linguistic analysis because they classify questions according to rigid designators—names for entities that have been assigned (through whatever means) and are used to refer to those entities. The semantic level of analysis is absolutely central to understanding the process of question negotiation in library reference. As Taylor (1968) points out, the first thing that must be negotiated between a librarian and a patron is the "determination of subject" (p. 183). It is only because rigid designators exist—only because two individuals can use the same name to refer to the same thing—that meaningful communication is possible, that one individual can communicate that he or she has a gap in his or her knowledge of a subject and another individual can assist him or her to fill that gap.

Discourse Level

While the lower five levels of linguistic analysis deal with language within the bounds of single sentences and utterances, discourse analysis deals with language as it is used in interpersonal interactions. The archetypal interpersonal interaction that discourse analysis deals with is the face-to-face conversation between two individuals. Schegloff and Sachs (1973) and Sachs, Schegloff, and Jefferson (1974) propose a model of turn-taking in face-to-face conversation, in which, as a rule, one speaker speaks at a time, overlaps are infrequent, and techniques are used by both parties to allocate turns. Roberts and Bavelas (1996) expand on this model of turn-taking and propose a three-step model based on contributions to conversation and feedback between the speakers. This model is based on the premise that meaning is negotiated between the speakers, and that therefore a minimum of three steps are necessary for any communication to be successful. Step one is the *Utterance*: simply, any single speech act, including non-verbal communication acts. The Utterance is what Austin (1999) refers to as a locutionary act, an "act of saying something" (p. 94), simply the act of uttering a series of sounds that have meaning. Roberts and Bavelas' second step is the *Reaction*: the addressee's response to the initial speech act (again, verbal or nonverbal). The Reaction reflects the addressee's interpretation of the initial speech act, which may be either correct (in that it corresponds to the speaker's interpretation of his or her own speech act), or incorrect. Roberts and Bavelas state that the Reaction "may consist of a request for clarification, a formulation or reformulation, or other explicit comment," or it may simply be an "appropriate continuation" (p. 142) to the conversation, thus signaling that the Utterance was understood and that the conversation can move on, assuming the Utterance as common ground. The third step is the *Confirmation*: the speaker's reaction to the addressee's reaction. Again, this may be verbal or nonverbal, and may be a request for clarification or a continuation. In the Confirmation, the speaker validates that the addressee's interpretation of the initial speech act was either correct or incorrect, and, with the Confirmation statement, may correct the addressee's interpretation at the same time. In this way, mutual acceptance of the initial speech act is accomplished.

The taxonomy of functions of expected answers to questions falls into the discourse level of linguistic analysis because it classifies questions according to the information need indicated by the stated question. An example will serve to illustrate this. The utterance "How does quinine work?" is a simple enough speech act on its face. But the reaction to that question may take several forms. It may be an answer, or it may be some form of clarification, such as to clarify whether the questioner wishes to know how quinine works chemically when ingested by a human being (a process answer), or how well it works under certain conditions (a value judgment answer). A reference librarian might argue that it is only when this clarification is accomplished that the original question can and should be answered. The confirmation, then, would serve to validate the interpretation of the question as being an indication of one or another category of information need: a request for an explanation of how quinine performs the actions it does when ingested, or the value of ingesting quinine under certain conditions. It is only through this turn-taking, which allows feedback to occur between the speakers engaged in a conversation, that the information need expressed by the questioner's initial utterance may be properly understood and mutually accepted as common ground.

Pragmatic Level

Pragmatics builds on discourse analysis in that it addresses the role of language beyond the conversation, and considers language use in a broad social context. Indeed, so many phenomena have been considered under the name of pragmatics that Bar-Hillel (1971) referred to it as a "wastebasket." The important consideration here is that pragmatics considers the interaction between language use and world knowledge.

The final two taxonomies discussed above fall into the pragmatic level of analysis: the taxonomies of forms of expected answers to questions, and of types of sources from which answers may be drawn. Both of these taxonomies partition entities according to their genre. Orlikowski and Yates (1994) define genre as "a distinctive type of communicative action, characterized by a socially recognized communicative purpose and common aspects of form" (p. 543). The socially recognized communicative purpose for which these taxonomies exist is the process of question answering; the common aspects of form differ for each taxonomy. As discussed above, the taxonomy of forms of expected answers classifies answers, and partitions these entities according to that content of the answer that affects the format in which the answer is provided. The taxonomy of types of sources classifies the information sources in which answers may be located, and partitions these entities according to characteristics of those sources—characteristics that are well known to reference librarians and are explicated in any number of textbooks on references service (see, for example, Bopp &

Smith, 2001; Katz, 1997a, 1997b). A considerable amount of schooling is required to teach librarians about the content contained in the various types of reference sources thoroughly enough to know which type of source to consult in answering various types of questions, not to mention the training required to completely and accurately answer questions and to structure that answer appropriately to the content. This world knowledge, shared by the community of reference librarians, has over time shaped the genres of answer forms and of information sources; these genres have likewise shaped the speech acts that reference librarians utilize in answering questions.

Discussion of Linguistic Analysis

A taxonomy was identified at each of the top four levels of linguistic analysis; it does not, however, follow that these taxonomies are the *only* ones that may exist at each of these four levels of analysis. Indeed, there are many classification schemes that currently exist at the semantic level of analysis—that is, many schemes that classify entities according to subject. The Library of Congress Subject Headings, the Dewey Decimal Classification, the ASIS Thesaurus of Information Science, and the Medical Subject Headings were among those mentioned above. There may also be equally large numbers of taxonomies at the other levels of linguistic analysis, which simply have not yet been developed. Further, it is possible to imagine question taxonomies formed along nonlinguistic dimensions. Just as the five taxonomies discussed above are by no means the only taxonomies according to which questions may be classified, so too these four levels of analysis are by no means the only levels of analysis along which it is possible to develop classification schemes. Future work is called for to identify these taxonomies and the dimensions along which they are formed.

Looking only at taxonomies formed along the levels of linguistic analysis, however, it becomes more difficult to classify questions the higher up in the levels of linguistic analysis one goes. It is simple enough, for example, to automatically identify which wh- word is used in a question. Automatic identification of the subject of a question can also be accomplished with great success; QA systems that utilize natural language techniques have demonstrated by their success in the TREC QA track that this can be accomplished. It is more difficult, however, to identify the information need underlying a question or the patron's expectations about the form in which an answer will be presented, without some interaction with the patron. In reference work this interaction takes the form of the reference interview; in information retrieval this interaction takes the form of relevance feedback.

A new task was introduced in the Question Answering Track of the 2001 TREC conference (Voorhees, 2001): the "context task" was intended to test QA systems' ability to track context through a series of questions. As Voorhees writes, "the interpretation of a question later in the series could depend on the meaning or answer of an earlier question in the series" (The Context Task section, ¶ 2). This ability to track context was designed as a first step towards more human-like dialogue between the user and the QA system. In other words, the ability to track context would allow a QA system to pose questions to the user and collect relevance feedback, in a manner that could begin to emulate the reference interview. As it turned out, the context task as conceived for the 2001 TREC conference was "not a suitable methodology for evaluating context-sensitive processing" of questions (The Context Task section, ¶ 4). Nevertheless, the idea of context-sensitive processing of questions is a useful one, and one that has the potential, once a suitable methodology is formulated, to allow QA systems to interact with the patron to identify elements that are currently difficult to automatically identify, such as the information need underlying a question or the patron's expectations about the form in which an answer will be presented.

For automatic context-sensitive processing of questions to be possible, it is necessary for questions to be interpreted as thoroughly as possible. For questions to be interpreted thoroughly by QA systems, it is necessary that they be interpreted on as many levels of analysis as possible. For questions to be interpreted on many levels by QA systems, it is necessary to identify those levels, and the question classification schemes that exist at those levels. The levels of linguistic analysis presented above are important, if not the only, levels of analysis according to which questions can be interpreted. These four levels of linguistic analysis are the criteria according to which the five question taxonomies presented above are formed. Question answering systems that classify questions according to these levels of linguistic analysis have the potential to arrive at sophisticated and human-like interpretation of questions. A symbiosis between digital reference and QA systems will rely on such human-like interpretation of questions.

Conclusion

The NIST Roadmap document (Burger et al., 2001) lays out a vision for the development of QA systems participating in the TREC QA track, according to which QA systems will ultimately have the functionality to provide expert-level answers to expert-level questions. Some QA systems are already well on their way to this goal. The Center for Natural Language Processing at Syracuse University, for example, is currently developing a system to answer questions in the domain of aerospace engineering posed by undergraduate students in that field (Diekema et al., 2003). Cohen et al. (1998) describe several systems developed for the Defense Advanced Research Projects Agency (DARPA) High-Performance Knowledge Base (HPKB) Project for question answering and problem solving in the domain of military systems.

These systems are impressive, and do indeed fulfill the goal of providing expert-level answers to expert-level questions. These systems are, however, limited to narrow domains: materials science and military systems, for the systems just mentioned. The QA systems that participate in the

TREC QA track, on the other hand, are able to answer questions across domains, but the questions that they are able to answer are far from expert-level. The current trend in the development of QA systems appears to be to first address the extremes: simple, open-domain questions and complex, narrow-domain questions. For QA systems to gain a wider usefulness it is certainly necessary that they be able to provide expert-level answers to expert-level questions. But it is also necessary that they be able to provide acceptable answers to any questions. A combination of depth and breadth will allow QA systems to be implemented beyond the research test-beds in which they have to date been implemented.

A type of question answering service exists—has existed for well over a century—that specializes in providing answers to any question: library reference services. The next Roadmap document must lay out a vision for the development of QA systems to emulate library reference services, in their ability to provide answers of varying levels of complexity, to questions of varying levels of complexity.

In the beginning of this article a call for automation to assist reference services to scale up to handle an increasingly large number of questions was voiced. The argument was made that as question answering systems are developing, they are moving closer towards human-intermediated reference services in that they are becoming more realistic, gaining the functionality to answer questions in open domains and to provide expert-level answers to expert-level questions. Part of this development of QA systems was a call for criteria according to which taxonomies of questions may be formed. The literature from several fields that deal with questions was reviewed, and question taxonomies that exist in these bodies of literature were identified. Specifically, five question taxonomies were identified in these bodies of literature, which occupy the top four levels of linguistic analysis.

Question taxonomies are crucial to the development of QA systems because such taxonomies allow the classification of questions as one of the first steps in the interpretation of the question. Question answering systems that utilize question taxonomies may further pair specific question classes with algorithms for handling questions within those classes, thus enabling the automation of question answering on a class-by-class basis. The NIST Roadmap document (Burger et al., 2001) calls for criteria according to which taxonomies of questions may be formed. It is suggested that the levels of linguistic analysis are important criteria for the formation of question taxonomies. Question taxonomies at the various levels of linguistic analysis will prove to be crucial to the development of QA systems because such taxonomies allow the classification of questions along a number of dimensions, much as human interpretation of questions is accomplished. This question classification may not be a conscious, but it is a necessary cognitive process in the interpretation of a question and the determination of the appropriate content and form of the answer. If QA systems are to become more realistic, if they are to emulate human expert question answering, they must utilize a range of techniques utilized by human expert question answerers.

References

Austin, J.L. (1999). How to do things with words (2nd ed.). Cambridge, MA: Harvard University Press.

Bar-Hillel, Y. (1971). Out of the pragmatic wastebasket. Linguistic Inquiry, 2, 401–407.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982). ASK for information retrieval: Part I. Background and theory. Journal of Documentation, 38(2), 61–71.

Belnap, N.D., Jr., & Steel, T.B., Jr. (1976). The logic of questions and answers. New Haven, CT: Yale University Press.

Bopp, R.E., & Smith, L.C. (2001). Reference and information services: An introduction. Englewood, CO: Libraries Unlimited, Inc.

Brown, D.M. (1985). Telephone reference questions: A characterization by subject, answer format, and level of complexity. RQ, 24(3), 290–303.

Bunge, C.A. (1990). Factors related to output measures for reference services in public libraries: Data from thirty-six libraries. Public Libraries, 29, 42–47.

Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., et al. (2001). Issues, tasks and program structures to roadmap research in question & answering (q&a). National Institute of Standards and Technology. Retrieved January 13, 2005, from http://www.nlpir.nist.gov/projects/duc/ papers/qa.Roadmap-paper_v2.doc

Carnap, R. (1957). Meaning and necessity. Chicago: University of Chicago Press.

Carter, D.S., & Janes, J. (2000). Unobtrusive data analysis of digital reference questions and service at the Internet public library: An exploratory study. Library Trends, 49(2), 251–265.

Chen, J., Diekema, A.R., Taffet, M.D., McCracken, N., Ercan, N., Ozgencil, O.Y., et al. (2001, November). Question answering: CNLP at the TREC-10 question answering track. Paper presented at the Tenth Text REtrieval Conference (TREC 10), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec10/papers/cnlptrec10.pdf

Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.

Chomsky, N. (1977). On WH-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), Formal syntax (pp. 71–132). New York: Academic Press.

Clark, H.H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22(1), 1–39.

Cohen, P., Schrag, R., Jones, E., Pease, A., Lin, A., Starr, B., et al. (1998). The DARPA high-performance knowledge bases project. AI Magazine, 19(4), 25–49.

Conner, M. (1927). What a reference librarian should know. The Library Journal, 52(8), 415–418.

Crews, K.D. (1988). The accuracy of reference service: Variables for research and implementation. Library & Information Science Research, 10(3), 331–355.

Dervin, B. (1977). Useful theory of librarianship: Communication, not information. Drexel Library Quarterly, 13(3), 16–32.

Dervin, B. (1983, May). An overview of sense-making research: Concepts, methods, and results to date. Paper presented at the International Communication Association Annual Meeting, Dallas, TX.

Dervin, B., & Dewdney, P. (1986). Neutral questioning: A new approach to the reference interview. RQ, 25(4), 506–513.

Dewdney, P., & Michell, G. (1996). Oranges and peaches: Understanding communication accidents in the reference interview. Reference & User Services Quarterly, 35(4), 520–534. Retrieved January 13, 2005, from http://www.ala.org/Content/NavigationMenu/RUSA/Publications2/Reference_and_User_Services_Quarterly/Special_Features/RSP_Award-Winning_Articles/19981/1998.htm

Diekema, A., Liu, X., Chen, J., Wang, H., McCracken, N., Yilmazel, O., et al. (2000, November). Question answering: CNLP at the TREC-9 question answering track. Paper presented at the Ninth Text REtrieval Conference (TREC-9), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec9/papers/cnlptrec9.pdf

Diekema, A.R., Chen, J., McCracken, N., Ozgencil, N.E., Taffet, M.D., Yilmazel, O., et al. (2002, November). Question answering: CNLP at the TREC-2002 question answering track. Paper presented at the Tenth Text REtrieval Conference (TREC 11), Gaithersburg, MD. Retrieved

January 13, 2005, from http://trec.nist.gov/pubs/trec11/papers/ syracuse .diekema.pdf
Diekema, A., Yilmazel, O., Chen, J., Harwell, S., He, L., & Liddy, E.D. (2003). What do you mean? Finding answers to complex questions. In Proceedings of the AAAI Spring Symposium: New Directions in Question Answering, Palo Alto, CA. Retrieved January 13, 2005, from http://www.cnlp.org/publications/AAAI2003.pdf
Eichman, T.L. (1978). The complex nature of opening reference questions. RQ, 17(3), 212–222.
Evans, G. (1985). The causal theory of names. In A.P. Martinich (Ed.), The philosophy of language (4th ed., pp. 296–307). Oxford, UK: Oxford University Press.
Fogarty, E. (1995). Reference questions: Who, what, where, when, how, and why? New Jersey Libraries, 28, 19–21.
Graesser, A.C., & Black, J.B. (Eds.). (1985). The psychology of questions. Hillsdale, NJ: Erlbaum.
Graesser, A.C., Lang, K., & Horgan, D. (1988). A taxonomy for question generation. Questioning Exchange, 2(1), 3–15.
Graesser, A.C., Person, N., & Huber, J. (1992). Mechanisms that generate questions. In T.W. Lauer, E. Peacock, & A.C. Graesser (Eds.), Questions and information systems (pp. 167–187). Hillsdale, NJ: Erlbaum.
Graesser, A.C., McMahen, C.L., & Johnson, B.K. (1994). Question asking and answering. In M.A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp. 517–538). San Diego: Academic Press.
Hermjakob, U. (2001, July). Parsing and question classification for question answering. Paper presented at the Association for Computational Linguistics 39th Annual Meeting and 10th Conference of the European Chapter Proceedings of the Workshop on Open-Domain Question Answering, Seattle, WA. Retrieved January 13, 2005, from http://www.isi.edu/~ulf/papers/acl01-qa-parsing.pdf
Herner, S., Vellucci, M.J., & Leyman, E. (1972). National inventory of library statistics practices. Volume I: Data collection on the national, state, and local levels. Washington DC: U.S. Department of Health, Education, and Welfare, Office of Education.
Hovy, E., Hermjakob, U., & Lin, C.-Y. (2001, November). The use of external knowledge in factoid QA. Paper presented at the Tenth Text REtrieval Conference (TREC 10), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec10/papers/TREC10-webclopedia.pdf
Kaske, N.K., & Aluri, R. (1980). Analysis of reference statistics reported in 1977 library general information survey. New York: American Library Association.
Katz, W.A. (1997a). Introduction to reference work: Basic information services (Vol. 1). New York: McGraw-Hill.
Katz, W.A. (1997b). Introduction to reference work: Reference services and reference processes (Vol. 2). New York: McGraw-Hill.
Keyes, J.G. (1996). Using conceptual categories of questions to measure differences in retrieval performance. Paper presented at the ASIS 1996 Annual Meeting: Global Complexity: Information, Chaos and Control, Baltimore, MD. Retrieved January 13, 2005, from http://www.asis.org/annual-96/ ElectronicProceedings/keyes.html
King, G.B. (1982). Try it—You'll like it: A comprehensive management information system for reference service. The Reference Librarian, 3, 71–78.
Kripke, S. (1980). Naming and necessity. Cambridge, MA: Harvard University Press.
Kwok, K.L., Grunfeld, L., Dinstl, N., & Chan, M. (2001, November). TREC2001 question-answer, Web and cross language experiments using PIRCS (DRAFT). Paper presented at the Tenth Text REtrieval Conference (TREC 10), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec10/papers/queenst2001.pdf
Lehnert, W.G. (1978). The process of question answering: A computer simulation of cognition. Hillsdale, NJ: Erlbaum.
Library of Congress. (1992). Library of Congress Subject Headings (15th ed.). Washington, DC: Library of Congress Catalog Distribution Service.
Liddy, E.D. (1998). Enhanced text retrieval using natural language processing. Bulletin of the American Society for Information Science and Technology, 24(4), 14–16. Retrieved January 13, 2005, from http://www.asis.org/Bulletin/Apr-98/liddy.html
Lynch, M.J. (1978). Reference interviews in public libraries. The Library Quarterly, 48(2), 119–142.
McClure, C.R., Lankes, R.D., Gross, M., & Choltco-Devlin, B. (2002). Statistics, measures and quality standards for assessing digital reference library services: Guidelines and procedures. Syracuse, NY: Information Institute of Syracuse.
Mitchell, J.S., Beall, J., Matthews, W.E. Jr., & New, G.R. (1996). Dewey decimal classification and relative index (21st ed.). Albany, NY: Forest Press.
Miyake, N., & Norman, D.A. (1979). To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 18(3), 357–364.
Moldovan, D., Harabagiu, S., Pasca, M., Mihalcea, R., Goodrum, R., Gîrju, R., et al. (1999, November). LASSO: A tool for surfing the answer Net. Paper presented at the Eighth Text REtrieval Conference (TREC 8), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec8/papers/smu.pdf
Mooney, M.T. (1991). Sears list of subject headings (14th ed.). New York: H.W. Wilson Co.
Murfin, M.E., & Gugelchuk, G.M. (1987). Development and testing of a reference transaction assessment instrument. College & Research Libraries, 48(4), 314–338.
Orlikowski, W., & Yates, J. (1994). Genre repertoire: The structuring of communicative practice in organizations. Administrative Science Quarterly, 39(4), 547–574.
Phenix, K. (1983). Analysis of the 1981/82 HEGIS/LIBGIS Responses of Illinois Academic Libraries (Illinois Library Statistical Report No. 9). Springfield, IL: Illinois State Library.
Prior, M., & Prior, A. (1955). Erotetic logic. The Philosophical Review, 64(1), 43–59.
Ranganathan, S.R. (1957). Prolegomena to library classification. London: The Library Association.
Ranganathan, S.R. (1965). The colon classification (Vol. IV). New Brunswick, NJ: Rutgers University Graduate School of Library Service.
Richardson Jr., J.V. (1995). Knowledge-based systems for general reference work: Applications, problems, and progress. San Diego: Academic Press.
Roberts, G.L., & Bavelas, J.B. (1996). The communicative dictionary: A collaborative theory of meaning. In J. Stewart (Ed.), Beyond the symbol model: Reflections on the representational nature of language (pp. 135–160). Albany, NY: State University of New York Press.
Robinson, W.P., & Rackstraw, S.J. (1972a). A question of answers (Vol. I). Boston: Routledge & Kegan Paul.
Robinson, W.P., & Rackstraw, S.J. (1972b). A question of answers (Vol. II). Boston: Routledge & Kegan Paul.
Rothstein, S. (1964). The measurement and evaluation of reference service. Library Trends, 12(3), 456–472.
Ryle, G. (1949). Meaning and necessity. Philosophy, 24, 69–76.
Sachs, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50, 696–734.
Schegloff, E.A., & Sachs, H. (1973). Opening up closings. Semiotica, 8, 289–327.
Searle, J.R. (1995). Speech acts: An essay in the philosophy of language. New York: Cambridge University Press.
Seng, M. (1978). Reference service upgraded using patrons' reference questions. Special Libraries, 69(1), 21–28.
Slavens, T.P. (Ed.). (1985). Reference interviews, questions, and materials (2nd ed.). Metuchen, NJ: The Scarecrow Press, Inc.
Smith, C.A. (2002). The clinical document architecture: XML for enhanced clinical information retrieval. Unpublished doctoral dissertation, University of Pittsburgh.
Stalker, J.C., & Murfin, M.E. (1996). Why reference librarians won't disappear: A study of success in identifying answering sources for reference questions. RQ, 35(4), 489–503.
Stavri, P.Z. (1996). Medical problem attributes and information-seeking questions. Bulletin of the Medical Library Association, 84(3), 367–374.
Taylor, R.S. (1968). Question-negotiation and information seeking in libraries. College & Research Libraries, 29, 178–194.

U.S. Department of Education. (1981). LIBGIS—Public library universe (1981). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Vickery, B.C. (1966). Faceted classification schemes (Vol. V). New Brunswick, NJ: Rutgers University Graduate School of Library Service.

Voorhees, E. M. (1999, November). The TREC-8 question answering track report. Paper presented at the Eighth Text REtrieval Conference (TREC 8), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec8/ papers/qa_report.pdf

Voorhees, E.M. (2000, November). Overview of the TREC-9 question answering track. Paper presented at the Ninth Text REtrieval Conference (TREC 9), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec9/papers/qa_overview.pdf

Voorhees, E.M. (2001, November). Overview of the TREC 2001 question answering track. Paper presented at the Tenth Text REtrieval Conference (TREC 10), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec10/papers/qa10.pdf

Voorhees, E. (2002, November). Overview of the TREC 2002 question answering track. Paper presented at the Eleventh Text REtrieval Conference (TREC 11), Gaithersburg, MD. Retrieved January 13, 2005, from http://trec.nist.gov/pubs/trec11/papers/QA11.pdf

White, H.D. (1981). Measurement at the reference desk. Drexel Library Quarterly, 17(1), 3–35.

White, M.D. (1989). Different approaches to the reference interview. The Reference Librarian, 25/26, 631–646.

White, M.D. (1998). Questions in reference interviews. Journal of Documentation, 54(4), 443–465.

White, M.D. (2001). Digital reference services: Framework for analysis and evaluation. Library & Information Science Research, 23(3), 211–231.