Chapter 5. Discussion

5.1. Introduction

This study is one step in a larger research agenda exploring two issues: 1) the "lifecycle" of questions as they are handled by and pass through a digital reference service and 2) the appropriate balance between automation and human intermediation in digital reference. By extension this research agenda encompasses desk reference service, as the fundamental task of digital reference is the same as that of desk reference: to provide human-intermediated assistance to users in fulfillment of users' information needs. The second issue, the balance between automation and human intermediation, may also be extended to encompass all library service; there has always been such a balance in library service, as automated systems such as interlibrary loan and circulation systems have been in common use for many years, and systems for automated indexing (Maron, 1961; Salton, Wong, and Yang, 1975; Hlava, 2002), classification (Borko, 1964; van Rijsbergen, 1979), and abstracting (Sparck Jones, 1999) have been objects of study and experimentation for even longer. Thus, the results of this study have implications for all library services, both in physical and digital environments.

The purpose of this chapter is not to answer any questions, but to explore the issues raised above. Figure 1-1 presents a general model of digital reference; this study has focused in on one step in that model, the triage process. This chapter will explore how the findings from this study may be applied to investigation of other steps in that model as well as to processes in desk reference. This chapter will also explore how the findings from this study may inform the ongoing dialogue in the field of library science concerning the appropriate role and utility of automated systems in library work.

5.1.1. Defining the Term "Question" for Triage

In section 2.2, a number of approaches to understanding the nature and function of questions were discussed. A definition of the term "question" was advanced that drew from these different approaches. This definition is as follows:

1. *From Erotetic logic:* Questions may be decomposed into a subject and a request.
2. *From Speech Act Theory:* A question has force: in asking a question, the questioner expects that the person questioned will respond to the question.
    a. It is irrelevant whether the person questioned is face-to-face with the questioner, or remote.
3. *From Discourse Analysis:* A question does not stand alone, but is rather one step in a larger conversation.
    a. Questions are not necessarily phrased as questions.
    b. Statements phrased as questions are not necessarily questions.
4. *From the study of Information Needs and Uses:* A question is "an observable behavioral act reflecting information need" (Horne, 1983, p. 5).

This multi-part definition of the term "question" was advanced in order that questions received by digital reference services could be identified for the purposes of this study. This definition may also prove useful to future research on triage and other processes in digital reference, by illuminating the function of questions as entities to be handled in these processes. Further, any method for identifying questions would aid in the development of automated systems for handling questions.

Erotetic logic decomposes questions into a subject and a request; it is this subject that in large part determines whether the question is in scope for a specific digital reference service or appropriate for a specific answerer. The request provides a question with perlocutionary force: in asking a question, the questioner expects that the person questioned will respond to the question. An automated system that could recognize the subject and request of a question could greatly assist in the triage and question answering

processes, by suggesting to which service or individual a question should be triaged, and what form the answer should take. Discourse analysis contributes the ability to determine what speech acts are or are not questions, given ambiguous phrasing, by analyzing the putative question in light of the larger discourse between the patron and reference service or librarian. There may be a discourse in the traditional sense of semantic collaboration (Roberts and Bavelas, 1996) in real-time reference, but often there is no such collaboration in asynchronous reference. The study of information needs and uses contributes the ability to identify the information need underlying the question, even in the absence of collaboration between interlocutors. The development of automated systems for handling questions, utilizing these levels of linguistic analysis, will be discussed in more detail in this chapter.

5.2. The Utility of Question Taxonomies

This study was inspired by three papers: Jesse Shera's (1964) article predicting that automation will redefine reference service, the National Institute of Standards and Technology (NIST)'s "Roadmap" document to guide research in question answering and text summarization (Burger et al., 2001), and Marilyn Domas White's (1998) study of questioning behavior in reference interviews.

As discussed in chapter 1, Shera (1964) not only foresaw the use of computing to take some of the menial labor out and to raise the intellectual level of library reference work, but he also proposed a method for achieving that goal. Shera's agenda for action was to first, analyze the processes involved in reference service, and how questions are handled by humans in those processes, and second, create algorithms to represent these processes. This study was situated in Shera's first step: one step in the process of digital reference service was analyzed.

The NIST Roadmap document (Burger et al., 2001) provides a vision, almost a mission statement, of the direction in which Question Answering (QA) research and development should take over a five 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). The NIST Roadmap document was written to guide QA, and not digital reference research and development; however these two fields have for approximately the past five years been moving from different directions towards a similar goal: the automation of answer formulation. This study does not investigate question *answering*, but as is shown in the general process model of digital reference service, Figure 1-1, answer formulation is only one step in the larger process of question *handling* in digital reference. Any question taxonomies that are useful in one step in the process of digital reference service are likely to also be useful in other steps.

White's (1998) study of questioning behavior in reference interviews utilized one of the taxonomies discussed in the NIST Roadmap document (Burger et al., 2001) to analyze question types asked by both the patron and the reference librarian during the reference interview preceding electronic searches. To the author's knowledge, White's study was the first in the library literature to classify reference questions utilizing a taxonomy that was developed outside of the library arena. A number of schemes have been developed to classify reference questions based on the type of source from which an answer may be drawn and the form of the expected answer – classification schemes at the pragmatic level of linguistic analysis. White's study, however, utilizes a classification scheme at the discourse level of linguistic analysis, demonstrating that one useful criterion along which question taxonomies should be formed is level of linguistic analysis.

Thus the idea for this study was hatched: investigate what question taxonomies are utilized in digital reference, and how these taxonomies influence the actions performed by individuals in digital reference services who handle questions.

5.2.1. Linguistic Analysis

The idea of classifying questions is not a new one; indeed, this study would not have been possible if others had not already created a number of question taxonomies. This

study, however, identified four question taxonomies from diverse fields, and has made the original contribution of determining that these taxonomies classify questions according to the top four levels of linguistic analysis. These four taxonomies and their corresponding levels of linguistic analysis are presented in Table 5-1.

Table 5-1: Question Taxonomies and Levels of Linguistic Analysis

Level of Linguistic Analysis	Question Taxonomy
Pragmatic	The forms of expected answers to questions
Discourse	The functions of expected answers to questions
Semantic	Subjects of questions
Syntactic	Wh- words

It is significant that the existing question taxonomies from different fields fall so neatly into these levels of linguistic analysis. First, it is an indication of the utility of levels of linguistic analysis as a "technique… for analyzing and representing naturally occurring texts" (Liddy, 1998, p. 14). The taxonomy of wh- words is intuitive, derived naturally from the spelling of words used to form questions in English. The taxonomy of functions of expected answers has its basis in theories of human cognition and information needs. The taxonomy of forms of expected answers is derived from the categories utilized for measurement and evaluation of reference services. Many taxonomies of subjects exist, based both in theory and practice in a number of fields. The diversity of the origins of these taxonomies is a tribute to the logic of the above levels of linguistic analysis, and their usefulness for describing ways of analyzing questions as "naturally occurring texts."

5.2.2. Question Answering

These question taxonomies may be useful in QA research and development. In the NIST Roadmap document Burger and others (2001) state that:

> "What is difficult about answering questions is the fact that before a question can be answered, it must be first understood. One level of the interpretation process is the classification of questions. This classification should be determined by well defined principles" (p. 7).

Burger and others argue that these principles may be defined through the identification of "criteria along which question taxonomies should be formed" (p. 7). This study has identified four such criteria: the top four levels of linguistic analysis. Moreover, this study has not only identified criteria for the formation of question taxonomies, but has identified a taxonomy along each of those levels, and has demonstrated their usefulness for classifying questions. In identifying four criteria along which question taxonomies should be formed, and a taxonomy along each of those levels, this study has made a contribution to the vision for the future of QA research and development as detailed in the NIST Roadmap document.

Another future direction indicated by these question taxonomies is also a task called for in the NIST Roadmap document: to "correlate question classes with question complexity" (Burger et al., 2001, p. 8). As stated above, the NIST Roadmap document was written to guide QA research and development, and this task is important for QA so that the question processing and answer extraction mechanisms involved in answering a question can be trained appropriately for the complexity of the questions that the system is likely to handle. The classes in the taxonomies identified in this study – and the question types at the intersection of two or more classes – may provide a starting point for correlating question classes with question complexity.

5.2.3. Digital Reference

Another field that would benefit from the correlation of question classes with question complexity is digital reference – more specifically, the implementation of automation in digital reference. Before algorithms may be designed to automate the performance of any processes, it must first be determined which processes are amenable to automation.

Generally speaking, simpler tasks have historically lent themselves to being automated more readily than complex tasks: assembling inanimate objects such as cars is largely automated where surgery on living creatures is not, and there are a great many systems that automatically retrieve documents where there are few that automatically answer questions. Given the need for automation in digital reference services discussed in chapter 1, it would be a great benefit to the field of digital reference if it could be determined which types of questions are complex and therefore require human intermediation, and which are straightforward enough that they could be answered automatically. Indeed, the author is currently involved in a project to do precisely that: this project, supported by the National Science Foundation under Grant No. 02-054, seeks "to develop an ontology of questions representing at some level of detail questions needing human intermediation and those best handled by automated means" (Croft, Lankes, and Koll, 2002, p. 3). Utilizing this ontology, a system will subsequently be designed and built to handle questions appropriately: to either answer questions automatically utilizing a QA system or to triage questions to an appropriate human expert.

This study, in identifying question taxonomies at different levels of linguistic analysis, has made contributions to theory in linguistics and practice in both QA and digital reference. The theoretical contribution is the identification of the levels of linguistic analysis as "criteria along which question taxonomies should be formed" (Burger et al., 2001, p. 7), and the identification of specific question taxonomies at these levels of analysis. The practical contribution is the possibility that these taxonomies may be used as the basis for correlating question classes with question complexity in both QA and digital reference, and further as the basis for development of systems to automatically answer at least the simpler types of those questions.

5.2.4. Faceted Classification

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 is classification as a component of the interpretation of questions. While they do not state it explicitly, it is probable that Burger and others would agree that, since understanding is a complex task, more than one taxonomy would be useful to enable the full range of interpretations of questions.

As discussed in chapter 3, 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. Other classificationists have expanded Ranganathan's ideas on faceted classification schemes in two significant ways: 1) faceted classification schemes are "locally produced" (Vickery, 1966, p. 16), and so may be specifically designed to fulfill particular classificatory requirements, and, as a corollary, 2) the facets may be any categories that are logical for the entities being classified.

The four taxonomies identified in this study may be treated as a faceted classification scheme: each of the top four levels of linguistic analysis forms a facet, and the four taxonomies are the classification schemes along those facets. This faceted scheme is in a sense locally produced, as it was designed to fulfill the requirements of classifying questions, and as has already been argued, the facets are along dimensions of analysis that are logical for questions, as linguistic acts.

5.2.5. Other Possible Taxonomies

It should be pointed out, however, that just because a taxonomy was identified at each of the top four levels of linguistic analysis, it does not 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 Dewey Decimal Classification, the Library of Congress Subject Headings, and the ERIC Thesaurus, to name only a few. There may also be equally large numbers of taxonomies at the other levels of linguistic

analysis. Further, it is possible to imagine question taxonomies according to non-linguistic dimensions.

One taxonomy that has already been partially developed in the library literature is a taxonomy of genres of sources from which answers may be drawn. Several texts (Slavens, 1985; Richardson, 1995; Bopp and Smith, 2001) list such classes in this taxonomy as dictionaries, encyclopedias, almanacs, government documents, and indexes and abstracts. This taxonomy corresponds to the appropriate information sources that can be consulted in answering the question, and could be particularly useful in library education.

One useful taxonomy would be one that classifies questions according to their difficulty to answer. The problem with such a taxonomy is that it would be highly subjective, and bound to change from individual to individual and service to service. A ready reference question in a medical library might require a great deal of research in a public library, for example, or one individual might find a question difficult that another individual would find quite easy. Further, a difficult question is only difficult the first time, provided the answer is archived, searchable, and retrievable. The challenge of developing a taxonomy of the difficulty of questions would be in determining what criteria of difficulty are universal to any possible answerer and the context in which the answerer may formulate an answer.

Childers, Lopata, and Stafford (1991) propose a number of measures that may be used *a posteriori* to describe the difficulty of questions, including number of sources consulted, time spent in answering, types of sources consulted, and ease of access to sources. The problem with these variables is precisely that they are *a posteriori*, and therefore the difficulty of a question cannot be determined until after it has already been answered. While this may be a realistic way to determine question difficulty, it is not useful for processes in which questions are handled prior to answer formulation – such as the triage process.

A number of researchers and practitioners have proposed *a priori* measures of question difficulty. These include:

- The "searchability" of a question, operationalized as "the distinction between words (terms) that can or cannot be searched in an IR system as they were stated in a question." (Saracevic, 1980)
- The predictability of the source of the answer (White and Iivonen, 2002)
- The answerer's prior knowledge of the subject and of sources (Childers, Lopata, and Stafford, 1991)
- Whether the question is open or closed (White and Iivonen, 2002) (to this could be added neutral (Dervin and Dewdney, 1986))

Such *a priori* measures of question difficulty may be useful for processes in which questions are handled prior to answer formulation, such as the triage process. Such measures, however, are highly context dependent; one answerer's prior knowledge of a subject is different than another answerer's prior knowledge of the same subject, and the searchability of a question depends on the particular IR system used. Research is therefore required to determine which *a priori* measures of question difficulty are appropriate for use in what contexts. A project was discussed above with which the author is currently involved, to determine which question types require human intermediation and which may be answered automatically. Part of this project will be to determine the criteria according to which that human/automation judgment call may be made. That project will therefore make strides towards determining which *a priori* measures are appropriate for use in the context of one specific digital reference service. Two of the taxonomies identified in Phase 2 of this study correspond to factors that affect triage: the taxonomy of functions of expected answers corresponds to the question type attribute, and the taxonomy of forms of expected answers corresponds to the form in which the patron specifies that he or she would like the answer presented. By not classifying questions according to subject in Phases 2 and 3, however, this study lost some analytic power. Research question 2 asked how question type correlates with the action taken on a question in the triage process, where a question's type was defined as a question's class according to one or more taxonomies. By classifying questions according

to only three taxonomies, this study lost the ability to answer Research question 2 for question types at the semantic level of linguistic analysis. This is surely a loss for this study. As discussed above, however, each of the thirty-eight factors that affect triage may correspond to an entire taxonomy. This study only utilized three of the thirty-eight possible dimensions of analysis; the development of the remaining factors that affect triage into fully developed taxonomies would be a useful avenue for future research towards developing systems to automate the triage process.

5.3. Attributes that Affect the Triage Process

In addition to identifying dimensions along which questions may be analyzed, and taxonomies of questions along those dimensions, of course, this study has identified attributes of questions, of answers, of patrons, of services, and of answerers that affect a process of handling questions. These attributes may be useful outside of the digital reference environment, in other arenas in which questions are handled.

5.3.1. Training for Digital Reference Triagers

Training has always been an important issue in the field of librarianship (Woodard, 2001). It is no trivial thing to train to become a librarian, which is why it requires a Masters degree to become one professionally. Training to become a reference librarian is more complex still, as on top of all of the other skills necessary to become a librarian, one must master the communication skills necessary to interact with patrons on a regular basis. One of the skills that is necessary for a reference librarian to possess is when to give up – that is, when to know that the library's resources or the librarian's expertise are insufficient to fulfill the patron's information need, and thus when to refer the patron to another librarian, library, or organization. Some have argued that the referral is the mark of an unsuccessful reference transaction (Hernon and McClure, 1986). It is inevitable, however, that eventually a patron will ask a question at a reference desk that cannot be answered utilizing available resources, and that could be answered better by another library or organization.

The same is equally true in digital reference services; Silverstein and Lankes (1999) and McClennen and Memmott (2001) point out that digital reference services inevitably receive questions that are outside the scope of the service. Some of these questions may be within the scope of another digital reference service, which is of course one of the primary purposes of the triage process – to ensure that a question is answered by the organization or individual best qualified to answer it.

Knowing when to refer a question asked at a reference desk, or to triage a question received by a digital reference service, however, is not always straightforward. McClennen and Memmott (2001) state that this decision "often involves subtle questions of policy" (p. 144), and is intimately tied up with the scope of service provided by the service: e.g., the service's subject specialization, whether it provides subject-specific or general reference, the scope of its collection, and the depth of assistance it provides. It therefore seems clear that training is important for triagers in digital reference.

A limitation of this study is the very simple categorization of triage recipients that was utilized: questions received by digital reference services can be triaged to an answerer within the service, or to another digital reference service. This simple categorization was utilized because of the wide range of recipients to which different services triage questions; it was difficult to make any generalizations about these recipients or their characteristics, except to determine whether the recipient is internal or external to the digital reference service itself or the organization with which the service is affiliated.

One possible remedy to this limitation would be to conduct another study, similar to this one but in one service only, longitudinally over a span of weeks or months. The present study utilized a simple binary categorization of triage recipients, internal or external to the digital reference service itself. From a study of the triage process in one single digital reference service it would be possible to develop a more sophisticated categorization of triage recipients. Indeed, it might not even be necessary to identify these recipients through the think-aloud method; the triager in the service studied might be able to

articulate the recipients him- or herself. For example, AskERIC central (the AskERIC Clearinghouse on Information & Technology, the Clearinghouse that performs triage for the entire AskERIC service and the Clearinghouse in which a think-aloud study was conducted for this study) provided the author with a sheet of paper on which is listed the sixteen AskERIC Clearinghouses and the eight other services to which AskERIC triages questions, and the quota of questions that may be triaged to those services per day. Similarly, several academic libraries in which think-aloud studies were conducted provided the author with lists of reference and subject experts employed by the library to whom questions may be triaged.

Additionally, now that this study has identified a set of attributes that affect the triage process, it would be valuable in this proposed future study to determine which of these attributes actually do affect triage in individual digital reference services, and the correlation between question type and triage action in individual services. Such a study could test the hypothesis that there is a strong correlation between question type and triage action in a single digital reference service, stronger than the correlation across services. The author believes that this will prove to be the case. With a detailed list of triage recipients, both internal and external to the service, and a strong correlation between question type and triage action, it would be possible to articulate a detailed set of rules for performing triage. Such a detailed set of rules could be utilized to create detailed training for the triagers in an individual service.

5.3.2. Machine-Assisted Triage

The creation of such a detailed set of rules for performing triage was, of course, the second goal of this study. The correlation between question type and triage action was determined as an indication of whether such a set of rules could be drawn up; a strong correlation would indicate that such a set of rules could be drawn up, and a weak correlation would indicate not. Due to the low values for Cramér's V for the correlation between question type and triage action it does not currently appear possible to draw up such a set of rules.

One possibility for why the values for Cramér's V were not higher for the correlation between question types and question attributes, and for the correlation between question type and triage action, reported in section 4.4.4, is that this study investigated triage across several digital reference services. These digital reference services were located in five different countries, were affiliated with different types of institutions, received different volumes of questions per unit time, and had different subject specializations. Indeed, the only thing that these services had in common was that they were located in nations in which English is spoken. It is possible that such different types of services would perform triage according to different sets of criteria. Thus, this study was able to answer Research question 1, what attributes of questions affect the triage process? The answer is, thirty-eight attributes in eight categories. This study was not, however, able to fully answer Research question 2, how does question type correlate with the action taken on a question in the triage process? The partial answer to this question based on the results of this study, is: There is evidence that question type correlates with triage action moderately strongly, for all services surveyed. Given that different services perform triage differently, however, this study was unable to determine how strong the correlation may be between question type and triage action for any one single service.

As discussed in the previous section, another study is proposed, similar to this one but in one digital reference service only, longitudinally over a span of weeks or months. The results of such a study, the author believes, could be a detailed set of rules for performing triage. These rules would be specific to the service: based on the taxonomies and factors that affect triage that were identified in this study, but customized for the specific requirements of the service. Such a detailed set of rules could be utilized as the basis for designing and building a system to automate part or all of the triage process, by mapping specific triage actions to question types. When a question was received by the service, it could be automatically classified, and a triage action recommended based on that classification. A model of such a system is represented in Figure 5-1.

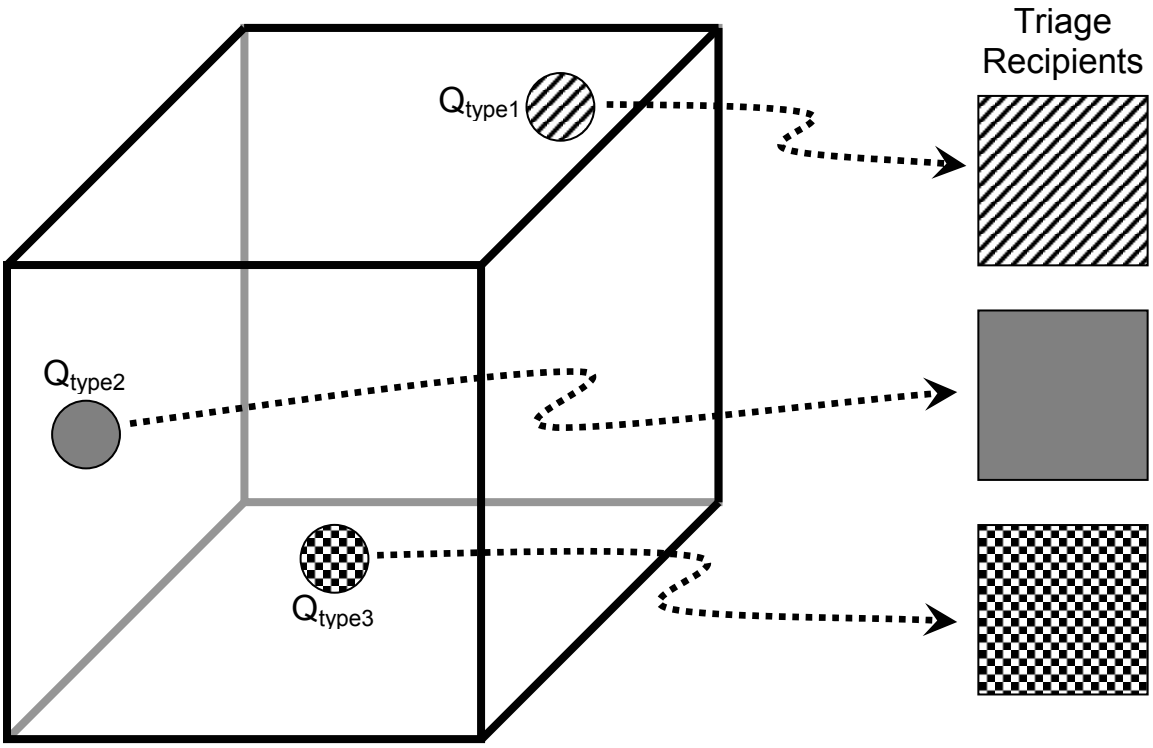

Figure 5-1: Taxonomy Space as the Basis for Machine-Assisted Triage

The system represented in Figure 5-1 is, of course, the system that the present study leads up to, but cannot develop specific rules for, let alone build. By studying multiple services and multiple types of services, the present study identified a set of factors that affect triage in multiple services and types of services. This is useful in that these thirty-eight factors are universally applicable to digital reference services. This is a limitation, however, in that these thirty-eight factors are universal, and not specific to any one service. Thus, the present study has laid the groundwork for developing "generic" algorithms for automatic triage systems. Ultimately, one possible goal for this research agenda could be to develop a system that includes all possible question types and factors that affect triage, and individual services could customize the system to fit their own particular triage requirements.

It is necessary to state at this point that the author does not believe that the triage process will be fully automated any time in the foreseeable future. Rather than "automated

triage," the author suggests the term "machine-assisted triage," echoing the terms more familiar in librarianship, "machine-assisted indexing" and "machine-assisted classification." It is reasonable to assume that some processes in digital reference may be entirely automated in the near future – indeed, the question acquisition process is nearly there even now (except for the fact that it requires a human user to submit a question), with the use of web forms for question submission and storage of questions in databases. It is easily foreseeable that the tracking process may be automated through the use of bibliomining techniques (Nicholson and Stanton, 2002). It seems less likely, however, that the triage and answer formulation processes will be fully automated any time soon, as these processes require a great deal of human judgment. Arms (2000), usually a strong advocate of automation, provides an excellent example of why such processes are difficult to automate. He asks:

> "How would we create a computer system to answer questions such as, 'Why was the space station a bad idea?' … Even with the most advanced parallel computers, nothing on the horizon approaches human judgment in understanding such subtleties" (Reference librarianship section, ¶5).

Arms' question "Why was the space station a bad idea?" is a straw man, as questions that require value judgments are a small percentage of questions received by digital reference services (see sections 4.3.1.2 and 4.4.1.2). His point, however, remains a sound one: some questions are not presently and may never be amenable to automation. Tasks that involve routines and regular actions are amenable to automation, while tasks that require human judgment are not, or at least are only partially amenable to automation. The difficult task is in determining, for tasks of this latter type, where the appropriate balance is between automation and human intermediation. To determine this, more studies are required of such tasks in different environments and contexts, to discover what patterns and regularities exist in the actions performed by human intermediaries.

5.3.3. Grammars of Actions

Pentland (1995) and Pentland and Rueter (1994) offer a useful framework for discovery of patterns and regularities in the actions performed by humans in organizational settings, proposing a grammatical model for describing such actions. Pentland proposes that a grammatical metaphor encompasses individual actions ("moves") that may be combined and recombined into sequences, just as words may be combined and recombined into sentences. Like natural language too, there are constraints on the sequences of moves that are possible within specific contexts. To discover a grammar of moves and the constraints on them in an organizational setting, Pentland suggests ethnographic techniques combined with the constant comparative method (though he does not refer to it as such) "to arrive at a set of syntactic constituents that can be generalized across settings" (p. 550).

Pentland and others (1994) apply this approach to decomposing actions performed by reference librarians in a desk reference service, though they present the frequency and variety of tasks, rather than an enumeration of the specific tasks themselves. They propose that "lexical variety" is the grammar of moves that may be performed, and "sequential variety" is the set of sequences according to which these moves may be combined. Pentland and others' results indicate that reference librarians in a desk reference service have high task variety, but low sequential variety: as they put it, "reference librarians handle a huge number of different requests in the course of their work, but the details of how they go about handling these requests do not vary much from instance to instance" (p. 30). The applicability of this grammatical model of actions to automating triage – or to automating any process in digital reference – should be clear: once a grammar of moves and the set of sequences according to which they may be combined are clearly defined, automation of these sequences becomes feasible.

5.3.4. The Equivalence of Triage and Answer Formulation

One task for which research is currently underway to determine where the appropriate balance is between automation and human intermediation is a project funded under the National Science Digital Library (NSDL) initiative, with which the author is currently involved. This project, entitled "Question Triage for Experts and Documents: Expanding the Information Retrieval Function of the NSDL," is an effort to merge the information retrieval (IR) and digital reference components of the NSDL. The goal of this project is to design and build a system to automatically determine whether a question submitted to the AskNSDL service (the AskA service for the NSDL) may be answered automatically, or needs to be answered by a human intermediary.

The premise of the Question Triage for Experts and Documents project is that triage and answer formulation are equivalent: the purpose of both processes is to match a question with the source that is best suited to provide an answer. This project treats answer formulation from the document retrieval perspective taken by IR rather than from the answer retrieval perspective taken by QA: a response to a query is a set of one or more documents, rather than a passage that contains an answer. From this perspective, the purpose of triage is to match a question with a service and ultimately an individual who will answer it; the purpose of answer formulation is to match a question with an information source or sources that contain an answer. For both of these processes, the goal is to match a question to an information container – whether that container is a document or a human. While these two processes may occur at different points in providing digital reference service, and may involve a different set of cognitive processes to perform manually, the Question Triage for Experts and Documents project is based on the premise that these two processes are equivalent in terms of the type of actions that may be taken upon a question, and therefore in terms of how they may be automated.

The problem of selecting information sources raises an issue that is not addressed in the present study: how to achieve an understanding or create a model of the content of an information source. There are several methods for creating such a model in document

retrieval: vector space modeling (Salton, Wong, and Yang, 1975), natural language processing (NLP) (Liddy, 1998), and others. Where the information source is a human intermediary, however, these methods are likely to be inappropriate. What is required instead is to create a model of the digital reference service or individual. Such models clearly exist, at least as tacit knowledge in triagers' heads (Polanyi, 1974), or Phase 1 of this study could not have been conducted. What this study does not address, however, is how those models (tacit or explicit) are created. In some cases, a model – or at least a simplified version of it – was made explicit: as mentioned above, the triager from AskERIC provided the author with a sheet of paper on which is listed the sixteen AskERIC Clearinghouses and the eight other services to which AskERIC triages questions, and several academic libraries provided the author with lists of reference and subject experts employed by the library to whom questions may be triaged. These lists were merely simplified versions of a more detailed cognitive model of the available triage recipients; they included such data as contact information, subject specialization, and quota of questions accepted per day, but did not include most of the rest of the thirty-eight factors that affected triage. In order to be made explicit, these had to be discovered inductively, in this case through think-aloud studies.

For the Question Triage for Experts and Documents project to be successful, models of individual experts are required so that the appropriate expert may be selected automatically. Such model creation is reminiscent of user modeling and automatic profiling systems. Rich (1979, 1986) describes a system called Grundy that utilizes "stereotypes, or clusters of characteristics" (1979, p. 332) as the basis for user profiles. Cohen and Perrault (1979) discuss speech acts as a basis for developing user models. Kobsa (2001) describes several more recent commercially available systems based on these techniques.

There are two differences between the models discussed in the user modeling and automatic profiling literature and those developed for the Question Triage for Experts and Documents project: 1) the Question Triage profiles will be of experts rather than of users, and 2) these profiles will be built from documents rather than from interactions with

users. These models – these expert profiles – will be built inductively, utilizing basic information, such as contact information, institutional affiliation, etc., as well as documents produced by the experts themselves, such as vita, published papers, home pages, favorite web sites, personal statements, etc. The most useful sources of information for this profile development will be determined through experimentation as part of the project. This inductive expert profile development is in a way analogous to the natural language processing of documents: Liddy (1998) states that the purpose of NLP for information retrieval is "achieving human-like language processing" (p. 14) to create representations of documents for use in matching and retrieval. While the purpose of inductive expert profile development is not language processing, it is to achieve human-like understanding, to create representations of individuals for use in matching and triage.

5.3.5. Correctness of the Triage Decision

The correlations between question types and question attributes, and between question types and triage action were calculated as part of this study. Correlations that were not calculated, however, were those between question types and correctness of triage, and between question attributes and correctness of triage. There are two ways of defining correctness of triage: from the perspective of the digital reference service, and from the perspective of the user who asked the question. From the perspective of the digital reference service, a question has been triaged correctly when the service or individual to which a question was triaged provides an answer to the question, rather than triaging the question again to another service or individual. From the perspective of the user who asked the question, a question has been triaged correctly when the user is satisfied with the answer provided.

This study did not collect data that would allow any statements to be made concerning the correctness of the triage decisions made during the think-aloud studies: the progress of questions was not tracked to determine if they were triaged additional times, nor were the users' satisfaction levels measured. The former would not be difficult to determine for services that store answered questions: the database of answered questions could be

searched to determine if the original recipient of a question answered it. For services that do not store answered questions, on the other hand, this would be impossible to determine after the fact; the question would have to have been tracked at the time that it was being triaged.

There has been little investigation of users' satisfaction with digital reference services. Bristow (1992), in possibly the first look at this issue, reported that patrons who used her library's email reference service were almost universally satisfied with it. In 1992, of course, email reference services in academic libraries were still few and far between, so it is difficult to know whether these users' satisfaction is the result of high quality of service, or simply the "gee whiz" factor of a new technology. Garnsey and Powell (2000) measured satisfaction as a user's willingness to use the service again, and according to this criteria found that 100% of users were satisfied. Both of these studies focus on users' satisfaction with the digital reference service as a whole, however, rather than specifically with the information provided.

Saxton and Richardson (2002) provide an excellent overview of studies of satisfaction with desk reference services (pp. 40-44), so a similar review will not be repeated here. The upshot, however, is that these studies almost exclusively investigate patrons' satisfaction with the *service* provided – the helpfulness and friendliness of the reference librarian, the librarian's degree of interest in the patron's information need, and other interpersonal factors (Nitecki, 1996; Cullen, 2001) – rather than with the *information* provided – its relevance and usefulness to the information need. There are therefore few studies of desk reference services that could be used as guides in constructing a study of digital reference service users' satisfaction with the information provided by the service, as an indirect way of evaluating users' satisfaction with the triage recipient for that question.

In their systems analysis model of the reference process, Saxton and Richardson (2002) describe two forms of user satisfaction with information provided: short-term and long-term (pp. 112-113). The criteria for short-term satisfaction is that the user is provided with data that is relevant and pertinent to their query, and that the system is able to

"review the results and bring about closure" to the transaction (p. 112). The criteria for long-term satisfaction is that the user finds the answer useful when it is "applied to the information problem" (p. 113) – perhaps days or even weeks later. Both of these criteria could be used in an evaluation of users' satisfaction with the information provided by a digital reference service – or, for that matter, by any reference service.

5.3.6. The Zipfian Distribution

Zipfian distributions appear in several places in the results of this study: the distribution of questions classified according to all of the taxonomies, and the distribution of attributes that affect triage, both within each category and across all categories. This type of distribution was first described by Vilfredo Pareto, and has since come to be known as the Pareto Principle, or the 80/20 Rule. This principle states that a small percentage (the 20%) of entities or actions is responsible for a large percentage (the 80%) of observed effects. Pareto observed that eighty percent of the wealth in his home nation of Italy was held by twenty percent of the population; similar observations have subsequently been made of almost every natural phenomenon conceivable. In the field of information science, the 80/20 Rule has been applied to, among other phenomena (Chen and Leimkuhler, 1986), the circulation of books in a collection (Buckland, 1975), citation behavior (Heine, 1978), author productivity (Lotka, 1926; Newby, Greenberg, and Jones, 2003), and the distribution of articles on topics through the journal literature (Bradford, 1948).

Indeed, Bates (1998) states that "most information-related phenomena have been found to fall into [this] class of statistical distributions;" (p. 1193) it would therefore be easy to state that this study has merely uncovered another set of information-related phenomena to which this distribution applies and leave it at that. There is no generally accepted explanation for why so many natural phenomena fit the Pareto distribution (Bates, 1998; Ferrer i Cancho and Solé, 2003). It is possible, however, that the reason this distribution appears in the results of this study is related to Zipf's (1949) original application of this distribution, that the frequencies of words in natural language texts follows this

distribution. Zipf suggested that this distribution represents a "principle of least effort," as the speaker or writer of a text tends to select the (relatively few) most common words most often.

Ferrer i Cancho and Solé (2003) mathematically modeled the requirements of both the speaker and listener of natural language. The speaker requires a language in which it is easy to encode meaning; the ultimate speaker-friendly language is one in which there is a single sound that means everything. The listener requires a language in which it is easy to decode meaning; the ultimate listener-friendly language is one in which every entity and concept in the universe has a unique sound associated with it. Ferrer i Cancho and Solé found that word frequencies follow a Zipfian distribution when the requirements of both the speaker and the listener are weighted equally in their model. In other words, they suggest that natural language is a perfect trade-off between the requirements of the speaker and the listener.

If, as discussed above, triage recipients may be treated as a lexical unit for a grammar of triage action, then it would be natural if triage recipients followed a Zipfian distribution. (Unfortunately, given the simple categorization of triage recipients used in this study (internal or external to a service), it is impossible to determine if recipients indeed follow such a distribution; as discussed above, one possible remedy to this limitation would be to conduct another study similar to this one but in one service only, longitudinally over a span of weeks or months.) Taking the grammatical model of trage action one step further, Phase 1 of this study can be viewed as an attempt to discover the *reasons* for which individuals select specific lexical units, given a finite vocabulary. The selection of a particular lexical unit is itself an action; thus the selection of lexical units may be treated as a "meta-language," the content of which is the lexical units of a language. If the "language" with which this study is concerned is triage actions, and the lexical units of that language are triage recipients, then the meta-language is the reasons for selecting particular triage actions: in other words, the factors that affect triage decisions. Thus, again, it is natural that factors that affect triage follow a Zipfian distribution.

This does not, however, explain why the questions classified according to all of the taxonomies also follow a Zipfian distribution. Even here, however, Ferrer i Cancho and Solé (2003) may provide an explanation. First, however, it must be posited that there are a finite number of possible types of information needs. Dervin (1983) suggests precisely this, when she proposes six simple taxonomies of "gaps" in an individual's cognitive constructs, which she states are "translated in most studies as 'information needs'" (p. 9). These information needs are, therefore, entities to which names are assigned, as a natural language assigns names to conceptual entities. Indeed, there are multiple ways to refer to a single information need: the researcher suggests that there are at least four, one according to each of the top four levels of linguistic analysis. Thus, each of the four question taxonomies identified by this study corresponds to a language with an extremely limited vocabulary. If, as Ferrer i Cancho and Solé suggest, "Zipf's law is required by symbolic systems" (p. 791), then it is inevitable that the frequency of "words" (the question classes) in these "languages" follow a Zipfian distribution.

On a more practical level, the Zipfian distribution is useful in identifying those factors affecting triage and those question classes that are the most important by dint of being the most commonly used. This has implications for the future development of systems for machine-assisted triage. Systems for machine-assisted triage would ideally be able to uniquely identify all 2,520 possible question types, and triage those questions according to all thirty-eight factors identified in this study. However, it is almost always the case in software development that compromises need to be made, on account of budgets, time, or other factors. Thus, it is important in software development that the most common user practices must be supported, while the least common practices may be dropped from software specifications. In other words, the most important question types and triage factors must be supported by any future system for machine-assisted triage. As discussed above, there is a wide range of recipients to which different services triage questions, and it may also be that different services receive different types of questions. Thus, another study, similar to this one but in one service only, as proposed above, could provide the data necessary to develop systems for machine-assisted triage, customized to services' specific needs and uses.

5.4. The Uniqueness of the Triage Process

Lankes (1998a) refers to triage between digital reference services as "meta-triage," thus making a clear separation between intra- and inter-service triage. This distinction was collapsed for this study, and both intra- and inter-service triage was referred to as triage. The reason for this use of terminology is that both are the process of routing and assigning questions, regardless of the recipient. It is, however, useful to differentiate between intra- and inter-service triage, and this section will explore that difference.

In addition to intra- and inter-service triage, there is a third variation, which will be called multi-level triage. Actually, multi-level triage is not a variation of triage so much as it is a combination of intra- and inter-service triage; indeed, multi-level triage is an almost inevitable result of inter-service triage. Multi-level triage is triage in several stages; for example, one service may triage a question to another service, and the second service may triage the question to yet a third service, and that third service may triage the question to an individual expert. Thus, the triage process occurs in stages, rather than being a single act.

The discussion of referrals, above, draws a parallel between referrals at desk reference services and the triage process in digital reference. On the one hand, the purposes of the two processes are equivalent, in that they both attempt to insure that a question is answered by the organization or individual best qualified to answer it. On the other hand, the responsibility for completing the referral and triage tasks are quite different. In desk reference, if a patron is referred from one service to another, the burden is often on the patron to contact that other service. In digital reference, if a referral is made, it is not the patron who is sent from one service to another, but the patron's question. Thus, the burden is on the service that received the question to perform triage, and on both parties in the triage process to work out the details of the exchange.

It is in the details of the exchange of a question that triage truly differs from a referral. In the electronic environment, forwarding a question from one digital reference service to another is technically simple; every email application has a Forward button and an addressbook. What is more complex is developing policies and standards to govern the making of referrals. Lankes (1999) describes the Question Interchange Profile (QuIP), a proposed metadata standard for passing information between digital reference services along with a question: information about the question, about the patron, about the forwarding service, etc. Formalization of such a standard is currently being carried out by the National Information Standards Organization (NISO) Networked Reference Services Standards Committee AZ (www.niso.org/committees/committee_az.html). The possibility of the automatic exchange of not only a question, but also of metadata about the question, sets the triage process apart from referrals in desk reference services. Lankes (2001), states that there are two fundamental differences between traditional desk reference and digital reference: 1) the disconnect of user and physical space – that the library is remote from users, and 2) the fact that the process of providing digital reference service creates artifacts – documents that may be collected, stored in knowledge bases, and searched. To these two fundamental differences between desk and digital reference, the author believes that a third should be added: 3) the machine-assisted triage process, and the possibility of the automated exchange of rich metadata about the question between services.

It is not the possibility of a rich exchange of metadata about a question that is unique to digital reference, but the *automation* of that exchange. In traditional desk reference, when making a referral a librarian may telephone the referred-to service to inform them that a patron is being referred to them and to provide some information about the patron and the question. This information provided over the telephone may be considered to be metadata about a question, but the exchange of this information is a manual process. Further, this telephone call often is not made, and therefore, as mentioned above, the burden is on the patron to contact the other service. Thus, by automating this process of question and metadata exchange, not only will the reference service be able to scale up to handle an increasingly large number of questions, but this scaling will be modeled on the ideal

practice of reference. The utilization of automation will enable services to automate processes as they *should be* performed, rather than as they *are* performed in practice.

5.5. Triage is Dead, Long Live Triage

There is some irony in the possibility that the triage process as it is operationalized in this study – as a manual process – may be dying out, at least for intra-service triage in more high-tech digital reference services. Many services utilize web-based question management software: both commercially available software applications such as LSSI, 24/7 Reference, and LivePerson, and applications that have been developed by various digital reference services such as QABuilder, developed under the aegis of the Virtual Reference Desk Project, QRC, developed by the Internet Public Library (IPL) (Lagace and McClennen, 1998), and QuestionPoint, developed by the Library of Congress (Kresh, 2000). These applications store questions in what Lankes (1998a) refers to as a "triage area" (p. 106): a virtual space where questions are listed, and experts can select the questions that they wish to answer. Figure 5-2 is a screenshot of the triage area for the Learning Center, a digital reference service run by the Virtual Reference Desk (VRD). Every question must be classified according to a list of subject categories, including those shown: Social Issues, Arts, Philosophy, etc. The expert can click on a link to the question itself ("media influence on teens," "Shirley Temple," etc.), claim the question, and formulate and submit an answer.

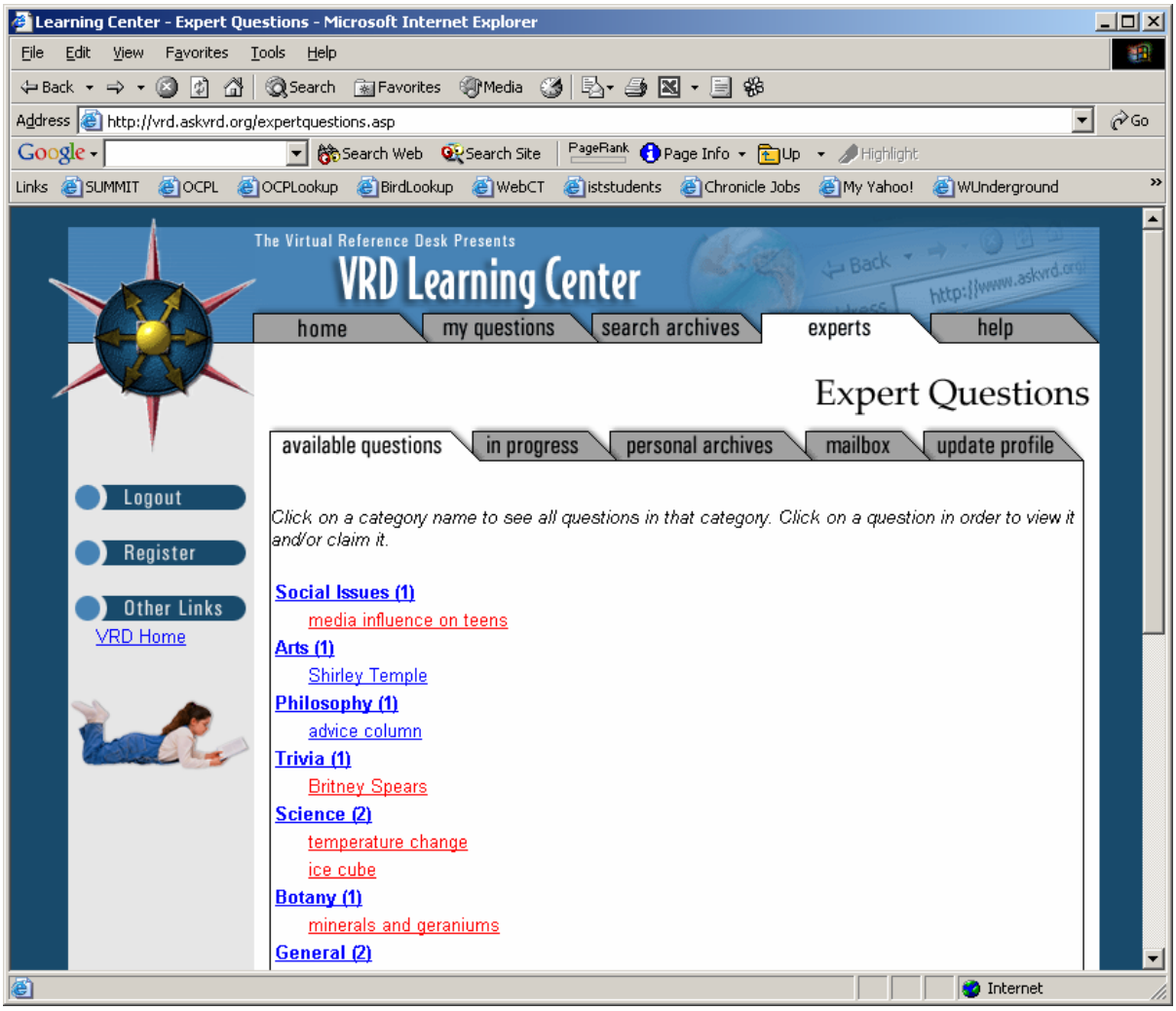

Figure 5-2: Screenshot of the Triage Area for a Digital Reference Service Using the QABuilder Software

While QABuilder and other applications all have different interfaces, they all seem to be converging on a similar set of functionalities (Hirko, 2002). Specifically, the functionality of the triage area is all quite similar: upon submission the patron may classify the question in one or more of a list of subject categories, questions may be re-classified by experts and/or the site administrator, and questions may be claimed by one and only one expert.

Pomerantz and others (forthcoming) state that at the time they were writing, the QABuilder was used by six services, including the VRD itself, QRC was used by five services, including the IPL, and one of the commercially available software applications, LiveAssistance, is used by "about 25" services (Sarkar, 2002). Of course this represents only a fraction of all of the digital reference services in the world, but the number of services utilizing some form of question management software is increasing rapidly (Hirko, 2002).

Pomerantz and others (forthcoming) report that 83% of services surveyed maintain a question submission web form. This is not in itself indicative of the use of a question management software application, since, while such applications do make use of webforms for question submission, such a webform may just as easily utilize a CGI back-end to send an email message to the digital reference service's inbox. On the other hand, some figures that Pomerantz and others did not report are that 39% of services surveyed store previously answered questions in a knowledge base, and that 6% automatically search a knowledge base of previously answered questions when a question is received. Again, this is not in itself indicative of the use of a question management software application, as Bry (2000) describes the MAD Scientist Network's archive of previously answered questions and use of CGI scripts to search that archive, without the use of such an application. Nevertheless, there is clearly a demand in the digital reference community for the functionality offered by question management software applications.

As digital reference services increasingly adopt question management software applications, they are likely to increasingly adopt the self-selection model of triage supported by these applications. Thus triage as an entirely manual process may be dying out. As can be seen in Figure 5-2, however, in QABuilder an expert sees questions listed under a set of subject categories. In setting up his or her profile, an expert can set the subject categories in which he or she wishes to see questions. Additionally, an expert can set the grade level of the patrons from whom he or she sees questions. These are only two factors according to which questions could be classified, and according to which an expert could select subsets of questions to see. It should also be pointed out that these two

factors are attributes that were identified by this study. Thus there are possibly as many as thirty-eight factors according to which expert profiles may classify and filter questions. It may be that a hybrid triage/self-selection model is evolving, where filtering of questions according to a set of criteria is performed automatically, so that the expert only sees the questions that he or she wants to see, that match the criteria set in his or her profile, and them the expert may self-select questions from within that subset of all questions received by the service.

5.6. The Question Interchange Profile (QuIP)

While triage may be dying out, or at least evolving within services, it is alive and well between services. Lankes (1999) describes the Question Interchange Profile (QuIP), a set of metadata elements that may be passed between services along with a question, to convey any information deemed necessary to provide context for the question. QuIP elements include, in addition to the question itself, the instigator of a communication interchange, the patron's affiliation, the patron's role, language of the question, subject of the question, subject specialty of the expert, etc. QuIP is currently undergoing review by Standards Committee AZ of the National Information Standards Organization (www.niso.org/committees/committee_az.html).

Pomerantz, Nicholson, and Lankes (2003) argue that QuIP, or any future standard based on it, should contain an element for question type. This study provides a faceted scheme for classifying question types, according to four levels of linguistic analysis. Providing data about a question's type could be useful when passing a question from one service to another, particularly if question type correlates strongly with triage action within the recipient service.

Another useful element that could be included in QuIP or any future standard based on it is triage factors: which factors out of the set of thirty-eight discovered by this study were relevant to the triager's decision to pass the question to another service. When a user submits a question to a digital reference service, or asks a question at a reference desk,

the reason why the user chose that particular service, or even more fundamentally, why the user asked the question of a reference service at all (as opposed to, for example, asking a friend or colleague) is generally not asked. Such a question may be seen to be irrelevant; the patron is standing in front of a reference desk, or a question is sitting in the email inbox or triage area, and the expert's job is to provide an answer. On the other hand, when a question is received from another digital reference service, rather than from a patron, it could be useful to know why that other service did not or could not answer it. The point of QuIP is to convey any information between services deemed necessary to provide context for the question; a service's reason for passing a question off to another service certainly could be a crucial piece of contextual information.

5.7. Real-time Reference

A technology that is currently being rapidly adopted by digital reference services is so-called "real-time" reference, also known as chat or live reference (Coffman, 2001; Francoeur, 2001; Janes, 2002; Marsteller and Neuhaus, 2003). This form of reference utilizes chat or instant messaging functionality to conduct a synchronous reference transaction, along with web page-pushing, graphical co-browsing, queue management, and a number of other functions (Hirko, 2002). The synchronous nature of this form of reference sets it distinctly apart from the asynchronous email- and web-based reference with which this study is concerned. There are a number of commercially available question management software applications that offer this functionality (Hirko, 2002), and at least one developed by a digital reference service – QABuilder – that is currently working towards implementing that functionality.

This study, and the general process model of digital reference service upon which this study is based (presented in Figure 1-1), does not take real-time reference into account. At the time that the general process model was developed, these technologies had not yet widely impacted the practice of digital reference. The focus of this study was the triage process, and there has been no literature on real-time reference indicating that any services are performing triage on incoming real-time requests, though Francoeur (2001)

states that more sophisticated question management applications allow an administrator to "transfer" incoming requests to available librarians (p. 193). In principle, there is no reason why triage *could not* be performed on incoming real-time requests, though given the synchronous nature of the transaction, a human intermediary would have to be standing by in the recipient service to receive the triaged request, just as Patterson (2001) and Kibbee, Ward, and Ma (2002) state that librarians were on duty to receive incoming requests in the services that they report on.

Because this study was concerned with only those services that receive questions *via* asynchronous media, the findings are unlikely to inform current real-time reference services, though should real-time services develop triage functionality in the future, this study's findings could be extended to triage in that environment. An interesting future application of the results of this study would be to conduct think-aloud studies with triagers or the equivalent in a digital reference service offering real-time reference (Francoeur's (2001) administrators, who transfer incoming requests to available librarians), to discover the factors that affect triage of real-time requests. Such a study could test the hypothesis that the same set of thirty-eight factors affects triage of real-time requests as affect triage of asynchronous email-based reference requests.

Another interesting future application of the results of this study would be to continue White's (1998) work in classifying the questions asked by both patrons and reference specialists during the reference transaction. White found that question types different significantly for questions asked by patrons and reference specialists. White did not report, however, on whether specific question types led to other specific question types: that is, if there is a correlation between a question's type and the question's type that follows it. This is, of course, the sort of study that can only be conducted for desk or real-time reference services, since, as Carter and Janes (2000) report, if an expert replies to a user's question with a request for clarification, 30% of users do not ever reply with that clarification. (Judging by the author's conversations with the triagers during Phase 1 of this study, Carter and Janes' finding is a remarkably small percentage.) If such a correlation exists, then this would be evidence that there are regularities in the reference

transaction, patterns that exist across conversations. Such patterns could be used to create standard "scripts" to automate the reference transaction, in much the same way that reference expert systems have attempted to guide the user through structured interviews (Richardson, 1995), but with more sophisticated criteria for following one question up with another.

5.8. Chapter Summary

This study posed two Research questions:

RQ1. What attributes of questions affect the triage process?

RQ2. How does question type correlate with the action taken on a question in the triage process?

The answer to Research question 1 is:

A1. Thirty-eight attributes were discovered that affect triage decisions. These thirty-eight attributes fall into eight categories. These attributes and categories are presented in Table 5-2:

Table 5-2: Categories and Attributes That Affect Triagers' Decisions

Category of Attributes	Attribute
Attributes of the question	Subject of the question
	Difficulty of the question
	Generality or specificity of the question
	Question type, according to one or more of the classes from the taxonomies of wh-words or functions of expected answers
	Interestingness of the question, in the opinion of the triager

	Language that the question is written in
	The fact that one patron submitted multiple questions
	The fact that the question has a prior history with the service: it is a follow-up question or has been forwarded to another service and the other service forwarded it back
Attributes of the answer	Form in which the patron specifies that he or she would like the answer presented: e.g., short or long answer, or according to one or more of the classes from the taxonomy of forms of expected answers
Attributes of the patron	Organizational affiliation
	Role, job description, or other capacity in which the question is being asked
	Age
	School level
	Country of residence
	State of residence, or other subdivision within the country of residence: e.g., Province in Canada
Attributes of the patron's current information need	Planned use of the answer
	Information sources already searched
	Date an answer is needed by
	Number of questions received on a given

Attributes of the triaging service (the service or triager that received the question)	day
	Number of questions on the same subject that is expected to be received by the service
	Habits that the triager has developed over time in triaging questions
Attributes of the receiving service (the service to which a question is triaged)	Scope of the collection
	Scope of the service: does the service provide general or subject-specific reference?
	Depth of assistance provided: what is provided as or with an answer: citations only, answers, bibliographic instruction, etc.
	Response rate: how many of the questions which the service receives get answered
Attributes of the receiving service or the answerer	Quota for number of questions that can be accepted per day
	Subject expertise: areas in which the service or answerer has expertise in answering questions
	Audience served: patrons' affiliation, age, school level, area of residence
	Availability of appropriate information sources that can be consulted in answering the question
	Past performance in quality of answers

	provided
	Turnaround time for providing an answer
	Country of residence
	State of residence, or other subdivision within the country of residence: e.g., Province in Canada
Attributes of the answerer (the individual expert to whom a question is triaged)	Subject interest beyond subject expertise
	Reference experience
	Customer service expertise
	Expected answer formulation: how the triager anticipates that the answerer will formulate an answer

The answer to Research question 2 is:

A2. There are moderately strong correlations between triage action and the taxonomy that is used to classify the question. There are some very strong correlations between triage action and specific question types (that is, specific cells in the taxonomy space). These very strong correlations are presented in Table 5-3:

Table 5-3: Very Strong Correlations between Intersections of Question Types and Triage Action

Cell in the taxonomy space	Triage action	Example question
----- Coverage Factual	Internal	"Is there a chart that can give you distance between one star and another"

		"How were scalp diseases treated in the late 19th Century?"
----- Quantification Factual	Internal	"Do babies see in color right away, if not when do they?" "Would you be able to tell me how much money from the Gross National Product goes into the Military?"
----- Coverage Citation list	Internal	"Is there any system or body of experience in teaching bridge to school children?" "Do have research on including marshal arts, wresting and/or boxing modules in alternative school settings?"
What-description Request -----	Internal	"name the Major River in Florida." "Can you tell me information about bronze items made in the Shang Dynasty?"
Where Request Factual	Internal	"Where was the first McDonald's restaurant franchise location in Seattle?" "I am seeking journals/articles relating to student test preparedness behaviors. Do you have any suggestions as to where I might find this information?"
How Explanation Factual	Internal	"How does quinine work?" "I know that Spanish Missions in California tended to follow the coastline. Is there a similar pattern of coastal missions in Mexico?"
-----	External	"Where can I find info on the Internet relating

Coverage Directional		to the European Economy during 1400 to 1700 AD?" "What is the website for the SCANS Report?"
What-selection Coverage Citation list	External	"Need documentation on schools associated with zoos." "Need lesson plans to assist with classes related to memorials for Sept. 11th."

The utility of question taxonomies and the set of factors that affect triage were explored, in both the fields of QA research and development, and digital reference. It was demonstrated that both the set of question taxonomies and the set of factors that affect triage that were discovered by this study have utility both for the human intermediaries in digital reference services, and as the basis for systems that will partially automate the triage process and other processes in digital reference. The term "machine-assisted" triage was proposed rather than automated triage. Other steps in the digital reference process were discussed, and it was argued that both the set of question taxonomies and the set of factors that affect triage may have utility for these purposes.

In conclusion, this study produced two useful intellectual tools that are original contributions to the fields of QA and digital reference. First, by expanding on the fifteen factors that Pomerantz, Nicholson, and Lankes (2003) discovered that influence the triage process, this study discovered eight attributes of questions, and a total of thirty-eight criteria in eight categories, that affect triage. Second, while the four question taxonomies existed prior to this study, this study determined that they correspond to the top four levels of linguistic analysis; further, these four levels of linguistic analysis are shown to fulfill Burger and others' call for "criteria along which question taxonomies should be formed" (p. 7). Finally, some very strong correlations were discovered between specific question types and the triage action taken on those questions, and it is suggested that

these very strong correlations may be utilized as the basis for designing systems to automate the triage process.