1984

Case Grammar And Functional Relations In Aboutness Recognition And Relevance Decision-making In The Bibliographic Retrieval Environment

Dee Ann Lewis

Follow this and additional works at: https://ir.lib.uwo.ca/digitizedtheses

Recommended Citation
https://ir.lib.uwo.ca/digitizedtheses/1351

This Dissertation is brought to you for free and open access by the Digitized Special Collections at Scholarship@Western. It has been accepted for inclusion in Digitized Theses by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED
CASE GRAMMAR AND FUNCTIONAL RELATIONS IN ABOUTNESS RECOGNITION AND RELEVANCE DECISION-MAKING IN THE BIBLIOGRAPHIC RETRIEVAL ENVIRONMENT

by

DeeAnn Emmel Lewis

School of
Library and Information Science

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
July, 1984

ABSTRACT

One of the most perplexing problems in Information Science has been definition of central concepts such as 'relevance' and 'information' at a theoretical level which incorporates and accounts for all known attributes of the concepts and the principles which underlie their particular application in this field. The objective of this thesis was to determine whether one definition of relevance, 'aboutness', can be based, at least in part, on textual characteristics of queries and abstracts. To this end, this study was conducted to determine to what extent a set of functional relations based on Fillmore's case grammar theory could be used to explain the correspondence between patterns of language behavior in aboutness recognition and language patterns in the texts of queries and abstracts. The test environment was a real-life bibliographic retrieval system. Thirty subject specialists (advanced graduate students and university faculty from the Sciences, Social Sciences and Humanities) submitted queries and performed relevance assessments on titles and abstracts retrieved from the DIALOG data system using 'aboutness' or 'topicality' as the operational definition for their judgements. The analysis of the functional relations between keyterms in the queries and in the titles and abstracts were compared and the results were then
compared to the subject specialists' decisions. The agreement between the two sets of decisions was 97% across all queries. A second test of the ability of secondary judges (with and without subject backgrounds in the subject of the query they were assigned) to predict the decision of the subject specialists showed only 65% agreement. However, the internal consistency of these judges' decisions based on their interpretation of the intended query showed 84% agreement with the functional relations assignments based on their own decisions. The functional relations assignments and interpretations of match and non-match of queries and abstracts were corroborated by a second language analyst. The conclusion reached on the basis of the findings is that abstracts judged to be about the topic(s) named in a query did contain the desired keyterms in functional relations which matched the functional relations between the keyterms in the query. Abstracts judged to be not about the topic(s) named in query did contain the same or equivalent keyterms, but not in functional relations which matched the functional relations in the query. The high level of agreement between the functional relations decisions and the judge's decisions also demonstrated the consistency of language behavior in relevance assessments where the definition on which the decisions are made is aboutness or match in topic.
These conclusions have implications for indexing systems, query negotiation, search strategy formation, retrieval system research and the development of interface mechanisms for on-line retrieval systems.
ACKNOWLEDGEMENTS

My first inclination here is to thank the whole world—all of the people in my own corner individually, and the rest of the world in general for continuing to limp along until I completed this seemingly endless task. Realizing that would seem pretty silly and very boring to anyone reading this, I have opted instead for a brief expression of my gratitude in more general terms. I first wish to thank everyone at S.L.I.S. — faculty, fellow doctoral students, and staff alike — whose friendship and support made this endeavor not only possible but also enjoyable. (I may be completely mad, but I actually did enjoy doing this degree.) I also wish to thank the nearly one hundred people who participated in this study, especially the searchers (Cynthia Singleton, Janet Rogers, David Nelson, and Peter Buehler) whose enthusiasm and professionalism ensured the success of the data collection. And, finally, a very special thank you to my advisors Dr. Jean Tague (chief advisor) and Dr. Gillian Michell for their support, both academic and personal, and for their friendship. I learned a great deal about research, writing, and discipline under your guidance.

On my part, I promise that I will try to take the best of what I learned from this experience at S.L.I.S. with me. The principles under which the school is run are probably the most important ideas I have encountered in my educational career. Thank you.
Saracevic's historical analysis of the treatment of relevance in information science is thorough, and his summations insightful. However, his conclusion that the assumptions underlying relevance cannot be empirically tested presumes, possibly erroneously, that all of the assumptions underlying the concept are known. The linguistic analysis being proposed here presumes that descriptive analyses of the factors which are thought to affect relevance judgements have not considered adequately deeper questions of language behavior which may lend themselves to empirical testing and which may help us to understand the process by which relevance judges make relevance decisions.
INTRODUCTION

One of the most perplexing problems in information science has been definition of central concepts such as 'relevance' and 'information' at a theoretical level which incorporates and accounts for all known attributes of the concepts and the principles which underlie their particular application in this field. The problem addressed in this thesis is the difficulty in stating an objective definition of 'relevance' which can be applied reliably in tests using relevance as a basis for measurement. The objective of the thesis is to determine whether one such measure - 'aboutness' - can be based, at least in part, on information provided by the textual characteristics of queries and abstracts. In order to accomplish this objective, it is necessary to determine whether it is possible to obtain empirical evidence of correspondence between patterns of human language behavior in aboutness recognition and language patterns in the texts of queries and abstracts. The purpose of the research undertaken here was to determine to what extent one linguistic theory, case grammar theory, can be used to explain how relevance judges discriminate between relevant and nonrelevant document representatives.
The research question explored here is:

if

a) a set of abstracts contain the same, or equivalent, keywords, and
b) subject specialists judge some of the abstracts to be about the query and others to be not about the query,

then

c) there will be a relationship between the subject specialists' decisions and the functional relations between the keywords in the query and in the abstracts.

That is, abstracts judged to be about the topic named in a query will contain the desired keywords in functional relationships which match the functional relationships between the keywords in the query. Abstracts judged to be not about the topic named in the query will contain the same, or equivalent keywords, but not in the functional relationships which match the functional relationships described in the query. Such an explanation could prove to be an important step in the development of a theory of relevance, and consequently, a stronger theoretical foundation for many aspects of information science research activities which have historically depended upon 'relevance' judgements as a basis of measurement. In addition, there
has long been interest in the use of some method of indicating the functional relations between words in indexing and in search strategy formation.

The method of indicating functional relations which is explored in this thesis is based on case grammar, a linguistic theory which uses deep structure syntagmatic relations. The most attractive feature of deep structure syntagmatic relations is that they get around the problem of variation in surface syntactic structures. That is, the relations between noun phrases and the verbal element in a sentence or proposition remain constant no matter how they are syntactically expressed - e.g., as in active and passive versions of the 'same' sentence. It is hoped that differences in functional relations will provide at least a partial explanation of contributing factors in language behavior in relevance decision-making, and further, that the functional relations by which users discriminate between texts which contain the same words, but do not carry the same meaning, can be identified. Finally, it is also hoped that analysis of document representatives whose non-relevance cannot be explained by functional relations differences will indicate other generalizations - e.g., semantic or pragmatic patterns - which can be explored further.

Determination that a major meaning-based factor is an important underlying mechanism in relevance decision
making could contribute significantly to the development of:

1. theories of relevance, query negotiation, search strategy formation, and indexing;
2. searching systems which more effectively reflect and exploit natural language behavior; and
3. more reliable definitions of various types of relevance for use in evaluation of the performance of bibliographic retrieval systems.
0.1. Relevance in Information Science.

In his excellent state of the art review of the development of relevance as a concept in information science, Saracevic states, "Relevance has been selected [as the measure in information retrieval system testing] as a result of the recognition that the prime objective of information retrieval systems is to provide relevant information to user queries (1975, p. 327)." Historically, Saracevic states, operational definitions of the relation between relevance as judged by a user and relevance as judged by an information system - i.e., document surrogates retrieved by the system in response to a query - have been very important for setting boundaries on experiments. However, these definitions have, for the most part, consisted of descriptive analyses of the factors which were thought to affect relevance judgements and discussions of the most appropriate or valid criteria under which relevance decisions would be made. The concept 'relevance' itself has not, however, really been dealt with adequately. Saracevic concludes that relevance definitions have really been a form of hypothesis for each experiment and neither the hypothesis nor the assumptions underlying it have been, nor can be, fully tested since the concept does not lend itself to empirical or scientific testing.
Saracevic's historical analysis of the treatment of relevance in information science is thorough, and his summations insightful. However, his conclusion that the assumptions underlying relevance cannot be empirically tested presumes, possibly erroneously, that all of the assumptions underlying the concept are known. The linguistic analysis being proposed here presumes that descriptive analyses of the factors which are thought to affect relevance judgements have not considered adequately deeper questions of language behavior which may lend themselves to empirical testing and which may help us to understand the process by which relevance judges make relevance decisions.
0.2. The Operational Definition of Relevance.

The operational definition of 'relevance' in information science involves two decisions. First, the relations which interact in a bibliographic retrieval situation and which will affect the relevance decision-making process must be explicated. The most important of these elements are the relevance judge, the information system, the query, and the document representatives retrieved from the database system in response to the search strategy representing the query. Second, the definition of 'relevance' itself must be determined. The definition of relevance is an explication of the basis upon which document representatives will be judged.

The concept 'relevance' has been widely discussed and each retrieval experiment or study involving relevance judgements has had to describe how the term was defined. Several distinct definitions of relevance have been postulated. All of these definitions are descriptions of the nature of the relationships between retrieved documents or document representatives, the query, and the relevance judge. The three most commonly used definitions of relevance are 'utility', 'pertinence', and 'aboutness' or 'topicality'. The choice of a definition of 'relevance' for a study depends upon the purpose for which the study is being conducted. In tests of efficiency or effectiveness of a system or features of a system, the appropriate definition is defined by the research question.
0.2.1. The Operational Definitions of Relevance Elements.

Two considerations have traditionally been involved in the construction of the operational definitions of each of the elements which affect the general operational definition of relevance for each experiment. The first is explication of the characteristics of those elements in the experiment which will affect relevance - e.g., real queries versus artificial queries, real-life users versus specially recruited users, etc. The second is explication of the effect these characteristics have on the relations between the elements, such as the relations between the user and the query, and the effect of the nature of that relation on relevance decision-making. The operational definition of relevance upon which the final performance measures for the experiment (traditionally recall and precision which are also relations) are based is, thus, in part determined by the operational definitions of the elements and their relations.

The operational definition of retrieval system elements for each experiment is determined by the opinion of the researcher about the validity and importance of various characteristics of elements and the relations between them, and by constraints imposed by the experimental situation itself. (Such environmental constraints may arise from the characteristics of the user population,
the data base, the data system's interface mechanisms, time, etc.). The result has been not only that it is difficult, if not impossible, to generalize conclusions drawn in one experiment to other situations, but also that the validity of the results of any one experiment can be called into question simply by altering the characteristics of any important variable and obtaining significantly different results.

Because the relevance judges' decisions are probably the most crucial element in both real-life and experimental retrieval situations, it is around this crucial element that much of the most serious debate about relevance has centered. The first characteristic of the relevance judges which should be considered is their level of subject knowledge and expertise. Saracevic states that "...subject knowledge seems to be the most important factor affecting the relevance judgements as far as human characteristics are concerned (1975, p. 341)."

While subject expertise is an important factor in relevance decisions, several other problems involving this element are also of interest here. The relevance judge is involved in important relations with two other elements in the bibliographic retrieval system as well - i.e., the query and the retrieved document representatives. These relations, more than any others, affect the nature of the relevance decisions which will be made.
The relation between the relevance judge and the query has a potentially determining effect on relevance judgements regardless of any instructions given for purposes of the experiment and beyond degree of subject knowledge. A researcher is faced with choosing among use of real queries in a real-life situation, use of artificial queries and neutral relevance judges in a completely artificial situation, use of real queries but neutral relevance judges - i.e., judges who did not submit the queries, use of real queries and judgements by users who submitted the queries but under controlled conditions, etc. Each variation can alter the response of the relevance judge to the set of retrieved document representatives. For example, according to a study by Berhydt (1967), neutral subject specialists performed no better than searchers in anticipating a user's decisions about the relevance or nonrelevance of retrieved items (about 65% agreement). Therefore, without an adequate theory of relevance or understanding of the relevance decision-making process, preference for the results of one experiment over another rests almost entirely on personal opinion.

The relevance decision situation is further complicated by the range of possible criteria which relevance judges can be asked to apply in their decision-making. The operational definition of relevance is probably the single most contentious issue in bibliographic retrieval experiments at present, which is a direct result of the lack of an
adequate theory of relevance which recognizes and accommodates all varieties of relevance judgement criteria.

As noted above, the three most commonly used definitions of relevance are 'utility', 'pertinence', and 'aboutness'. The narrowest of these definitions is 'utility'. As defined by Cooper (1979), utility is a measure based on the usefulness of items retrieved and on cost-effectiveness in terms of time, effort, cost and satisfaction. Utility is, therefore, a completely subjective measure.

'Pertinence' refers to the relations between the user's 'information need' and the retrieved records - e.g., each record is judged to be 'useful' or 'not useful', or records can be ranked from 'most useful' to 'least useful'. This definition is also very subjective and, like utility, includes more than whether or not a record is about or not about the topic described in the query. It includes comprehensibility, credibility, degree of importance, currency, style, journal, author preference, etc.

The third measure, 'aboutness', refers to the relation between the query and the retrieved document representatives and is usually defined in terms of a match in topic between a document representative and the query - i.e., the document is 'about' or 'not about' the topic described in the query. Relevance decisions based on aboutness are in a sense independent of 'utility' or 'pertinence' since
A document representative may be 'about' the topic described in the query, but not useful to the user for some reason.

The strongest position on relevance judgements currently is that the only truly valid relevance decisions are based on real-life user satisfaction and the usefulness of particular documents for answering the user's information need - i.e., 'pertinence' or 'utility'. Paradoxically, acceptance of this position and the assumptions underlying it may be partially responsible for the continued lack of a broader theoretical base for relevance. Obviously, user satisfaction is the final criterion by which any information service is judged, and a better understanding of the factors involved in user relevance decision-making is essential to the central purpose of the system - i.e., providing information which will satisfy the user's information need. But the development of a comprehensive theory of relevance must include the entire process of relevance decision-making, the beginning and end of which remain to be determined.

Another approach to relevance and to retrieval performance evaluation is to derive a measure of the effectiveness of a system in supplying all and only those document representatives which are 'about' the topic named in the query statement and which are, therefore, potentially relevant for any user presenting that query.
to the system. This definition of relevance - i.e., a match in topic, or 'aboutness', between a query and a document or a document representative - moves toward the concept of the 'concern set' discussed by Wilson. By Wilson's definition, "...a concern set of statements is a set of all the statements each of which is a possible answer to a question about some feature of concern, and among which one has preferences (1973, p. 461)."

In bibliographic retrieval system performance evaluation terms, Wilson's concept of the 'concern set' could be defined as 100 percent recall with 100 percent precision, or retrieval of all and only those document 'representatives' contained in a data base which express the concern, or topic, expressed in the query. For purposes of this discussion, the query statement and the information need will be considered to be the same, although it is recognized that in reality the user's need is often not fully or accurately expressed in the query statement. The user would then choose those document representatives from the concern set which are pertinent to his/her particular situation - or those which s/he 'prefers'.

Use of the notion of the 'concern set' as a basis for the operational definition of 'relevance' for an experiment would appear to remove the most serious research problem imposed by 'pertinence' and 'utility' - i.e., the inherent subjectivity of assessments of performance based
on these definitions. No two users presenting the same query and receiving the same retrieved set will make the same 'pertinence' or 'utility' judgements. However, objective means by which the concern set of documents for a query could be determined may also not be possible to define. Selection of the concern set would require human judgements and the criteria by which these judgements could be made more objective have not been established. One of the objectives of this study is to determine whether language-based criteria can be developed which will provide at least a partial basis for establishing an objective concern set for a query in a research situation.

Lancaster suggests that in bibliographic retrieval experiments:

"The relationships between document representation and a search strategy are in a sense internal to the information system, and decisions as to whether or not appropriate matches have occurred need not involve either system users or other subject specialists. It is probably best if we use neither the word 'relevance' nor the word 'pertinence' for these relationships but simply refer to document representations 'matching a search strategy' and documents 'matching an intended search strategy' (1978, p. 260)."
In other words, Lancaster is suggesting here that a 'match' between word tokens in the search strategy and in a document representative is not relevance per se. By the above definition, the sense and referents of the tokens, evaluation of which would require human judgement, are not involved in evaluations of internal matching. An operational definition of relevance based on matching would be valid if the accuracy with which the system responds to a search strategy, regardless of subject content or intention, was being tested.

Although Lancaster's statement does not address the problem of developing more objective criteria for other kinds of bibliographic retrieval system evaluation, it presents an early objective definition based on some commonality between a search strategy and retrieved document representatives, which can be discerned without the aid of interpretation or judgement by human users or subject specialists, as distinct from subjective measures.

However, as Lancaster says later in the same chapter:

"...we are concerned not only with matches between documents, or their representations, and searching strategies. We are also concerned with relationships between documents and request statements and between documents and the information needs of users (1978, p. 261)."
In other words, we must find these relationships as well, and the relationships between documents and queries or between documents and users' information needs involve meaning and subject content and human judgement.

Saracevic indirectly suggests, several times, a possible focus for exploration of the relationships between a query and a relevant document or document representative. Each of the following statements narrowed the suggested focus for studying the characteristics of this commonality.

"Although a number of factors are aligned with documents as variables, the most important of these factors affecting relevance judgements appears to be the subject content of documents as compared to the subject content of the query (1975, p. 340)."

This most general commonality between the query and relevant documents - i.e., subject content - is the focus which has received the most attention in construction of indexing systems. Saracevic goes on to say:

"A close similarity and correlation seems to exist between texts of queries and texts of relevant documents which cannot be found between texts of queries and texts of non-relevant documents (1975, p. 340)."
This statement suggests a focus on the content or structure of texts of document representatives themselves as opposed to the more general subject content approach. And, finally, he states:

"Document texts may be the most important factor in triggering relevance judgements in relation to stated queries, i.e., if one finds a statement in a document resembling a query statement, one is assured of a high probability that the document will be considered relevant by the user stating the query (1975, p. 340)."

However, Saracevic does not go beyond these statements to suggest what the commonality is between relevant documents, or document representatives, and query statements. That is, what characteristics do the query and relevant texts have in common which the query and nonrelevant texts do not have in common.

The most important factor in bibliographic retrieval which suggests a deeper investigation of language as the key to underlying generalizations about relevance decisions is the fact that both relevant and nonrelevant document representatives are retrieved by the same search terms. Changing the criteria for relevance assessments from 'usefulness' to a 'match in topic' between the query and the members of the retrieved
documents or document representatives does not eliminate all nonrelevant members of the set. Some document representatives retrieved by the search strategy are simply not about the topic named in the query even though they contain the same words - i.e., terms or phrases - as the query and/or relevant document representatives. Thus relevance appears to be based on what is meant, not simply on what words occur. Nonrelevance is due to a mismatch between the intended topic of the user and the searcher and the actual topic expressed in the retrieved document representatives. In other words, there is some significant difference in meaning, or semantics, either between the topic in the query and the topic expressed in the retrieved nonrelevant text, or between the statement containing the search terms in the retrieved nonrelevant document representative and the statement in the query.

As Maron and Kuhn point out in a very early article on relevance:

"One of the really remarkable characteristics of human beings is their ability to communicate with and operate on information formulated in ordinary language. We somehow are able to determine the meanings of words and sentences so as to make judgements about sameness of meaning,
redundancy, inconsistency, relevance, etc., in
spite of the fact that ordinary language is
extremely complex and fraught with vagueness
and ambiguity (1960, p. 216)."

Like most articles which discuss the issues raised in
this study, this description was in reference to the
indexing process. Again, as Maron says:

"...part of the process of understanding a
piece of writing is 'seeing' what it is about.
As we have said earlier, this inner ability to
recognize what a document is about is the very
heart of the indexing procedure (1977, p. 39)."

Wilson, who also has written on relevance, meaning,
and aboutness, discusses the language in the text of
documents and several of his comments are at the heart of
the study undertaken here. According to Wilson:

"...That two documents contain the same text
does not guarantee that they contain the same
information; nor does the fact that they contain
different texts mean that they must contain
different information...The ability to para-
phrase is just the ability to produce different
linguistic texts, in the same or different
languages, that carry the same information...
[However]...the information contained in a
document cannot be identified with the text of the document. But it may be identified with the sense, meaning, propositional content, or semantic content of the text (1978, p. 11)."

Continuing in this vein, he notes that the meaning of a text is not contained in the text itself, but in peoples' interpretations of the text when they read it. These interpretations are made possible by the structure of the text, the way in which the words are strung together and the way in which the sentences are organized. And these interpretations are at the heart of the indexing process and the relevance decision-making process which are central in bibliographic retrieval activities. It is this human ability to interpret and paraphrase the subject content of a document or a document representative in relation to textual characteristics that is of interest in this study.

In summary, useful generalizations about why terms which retrieve relevant records also retrieve nonrelevant records and how consistently people distinguish between relevant and nonrelevant records have not been adequately developed in information science. Proof that one or more language characteristics consistently play a part in discrimination between relevant and nonrelevant, or 'about' and 'not about', texts would be an important step
in the development of a theory of relevance. Such proof would explain a major factor affecting users' discrimination between relevant and nonrelevant retrieved document representatives.

It is the ability of the competent speaker-hearer-reader of English to distinguish among:

a) texts which contain the same words, but do not contain the same semantic content,
b) texts which contain the same words and do contain the same semantic content, and
c) texts which do not contain the same words, but do contain the same semantic content

that is of interest in this study. To this end, tests were conducted to determine how consistently readers can make these language distinctions, what variables affect the ability of people to make these distinctions in a bibliographic retrieval system environment, and to what degree a functional relations model based on a case grammar can be used to describe the distinctions the readers make.

The tests conducted in this study included the determinations of:

1. the consistency of subject specialists' language behavior in making aboutness decisions;
2. the ability of secondary judges to match the aboutness decisions of the subject specialists;
3. the consistency of the secondary judges' language behavior in making aboutness decisions;
4. the effect of subject knowledge on the consistency of language behavior in aboutness decision-making by the secondary judges;
5. the level of agreement between the aboutness decisions of the researcher (acting as first language analyst) and the aboutness decisions of a second language analyst using the functional relations model as a guide in decision-making; and
6. the level of agreement between the researcher and the second language analyst on the assignment of specific functional relations to keywords in query statements and abstract statements.
0.3. The Structure of the Thesis.

The description of the research conducted for this thesis is comprised of the following chapters:

CHAPTER 1 includes an introduction to language and linguistics as they are viewed in this study, a description of the assumptions underlying Fillmore's case grammar - i.e., the grammar used in the analysis of language behavior - and the definitions of the functional relations used in the analysis of query statements and abstracts.

CHAPTER 2 provides the context for this thesis by discussing three indexing systems which incorporated functional relations and a study by Courrier who tested the ability of subjects to assign functional relations consistently to sentences in abstracts.

CHAPTER 3 describes the methodology employed to obtain the data for the various tests and includes the definitions of the parameters and variables and the test procedures for the three tests conducted in this study.

CHAPTER 4 describes how the analysis of the data was accomplished including the analysis of the queries, abstracts, and relevance decisions, the assignment of functional relations, and the analysis of the data for all tests including the tabulation of the results.
CHAPTER 5 provides the description of the statistical tests applied to the results of the various analyses and the results of these tests.

CHAPTER 6 discusses the functional relations model, including definitions and the problems and strengths of the model when it is applied to data obtained from real human behavior in aboutness decision-making.

CHAPTER 7 provides the conclusions made on the basis of the results of this study and offers speculations about the meaning of these results for future research in Information Science.
CHAPTER 1
FILLMORE'S CASE GRAMMAR

This chapter provides a brief description of the theoretical background of Fillmore's case grammar and of the functional relations - i.e., the case roles and verb types described by Fillmore - which were used in the analysis of the relations between keyterms in queries and in abstracts. The discussion begins with a brief description of language, linguistic grammars, the language abilities of competent speakers, and linguistic assumptions and concepts underlying transformational grammars and, therefore, Fillmore's case grammar. The view of language, linguistics, grammars and human language behavior described in this chapter underlies the approach taken to the study of language behavior in this thesis. The rest of the chapter is devoted to a description of Fillmore's grammar, including the definitions of the case roles and verb types used in the analysis of the data.
1.0 Language and Linguistic Grammars.

For purposes of this project, a language will be described simply as a vocabulary - i.e., a set of words - and a grammatical structure which describes how members of the vocabulary are combined into phrases and sentences in meaningful human communication. The vocabulary and the syntax, or grammatical structure, of any natural language are finite. That is, at any one time there are a specific number of words and ways of combining words which exist, while the number of possible sentences is infinite.

Words have the qualities of sense and/or reference which define their possible range of meanings. Most words have at least one 'sense', - i.e., they have some inherent meaning. But not all words have reference. A referent can be an object, an experience, an event, etc., which can be shown to exist in reality - e.g., a ball, a cloud, a picnic, a child, etc. A referent can also be a mythical object or entity - e.g., a unicorn, a demon, god(s), fairydust, etc. - or an abstract concept - e.g., an idea, a theory, a belief, etc. Words such as a, the, neither, do not have referents, but are used to help specify the meaning - i.e., the sense or reference - of other words in a sentence or of the sentence as a whole. Other words serve as syntactic indicators with little, if
any, descriptive content or referential use as in the sentence "It annoyed the Russian that he couldn't hit the ball, where it and that have no descriptive content and do not refer to anything outside the sentence (Schachter, 1973, p. 3).

The grammatical structure of a language is the way in which words are combined into meaningful sequences. We 'know' that in English John gave Harry the book is meaningful while John book Harry the gave is not. We also 'know' that John ran is meaningful but The book drank is nonsensical. How we distinguish the meaningful from the nonmeaningful or the nonsensical is a subject of much investigation. The study of how we use vocabulary and grammatical structures to convey meaning is part of the discipline of linguistics. Philosophy and psychology and recent research in human brain function also study this phenomenon. But we are concerned in this project with linguistic theories.

A task taken on by some linguists is construction of abstract grammars which "...consist of a finite set of rules operating upon a finite vocabulary and... capable of generating an infinite set of sentences (Lyon, 1970, p. 4)". The assumption is that the same set of rules needed to handle simple sentences should be able to handle complex sentences as well through recursive application of the rules. (By recursive application is meant
that the same rule can be applied more than once in the
generation of the same sentence.)

There are two central features of the linguistic
capacity of competent speakers of a language which must
be considered when constructing an abstract grammar.
First, competent speakers are capable of generating
sentences they have never heard before and have never said
before. This is called the 'creative aspect' of language
behavior. Second, they are capable of understanding as
similar expressions which differ in structure - e.g.,
active and passive versions of the 'same' sentence -
and/or in vocabulary - e.g., the different words which
carry the same or similar meaning such as synonyms,
generically related words, etc.

There is much day to day evidence of these two
language abilities of human beings, both native and non-
native speakers of English, in normal bibliographic
retrieval situations. People ask questions and attempt
to describe concepts or relations between concepts that
they have never verbalized before. And users are able
to determine whether or not retrieved items are useful to
them even though the syntactic structure and vocabulary
used by different authors may differ significantly.

So far the discussion has been about language and
linguistics in general. The next section will describe
the assumptions about language underlying the theories of
Noam Chomsky and Charles Fillmore. Fillmore's case grammar theory was derived in part from Chomsky's "Standard Theory" (1965) and carries some of the same assumptions about the nature of language behavior and grammatical structure.
1.0.1. The 'Competence'/ 'Performance' Assumption.

One essential assumption underlying Chomsky's theory is that human beings internalize the structure of their language as children at a psychological level and that this knowledge is largely unconscious. He called this internalized knowledge of the structure of a language 'competence'. The actual utterances of a person, which are filled with false starts, errors, repetitions, corrections, etc., he called 'performance'. So 'competence' is the speaker-hearer's unconscious knowledge of his language; 'performance' is the actual use of language in real situations. It is the speaker's linguistic 'competence' that is most regular and systematic. Therefore, Chomsky claims that a linguistic grammar should describe 'competence', not 'performance'.

In Chomsky's view "a grammar of a language purports to be a description of the ideal speaker's intrinsic competence... and... the problem for the linguist is to determine from the data of performance the underlying system of rules that has been mastered by a speaker-hearer and that he puts to actual use in actual performance (1965, p. 4)."
1.0.2. Deep Structure, Surface Structure, and Transformational Grammar.

Two linguistic concepts upon which assumptions underlying the hypothesis are based are Deep Structure and Surface Structure. The Surface Structure of a sentence is the way in which the various units comprising the spoken sentence - e.g., noun phrases, verbs, adverbs, etc. - are combined by a speaker. For example, the following two sentences have different Surface Structures:

(1) The elephant is eating peanuts.

(2) Peanuts are being eaten by the elephant.

The Deep Structure of a sentence is a more abstract description than its Surface Structure and is very unlike the sentence actually spoken. This description is made up of a restricted set of abstract linguistic elements and includes the most basic structural and semantic information necessary for deriving the meaning of a sentence, independent of the superficial variation possible in the Surface Structure. Deep Structure, therefore, is a level of linguistic analysis at which meaning relations are explicit and unambiguous. Thus, although sentences (1) and (2) have very different Surface Structures, they have essentially the same Deep Structure and a competent speaker of English would
recognize that they have essentially the same meaning. The Deep Structure reflects, then, the language 'competence' of a speaker, what the speaker has to know about his/her language to discern that the two sentences above have much the same meaning.

A Transformational Grammar attempts to explain the relationship between Deep Structure and Surface Structure. The relationship between the Deep Structure of a sentence and the Surface Structure of a sentence is very complex and the process by which the Deep Structure of a sentence is transformed into various Surface Structures; or vice versa, need not be explained for our purposes. The important point here is what a Transformational Grammar must explain.

Sentences (1) and (2) show that a grammar must account for sentences which have different Surface Structures, but the same Deep Structure. A Transformational Grammar must also account for sentences which have the same Surface Structure, but different Deep Structures. For example, a competent speaker of English would see that the following sentence is ambiguous:

(3) Visiting professors can be boring.

That is, the sentence has the two possible meanings given in sentences (4) and (5) as follows:

(4) Going to visit professors can be boring.

(5) Professors who are visiting can be boring.
The ability to see these two possible interpretations of Sentence (3) is also part of a speaker-hearer's language 'competence'.

A Transformational Grammar must reflect, then, the abilities of competent speaker-hearers to distinguish among:

a) sentences which have different Surface Structures but the same Deep Structure and, thus, more or less equivalent semantic interpretations;
b) sentences which have the same Surface Structure but different Deep Structures and, thus, different semantic interpretations; and, of course,
c) sentences which have different Surface Structures and different Deep Structures and, thus, different semantic interpretations.

These three abilities of a competent speaker-hearer are also reflected in people's language behavior in a bibliographic retrieval situation. The following examples are analogous to points (a) and (c) above, but are based on search terms, queries, and retrieved abstracts.

The following query was included in the test sets for this study. (The example has been simplified for the purpose of this illustration.)

Query: I want articles that discuss teaching computer literacy to children.
The search terms entered for this query were COMPUTER
(Boolean AND) LITERACY (Boolean AND) TEACHING (Boolean AND)
CHILDREN. Let's say the following statements were included
in the texts of abstracts retrieved in response to the
search strategy:

(6) Children were taught computer literacy.

Sentence (6) has essentially the same semantic content as
the query statement, but has a different Surface Structure.

(7) A computer was used to teach children literacy
skills.

Sentence (7) is ambiguous. We do not know if literacy
skills refers to computer literacy skills or to language
literacy skills.

(8) A computer was used to teach language literacy
skills to immigrant secondary school children.

This sentence has a different Surface Structure from the
query statement and also has a different meaning, or Deep
Structure. That is, language literacy is being taught,
not computer literacy. These examples show that the
general requirements of a grammar outlined above also apply
in a bibliographic retrieval situation.

In sum, Chomsky defines a grammar and its requirements
as follows:

"A grammar can be regarded as a theory of a language;
it is descriptively adequate to the extent that it
correctly describes the intrinsic competence of the idealized native speaker. The structural descriptions assigned to sentences by a grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be immediately aware of this) in a substantial and significant class of crucial cases (1965, p. 24)."

Chomsky's *Aspects of the Theory of Syntax* (1965) described a kind of grammar, Transformational Generative Grammar, that met his requirements. It is a proposed variation of this grammar, case grammar, that concerns us here. The requirements of a grammar discussed in this section are the requirements that case grammar will be tested against in this study, that is, its adequacy for describing textual characteristics and relevance decision-making in the bibliographic retrieval environment.
1.1 Fillmore's Case Grammar.

Charles Fillmore's case grammar was proposed as a transformational grammar with the addition of an underlying set of cases which described the functional relations between a predicate (the verbal element) and its arguments (the nominal elements related to the verbal element). Fillmore's grammar, like Chomsky's, is intended to describe the 'intrinsic competence of the idealized native speaker'. Fillmore developed this theory to introduce a set of generalizations which he claimed people have internalized about the deepest level of categories used in grammatical structure, categories he called 'cases'. Fillmore suggests that:

"...the case notions comprise a set of universal, presumably innate, concepts which identify certain types of judgements human beings are capable of making about the events that are going on around them, judgements about such matters as who did it, who it happened to, and what got changed. (1968, p. 24)."

These judgements are made at the linguistic competence level, that is, at the psychological level of the basic knowledge all people have about the descriptive patterns of their language. By 'case' Fillmore means something other than grammatical subject or grammatical object that
have been traditionally referred to by this term in such case languages as German, Finnish, or Latin. Rather, he is referring to a more abstract notion that is not dependent on the word endings of any specific language. And by 'descriptive patterns' he means something other than syntactic structures. He is referring to the descriptions of the relations between things and states or activities that are being described in language expressions.

For example, a competent speaker can understand that the following sentences contain essentially the same general information. The same elements, elephants, eating, and peanuts, occur in them all and are related to one another in the same way:

(9) Elephants eat peanuts.
(10) Peanuts are eaten by elephants.
(11) That elephant is eating a peanut.
(12) Those elephants ate peanuts yesterday.
(13) Peanuts are among the foods elephants eat.
(14) My elephant likes peanuts for his afternoon snack. (Eat is implied by snack.)

The next set of sentences, on the other hand, are not about the same subject even though they contain the same three words. That is, the relationship between the elements is different:

(15) Tom is eating peanuts and watching the elephant.
(16) The elephant standing in the peanut field is eating a banana.

(17) The poachers ate elephant steaks and peanuts - for dinner.

(18) I ate peanuts while I rode the elephant.

Fillmore's grammar attempts to expand the explanations of how people differentiate between the various meanings of sentences which contain the same words, but do not carry the same meaning, and how people recognize as essentially equivalent sentences which have very different surface structures.

According to Fillmore, a proposition - i.e. a sentence or a clause - is made up of two components:

a) a predicate, or the verbal element, and

b) arguments, or noun phrases.

The meaning relations between the predicate and its arguments can be explicated unambiguously regardless of the surface syntactic relations used in the expression of the proposition because the deep structure relations between the noun phrase(s) and the verbal element hold constant despite variations in surface representation - e.g., active and passive versions of the 'same' sentence. Thus, the grammatical subject or grammatical object and the case notions of specific languages - e.g. genitive, dative, accusative, and so forth - are not in any way
implied by Fillmore's notion of 'case'. Rather, the case roles in this theory correspond more closely to the elements of a journalist's report: who, what, when, where, to whom, with what, and how.
1.1.2. The Assumptions Underlying Fillmore's Case Grammar.

There are several assumptions underlying the actual structure of Fillmore's case grammar:

1. that there are a finite, and small, number of case relations which can be used to describe the relations of noun phrases to the verb in a sentence - i.e., the set of cases comprising the grammar must describe all possible functions a noun phrase can assume in a sentence;

2. that each case role must be atomic - i.e., cannot be broken up;

3. that the relations which noun phrases can assume to the verbal element are governed by, and can be defined in terms of, the verbal element; and

4. that no case relations, except where noun phrases are conjoined and bear the same relation to the verb, can appear more than once in a proposition.

The case roles used in this study were described and developed by Fillmore over several years. Where changes have occurred in his original concepts, the 1971 definitions will be used. The descriptions of the case roles which follow are taken primarily from Fillmore's own definitions which are being accepted without alteration at this point. The set of cases is: Agent, Instrument, Experiencer, Object, Source, Goal, Location, Path and Time.
The adequacy of this set of cases and their definitions will be discussed after the description of the analysis of the data.
1.1.2. The Cases.

AGENT. "...the case of the typically animate perceived instigator of the action identified by the verb (1968, p. 24)." In the sentence John opened the door, John is the Agent.

INSTRUMENT. "...the case of the inanimate force or object causally involved in the action or state identified by the verb (1968, p. 24)." Later Fillmore extended this case to include "...the immediate cause of an event, or, in the case of a psychological predicator, the 'stimulus', the thing reacted to (1971, p. 42)." For example, in the sentences John opened the door with the key and The door was opened with a key, John is the Agent and the key is the Instrument. The central requirement of the Instrument case is, then, that some direct causal relationship must exist with the Instrument as the 'cause'. In addition, the Instrument role can be occupied by a sentence when "...that sentence identifies an event which is understood as having some other event or state as its consequence (1971, p. 42)."

EXPERIENCER. "...where there is a genuine psychological event or mental state verb, we have the Experiencer (1971, p. 42)." Typically, in a sentence describing a psychological event, we have the Experiencer and some other noun
phrase which indicates the cause or the content of the mental state or psychological event. In the sentence John was afraid of the dog, John is the Experiencer and the dog is the Agent. In the sentence The noise reminded me of the accident, the noise is the Instrument, me is the Experiencer and the accident is the Object. The first sentence is a description of a mental state and the second is a description of a psychological event.

**OBJECT.** "...the semantically most neutral case, the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself; conceivably the concept should be limited to things which are affected by the action or state identified by the verb. "The term is not to be confused with the notion of direct object, nor with the name of the surface case synonymous with accusative (1968, p. 25)."

Fillmore later further explicates the conditions under which we have the Objective case. The Object occurs when "...there is a nonpsychological verb which indicates a change of state, such as one of dying or growing (1971, p. 42)," or when an entity which moves is the Object and the thing to which it moves or upon which it impinges is the Goal. In the sentence Mary hit the fence with a stick, Mary is the Agent, the fence is the Goal and the stick is the Object.
GOAL and SOURCE. "...depending upon the type of predicator, the Source and Goal are interpreted as earlier and later locations, earlier and later states, or earlier and later time points (1971, p. 41)." Also, "...where there is a transfer or movement of something to a person, the receiver as destination is taken as the Goal (1971, p. 42)." The Goal case is also used to indicate the later state or end result of some action or change. That is, it specifies the end-result role of a thing which comes into existence as a result of the action identified by the predicators, as in I wrote a poem or I constructed a bridge where poem and bridge are both Goals. A 'sentence embedded as Goal is one which identifies the resulting state or event in a causative construction.

PATH. In addition to Source and Goal, Path is another complement to verbs of motion as in He walked from the cemetery gate to the chapel along the canal, where cemetery gate is the Source, the chapel is the Goal and along the canal is the Path. A sentence with the Path designated can contain an unlimited number of Path expressions, as long as these are understood as indicating successive stretches of the same Path.

LOCATION and TIME. These cases are optional complements of essentially any predicator, and indicate the place or time in which an event occurs. They are not to be
confused with Source and Goal constructions which designate relations in time or place with verbs of motion—e.g., in the sentence John walked from the porch to the corner, porch and corner are earlier and later locations, Source and Goal; in the sentence John is standing on the porch, on the other hand, the porch assumes the Location case role. The difference between Location and Time and Source and Goal is further evidenced by the fact that all can occur in the same sentence as in He walked from the Museum to the Park along Central Avenue in Anytown on Tuesday afternoon, where Museum is the Source, Park the Goal, Central Avenue the Path, Anytown the Location and Tuesday afternoon the Time.

Although there is disagreement among linguists about whether or not the above nine case roles constitute all and only those cases required to describe all possible functional relations between noun phrases and the verbal element in a proposition, they will be considered here to be both complete and atomic in the initial approach to the analysis of the data. After all of the data to be analyzed for this study have been obtained, the consistency with which the set of cases being used here can be applied by different persons will also be tested.
1.1.3. The Verb Types.

According to Fillmore's theory, the basic structure of a sentence or a clause (what he calls a proposition) is made up of:

"...a tenseless set of relationships involving verbs and nouns (and embedded sentences, if there are any), separated from what might be called the 'modality' constituent.

There are four types of predicates, or verbs, in this grammar: State, Process, Action, and Action/Process. These are called the 'verbal frame'.

**STATE**: a fixed condition in time and/or space as in:

a) The car *is parked* in the garage.

b) The house *overlooks* the river.

c) John *is tall* for his age.

**PROCESS**: an ongoing event which involves some change in the state of the nominal element, but not by direct action as in:

a) The leaves *are changing* color.

b) The children *are learning* to read.

c) John *is growing* old.

**ACTION**: an activity which does not produce a change in a nominal element as in:

a) Mary *is running* to catch a bus.

b) The car *sped* down the highway.

c) The child *is reading*.
ACTION/PROCESS: activity which produces a change in a nominal element as in:

a) The children are being taught to read.

b) Mary is building a house.

c) The cat is washing its fur.

The verbal element provides clues as to what case roles the nominal elements in a sentence may assume. For example, the case role Instrument occurs only with the verbal frame ACTION/PROCESS.
1.2. The Use of Case Grammar to Describe Functional Relations.

Returning to the earlier example of **peanuts** and **elephants**, examples (9) to (14) contain the verbal element **eat** which is an Action/Process verb, the nominal element **elephant** which assumes the role of Agent, and the nominal element **peanuts** which assumes the role of Object of the verb **eat**. Even though each proposition is expressed differently syntactically, the relations between the elements in the proposition do not change. The second set of propositions, sentences (15) to (18), contain the same three words, but each sentence is not only syntactically different from sentences (9) to (14), but also semantically different. That is, the nominal elements do not assume the same relationship, or case role, in relation to the verbal element **eat**. Thus the deep structures are different.

Using Fillmore's case roles to describe the reasons for the differences between the two groups of sentences we have:

Sentences (9) to (14), **Elephants eat peanuts**, where **Elephants** is the Agent, **eat** is an Action/Process verb, and **peanuts** is the Object of the verb **eat**.
Sentence (15), Tom is eating peanuts and watching the elephant, where

Peanuts is the Object of the verb eat, but Tom has replaced elephant as the Agent and elephant is the Object of the verb watch.

Sentence (16), The elephant standing in the peanut field is eating a banana, where

elephant is the Agent of the verb eat, but a banana is the Object of the verb eat. 'Peanut is an adjective attached to field and the clause standing in the peanut field specifies which elephant is being talked about.

Sentence (17), The poachers ate elephant steaks and peanuts for dinner, where

peanuts is the Object of the verb eat, but poachers is the Agent and elephant is part of the other Object of eat.

Sentence (18), Tom ate peanuts while he rode the elephant, where

peanuts is the Object of the verb eat, but Tom is the Agent and elephant is part of another clause, while he rode the elephant, which assumes the case role Time - i.e., the Time during which the peanuts were eaten.

We can see from these examples that, while the sense and reference of the three key elements remain the same, the
semantic content of the first group of sentences (9) to (14), is essentially the same, while that of each of the sentences in the second group, (15) to (18), is quite different.

Another example of the difference between the traditional definition of grammatical subject and grammatical object or other case notions of specific languages and the notion of case in this theoretical framework is demonstrated in the following sentences:

(19) John opened the door with the key.
(20) The key was used to John to open the door.
(21) The door was opened by John with the key.
(22) John used the key to open the door.

All are equivalent expressions according to Fillmore’s theory. The argument that these sentences differ semantically because the emphasis changes when the grammatical subject changes is certainly valid. However, for Fillmore, the important point is that the functional relations between the nominal elements and the verbal element remain the same. In all four expressions:

John (Agent) did the opening;
the door (Object) was opened; and
the key (Instrument) was used to open.

The variations in surface syntactic structure do not alter these functional relations.

The previously discussed query statement and sentences which could have been retrieved from a
bibliographic retrieval system in response to the query can now be used to illustrate how functional relations operate in a bibliographic retrieval situation.

Query: I want articles on teaching computer literacy to children.

The case grammar analysis of the embedded sentence in the query, the sentence containing the keywords, is:

(23) (Someone) - teaches - computer literacy - Agent Action/Process Goal
to children. Object

The retrieved statements could include the following:

(24) The teaching guide for children in grades 4-6 includes lessons in computer literacy.

This sentence contains the same words and the same semantic content as the query statement although the Surface Structure is quite different. To illustrate, this sentence could be restated as:

(24i) The teaching guide includes lessons for teaching computer literacy to 4-6 Grade children.

Sentence (24i) contains the same embedded sentence as the query statement. The case grammar description is, then:

(24ii) (Someone) - teaches - computer literacy - Agent Action/Process Goal
to children. Object
A competent speaker-hearer would have intuitively recognized the equivalency of the query statement and Sentence (24).

The next sentence illustrates another important ability of competent speakers:

(25) The curriculum guide includes a plan for instructing public school students in basic computer skills.

Sentence (25) does not contain the words literacy, teach, or children. In addition, the Surface Structure of this sentence is quite different from the Surface Structure of the query statement. However, the case grammar analysis of the embedded sentence is the same as for the Query as follows:

(25i) (Someone) - instructs - students - in basic computer skills.

A competent speaker-hearer of English would also know that:

a) instruct is a quasi-synonym for teach;

b) basic computer skills is a quasi-synonym for computer literacy; and

c) students is a quasi-synonym for children.

There is a very important difference between Sentence (25) and all of the other sentences discussed in this chapter. That is, the earlier examples described differences and
similarities in Deep Structures and Surface Structures that any competent speaker-hearer would recognize. Sentence (25) requires more than a general knowledge of the structure of language to identify it as relevant to the query. It requires a knowledge of vocabulary, or some knowledge of a subject and the terms used to name that subject. The important point here is that terms that are considered to be equivalent in meaning - i.e., that are considered to be in some way synonymous - must assume the same case role as their equivalents in a sentence for the sentence to be considered to be equivalent.

The purpose of the test conducted for this project was to determine:

1. How consistently readers can make distinctions of the kind described in this chapter;
2. what variables affect the ability of people to make these distinctions in a bibliographic retrieval situation; and
3. whether case grammar can be shown to reflect this language ability.
CHAPTER 2

RELATED RESEARCH ON FUNCTIONAL RELATIONS IN INFORMATION SCIENCE

The purpose of this chapter is to place this study in the context of related research in functional relations in information science. While the methodology employed here is founded on research in information system performance and studies of relevance, the related research discussed here represents work involving case grammar or case grammar-like notions including:

1. indexing systems employing role or relational indicators; and

2. a similar test of language behavior which involved a case grammar model.

The indexing systems discussed have attempted to incorporate role or relational indicators in pre-coordinate systems - i.e., the indicators are attached to keyterms at the time the keyterms are entered into the system. It is then the user's decision which indicators to include in the search strategy. These systems assume that indexers and searchers will be able to recognize the relations that are not pertinent to their purpose. However, as will be seen in the brief description of the study by Courrier, which tested the ability of subjects to specify verb types
in abstracts, the assumption that indexers and searchers will be able to determine easily the relations between keyterms in abstracts or in a query may be erroneous.

The study by Courrier demonstrates clearly that there are difficulties in applying the definitions of case roles and verb types when using a case grammar system proposed by Cook (1974). The structure of this system is different from that of Fillmore's grammar, which is being used in this study, but it is being assumed that some of the problems Courrier encountered must be taken into account here.

There are three levels of relations between words in the bibliographic retrieval environment which are of concern here. They are:

1. relations between two or more words forming a key phrase, a phrase which is treated as a single keyterm whether it contains one or more words;
2. relations between keyterms in a sentence which contains two or more keyphrases, usually involving titles and single sentence query statements; and
3. relations between keyterms in multi-sentence structures, usually involving abstracts and lengthy query descriptions.

At level 1 the problems facing system designers and system users is the determination of the method to be used in describing concepts which required two or more words, such as a noun and an attribute, which are considered to be inseparable when describing a concept. At level 2, the sentence level, system designers and users are concerned with the functional relations between keyterms and keyphrases in titles and in single sentences or strings within sentences in abstracts. At this level the necessary

*a clause or phrase which contains more than one keyphrase within a sentence.
KIS (location) - Location
KWG (direction from) - Source
KWC (direction to) - Goal
KIT (time occurring) - Time
KIG (location-geographic) - Location
KIB (organization involved) - Agent (collective)
KEP (person involved) - Agent

The more complete definitions and the examples which follow served as the guide for determining possible correspondences between Fillmore's case roles and the WRU role indicators. They are taken from the Glossary of Role Indicators provided by Perry (1958, p. 128ff).

The discussion will proceed in sections divided by the case roles. In order to make clear the relations which are being compared the example will be as follows:

(n) (KIB) Steel producers - are shipping steel. Agent

where the role indicator immediately precedes the nominal element which assumes that role, and the corresponding case grammar role is given under the nominal element which assumes the role. The rest of the sentence is not analyzed so that the focus is on the particular role indicator being discussed.
2.1 Indexing Systems Employing Role or Relational Indicators.

Efforts to use some method of indicating functional relations in the more traditional areas of information science have been directed toward manual indexing systems—i.e., toward indexing systems in which the relations between terms are assigned directly by human indexers before the items are entered into an automated bibliographic retrieval system. The three indexing systems which will be discussed here have some comparable features which are of special interest because they provide corroboration for the finding in this study that relations between concepts, or between keyterms, are perceived to be important in the bibliographic retrieval environment. They also display a difference in the complexity of the relations they employ for searching and indexing.

Index terms are really a form of paraphrase of the important ideas contained in a title, abstract or document. Indexing systems provide the rules under which keyterms in texts are to be chosen and
represented. Systems which use single word index terms provide no means for indicating anything more than that a term must appear in a record. The context in which a term appears or the intended relationship between two terms cannot be indicated in a search strategy in these systems. The indexing systems described here have attempted to provide a means by which an indexer or a searcher can indicate certain functional relations between two or more keyterms or, at least, the function a keyterm assumes in the larger text.

The systems will be described in the order of the complexities of the relations they attempt to preserve. They are:

1. the Western Reserve University (WRU) indexing system which assigns roles to single keyterms - i.e., keyterms with their relational indicators are entered as separate entities into the system and are searched as individual keyterms;
2. the PREserved Context Indexing System (PREGIS) which purports to use role, or relational, indicators which are based on case grammar and which preserves relations through use of word order and prepositions; and

3. Farradane's Relational Indexing System (FRIS) which is based on a model of human thinking and memory and which employs a complex set of symbols to indicate relations between two or more keyterms.

The first part of this section will consist of a description of the case roles and verb types which are covered by each of the three systems, how the role or relational indicators are applied, and the nature of the descriptions of functional relationships they provide.

The second part of this section will consist of a brief summary of the comparison between the three systems and Fillmore's case roles and
2.1.1.2. INSTRUMENT.

The WRU roles which correspond to the Instrument case role include the following:

3. KAL (influenced by) 'is prefixed to any term that indicates a material, process, etc. which influences a preceding property...' as in:

(4) different magnetic moments are present depending on

(KAL) the type of atomic distribution.

This role indicator corresponds most closely to the Instrument role when it is occupied by a sentence.

4. KQJ (by means of) 'is prefixed to a term that designates the actual agent or instrument effecting or used to effect a process, testing technique or function.'

(5) polishing with - (KQJ) abrasives.

(6) corrosion of tankers - (KQJ) by sea water...
2.1.1.1. AGENT.

The WRU roles which correspond to the Agent case role include two separate indicators which differentiate between organizations acting as Agents and persons or groups of persons acting as Agents.

1. KIB (organization involved) is prefixed to a name of an organization... From the example we can see that, although no mention is made of the organization as the instigator of an action, the correspondence seems to be quite clear.

   (1) an aluminum fabricating plant is being built at Alicante by - (KIB) the Aluminio Iberico, S.A. Agent

2. KEP (person involved) is prefixed to either the name of a person or persons considered as a group such as workers, unions, management...

   (2) (KEP) steel producers - are shipping steel. Agent

   (3) (KEP) Elihu Thompson - recognized the Agent principle of welding.

We can see from these examples that the two role indicators described are used to indicate nominal elements which would be considered to be Agents in Fillmore's model.
"Limiting the number and defining the scope of possible synthetic relationships to those which are considered most useful for searching metallurgical documents led to the necessity of fixing certain arbitrary boundary lines....Since metallurgy is concerned with materials (metals, alloys, etc.), properties of materials, and processes for dealing with materials, the chief areas of searching will be those areas which include relationships between materials, processes, and properties, along with conditions of processes, as designated by role indicators (1958, p. 102ff)."

In other words, the creators of this system are not claiming that the roles or relations they include are in any way universal to other subject disciplines or to other languages. The list of role indicators which will be discussed here and the brief definitions given by the author are as follows:

- KEJ (material processed) - Object
- KAJ (starting material) - Source
- KWJ (product) - Goal
- KQJ (by means of) - sometimes Instrument
- KAL (influenced by) - Instrument
- KAM (process) - Verbs
- KAH (condition) - No direct match
- KAP (property influenced) - Object
KIS (location) - Location
KWG (direction from) - Source
KWC (direction to) - Goal
KIT (time occurring) - Time
KIG (location-geographic) - Location
KIB (organization involved) - Agent (collective)
KEP (person involved) - Agent

The more complete definitions and the examples which followed served as the guide for determining possible correspondences between Fillmore's case roles and the WRU role indicators. They are taken from the Glossary of Role Indicators provided by Perry (1958, p. 128ff).

The discussion will proceed in sections divided by the case roles. In order to make clear the relations which are being compared the example will be as follows:

(n) (KIB) Steel producers - are shipping steel. Agent

where the role indicator immediately precedes the nominal element which assumes that role, and the corresponding case grammar role is given under the nominal element which assumes the role. The rest of the sentence is not analyzed so that the focus is on the particular role indicator being discussed.
2.1.1.1. AGENT.

The WRU roles which correspond to the Agent case role include two separate indicators which differentiate between organizations acting as Agents and persons or groups of persons acting as Agents.

1. KIB (organization involved) 'is prefixed to a name of an organization...' From the example we can see that, although no mention is made of the organization as the instigator of an action, the correspondence seems to be quite clear.

   (1) an aluminum fabricating plant is being built at Alicante by - (KIB) the Aluminio Iberico, S.A. Agent

2. KEP (person involved) 'is prefixed to either the name of a person or persons considered as a group such as workers, unions, management...'

   (2) (KEP) steel producers - are shipping steel. Agent

   (3) (KEP) Elihu Thompson - recognized the Agent principle of welding.

We can see from these examples that the two role indicators described are used to indicate nominal elements which would be considered to be Agents in Fillmore's model.
2.1.1.2. INSTRUMENT.

The WRU roles which correspond to the Instrument case role include the following:

3. KAL (influenced by) 'is prefixed to any term that indicates a material, process, etc. which influences a preceding property...' as in:

   (4) different magnetic moments are present depending on

   (KAL) the type of atomic distribution.

Instrument

This role indicator corresponds most closely to the Instrument role when it is occupied by a sentence.

4. KQJ (by means of) 'is prefixed to a term that designates the actual agent or instrument effecting or used to effect a process, testing technique or function.'

   (5) polishing with - (KQJ) abrasives.

   Instrument

   (6) corrosion of tankers - (KQJ) by sea water...

Instrument
2.1.1.3. OBJECT

The WRU role indicators which correspond to the Object case are as follows:

5. KEJ (material processed) "is prefixed to a term which designates a material that is being acted upon or is being made to perform an action.

   (7) (KEJ) Fe-Ni-Al alloys - are cooled.
   Object

   (8) corrosion of - (KEJ) tankers.
   Object

6. KAP (property influenced) "is prefixed to a term that designates a property stated to be influenced by, dependent upon, variable with, or a function of another factor...."

   (9) (KAP) critical shear stress -
   Object

   is dependent on temperature.
2.1.1.4. SOURCE and GOAL.

The role indicators which correspond to the Source and Goal case roles include:

7. KAJ (starting material) 'is prefixed to a term that designates a material which is subjected to processes of fabrication...'

(10) The production of aircraft components –
(KAJ) from Al alloys...
Source

8. KWJ (product) 'is prefixed to a term to indicate that the term designates the material product of some process...'

(11) (KWJ) Pa metal – was obtained by reduction
Goal
of PaF₄ with metallic...

9 and 10. KWB (direction from) and KWC (direction to) are 'prefixed to terms that designate the place from which and to which a movement occurs.

(12) the emission of electrons –
(KWB) from the source – (KWC) to the target.
Source Goal
11. KIT (time occurring) 'is prefixed to a term that designates a date or inclusive dates....'

(13) butt-welding - (KIT) from 1877 to the present.

Source Goal

As can be seen in the above definition, KIT is also used to designate TIME alone.
2.1.1.5. LOCATION.

The role indicators which corresponds to the Location case role are:

12. KIG (location-geographic) which 'is prefixed to a place name when it is necessary to designate a geographic location' as in:

   (14) smelting in - (KIG) Spain.

13. KIS (location) which 'is prefixed to a term that designates where a process occurs... This role indicator is not used for geographic locations.'

   (15) casting - (KIS) in the foundry.
As is shown in the preceding examples, the WRU indexing system incorporates the following case roles:

1. Agent,
2. Instrument,
3. Object,
4. Source: as earlier times, earlier states, or earlier locations,
5. Goal: as later time, later state, or later location,
6. Location, and
7. Time.

However, more than one role indicator is included under many of these case roles—a result of the fine distinctions made in this system.

The most important feature of this system which differentiates it from the systems which follow is that the role indicators are assigned to individual keyterms which are then entered singly. As Hutchins notes, "...role indicators are entered separately in index files. The relational function of roles is evident only when all the terms descriptive of a particular document are brought together, either in the process of indexing or in searching (1975, p. 97)." The result, as noted by Aitchison, is that in at least one instance, "...Even if role indicators had been used in the search programme, they would almost certainly have included [the same roles, but
not in the appropriate relationship between the desired terms] so that they would not have excluded the document... (196, p. 42)." The point is that the role indicators specify the role or roles a concept assumes in the context of the entire document and not in relation to other specified keyterms. The next two systems preserve the relations between keyterms.
2.1.2. The PREserved Context Indexing System (PRECIS)

PRECIS was created by Derek Austin for use in the British National Bibliography. PRECIS is a string indexing system - i.e., an indexing system which preserves context, in this case the relations between concepts, by stringing together keywords in phrases which maintain the relations between keyterms by such devices as prepositions and punctuation. These strings, although normalized expressions, are obviously much closer to natural language expressions than single keyterms connected by a Boolean AND or even another string indexing system (Farradane's Relational Indexing System) which requires the use of symbols to indicate functional relations between terms.

The feature that Austin claims differentiates his system from other string indexing systems and which, therefore, makes PRECIS most suitable for use in indexes in languages other than English, is the use of case grammar based role operators which correspond to language universals - i.e., language characteristics which are common to all languages - such as nouns, verbs, questions, etc. The issue of whether or not Fillmore's case roles, or any other case grammar based roles, are truly linguistic universals is not at issue here. But Austin's choice of roles to be used to express the functional relations
between key terms in a string is, of course, of interest here because of the language behavior patterns described in this study.

A thorough analysis of the role operators in PRECIS was carried out by Michell (1978). The analysis carried out here does not alter Michell's description or conclusions except that Fillmore's 1971 case definitions will be used in this analysis.

The PRECIS role operators are based on a case system proposed by Johansen. According to Sorensen (1976, p. 112), the following sentence exemplifies Johansen's case role definitions as:

With a hammer, John built his brother
Instrument Agent Secondary Object

a shed from corrugated iron, in the garden
Primary Object Preliminary Object Place

yesterday.
Time

These roles appear to be equivalent to Fillmore's 1971 roles as follows:

Johansen  Fillmore
Agent  Agent
Instrument  Instrument
Preliminary Object  Source
Primary Object  Goal
Secondary Object  Object or Beneficiary
Place  Location
Time  Time
However, in Austin's subsequent description of his interpretations of several of these roles and their application in the indexing scheme, it becomes apparent that several of the role operators have been defined in ways that are not consistent with their definitions in the case grammar.

The initial comparison of Austin's role operators and Johansen's case roles is as follows:

<table>
<thead>
<tr>
<th>Austin: role operators</th>
<th>Johansen: cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0) Location</td>
<td>Place</td>
</tr>
<tr>
<td>(1) Key system</td>
<td>Agent/Preliminary Object/Primary Object</td>
</tr>
<tr>
<td>(2) Action/Effect</td>
<td>Primary Object/Actions</td>
</tr>
<tr>
<td>(3) Agent</td>
<td>Agent/Instrument/Secondary Object</td>
</tr>
<tr>
<td>(4) Viewpoint as form</td>
<td>xxxxxx</td>
</tr>
<tr>
<td>(5) Sample population</td>
<td>xxxxxx</td>
</tr>
<tr>
<td>(6) Target/form</td>
<td>xxxxxx</td>
</tr>
<tr>
<td>(d) Date as difference</td>
<td>Place/Time</td>
</tr>
</tbody>
</table>

The definitions of the role operators and the examples of the application of the role operators provided by Austin do not agree, however, with the definitions implied in his analysis of Sentence 1.

The first departure from case roles, which by definition are applied to noun phrases only, is in the definition of role operator 2, Action, which is actually used to describe the main verbal element in an indexing string. Since verbal elements cannot assume a case role in a
sentence, this operator cannot be considered to represent a deep case. Of the remaining role operators, roles 4, 5, and 6 also do not conform to true case roles. Role 4 is used to identify the viewpoint being discussed - e.g., democracy as viewed by a marxist. Role 5 is used to identify the sample population or the place in which a study was conducted. Role 6 is assigned to such characteristics of documents as bibliography, maps, anthology, etc. As Austin says, the concepts represented by these role operators are not part of the topic sentence.

The four remaining role operators do correspond to Fillmore's and Johansen's deep case roles. They are:

Role 1, key system, which corresponds to Agent of an intransitive action or Object of a transitive action. The following examples provided by Austin illustrate these two uses of Role 1:

(1) Prospecting for mineral deposits by government contractors

Role (1) mineral deposits
Role (2) prospecting
Role (3) government contractors.
This example shows the use of Role 1 as Object of the action - i.e., mineral deposits would be the Object in Fillmore's model as in:

(1a) Prospecting for - mineral deposits
Object
by government contractors.
(2) the migration of birds.

(1) birds

(2) migration

This example shows the use of Role 1 as the Agent of the action. The correspondence to Fillmore's cases is:

(2a) migration - of birds.

Agent

(3) Government regulation of capital expenditures by local authorities.

(1) local authorities

(2) expenditure of capital

(2) regulation

(3) government

In this example, we have two phrases, one of which, expenditures by local authorities, is the Object of the verb regulation. The correspondence with Fillmore's case roles would be:

(3a) Government - regulation - of expenditure

of capital by local authorities.

Object

local authorities - expend - capital.

Agent

Therefore, in this example, local authorities is both an Agent and an Object.
Another typical example of the use of Role 1 for the Agent, but this time with a transitive verb, is:

(4) Israel imports from the United States.

(1) Israel
(2) imports from
(3) the United States

(5) The United States exports to Israel.

(1) the United States
(2) exports to
(3) Israel

The roles assigned in this example will be discussed further after the discussion of Role 3. But at this point it can be said that Israel is the Agent in the first string and the United States is the Agent in the second string. This assignment pattern is typical of certain kinds of strings which will be described more fully under Role 3. The assignment of roles with verbs such as import and export is a problem for case grammar, since the importing party is also the later location, or Goal, and the exporting party is also the earlier location, or Source. However, for our purposes it will be assumed that the importing country must initiate the Action and is, therefore, the Agent, and the exporting country must also initiate the Action and is, therefore, the Agent.
Role 2, Action, was applied to some noun phrases — i.e., as opposed to the main verbal element — in several similarly structured examples in the PRECIS Manual (1974). One of the typical examples is:

(6) Planning medical research.

(2) Medicine

(p) research (p = part-whole relations)

sub(2) Medical research (indicates this phrase should be used when planning assumes the first position in the indexing string)

(2) planning

In this example it would appear at first that the assignment of the roles should more closely correspond to the assignments in the previous examples — i.e., Medical research is a noun phrase which is affected by (i.e., is the Object of) the main verbal element planning. However, medical research can also be seen as an activity which would be consistent with Austin's interpretation. That is, an activity — planning — affects another activity — medical research.

Role 3, Agent, is assigned to the nominal element which would assume the Agent or the Instrument role in Fillmore's grammar as in:
(7) Applications of queuing theory in scheduling of bus services.

(1) bus services
(2) scheduling
(3) queuing theory

where queuing theory is an Instrument.

(8) Prospecting for mineral deposits by government contractors.

(1) mineral deposits
(2) prospecting
(3) government contractors

where government contractors would be assigned the Agent role in Fillmore's grammar.

Role 3 is also used for nominal elements which would be assigned the Source or the Goal case under Fillmore's grammar as in:

(9) The United States exports to Israel

(1) the United States
(2) exports to
(3) Israel

where Israel would be the Goal, or later location.

(10) Israel imports from the United States

(1) Israel
(2) imports from
(3) the United States
where the United States would be the Source, or earlier location.

Role 0, Location, is assigned to the Location or environment in which the nominal element exists unless the location is the boundary on the study sample, in which case the nominal is assigned Role 5, Study Region or Sample Population.

Operator (d), date as difference, is assigned to nominal elements which would assume the Time case role in Fillmore's grammar.

The correspondences between Austin's roles and the case roles in Fillmore's grammar can be summarized as:

<table>
<thead>
<tr>
<th>Austin</th>
<th>Fillmore</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0) Location</td>
<td>Location</td>
</tr>
<tr>
<td>(1) as Object</td>
<td>Object</td>
</tr>
<tr>
<td>(1) as Agent</td>
<td>Agent</td>
</tr>
<tr>
<td>(2) Action</td>
<td></td>
</tr>
<tr>
<td>(3) as Agent</td>
<td>Agent</td>
</tr>
<tr>
<td>(3) as Instrument</td>
<td>Instrument</td>
</tr>
<tr>
<td>(3) as Source</td>
<td>Source</td>
</tr>
<tr>
<td>(3) as Goal</td>
<td>Goal</td>
</tr>
<tr>
<td>(d) Time</td>
<td>Time</td>
</tr>
</tbody>
</table>
In summary, even though Austin actually used the term 'case roles' to cover verbal elements and descriptions of form (e.g., bibliographies, manuals, etc.) which do not explicate functional relations between terms, his is the only system (in English at any rate) which attempts to apply linguistic theory to indexing in the bibliographic retrieval environment. First attempts do not need to be perfect. The most important contribution of this system is that an attempt was made to incorporate a linguistically based language model into indexing practice.
2.1.3. Farradane's Relational Indexing System (FRIS)

Farradane's Relational Indexing System (FRIS) is also a string indexing system - i.e., a system which 'strings together related terms' - which uses relational indicators to specify the functional relations between key terms in a string. The organizing principle underlying the structure of the relational system is a model of the psychology of human thinking and memory which is purported to reflect 'universals' in the association of concepts in the human mind. The truth of the claim that the relations included are based on a valid psychological model is not at issue here because, in any case, the relations are applied to language expressions. The discussion of this system will include a description of the relational indicators which are analogous to case roles only, since generic or attributive relations are not of concern here. The relations and their definitions have changed over the years as Farradane developed the system and attempted to clear up difficulties in application of the relations. The relations given in "Relational indexing, Parts I and II" (1980) are used in this discussion.

As will become readily apparent, Farradane's system provides descriptions of subjects which are very much like the query statements in this study. That is, one or more nominal elements is related to a particular verbal element
which describes the nature of the action or process affecting or being affected by the nominal element(s). Before giving the relations it is necessary to provide the rules and assumptions underlying this complex system. According to Farradane, "Relational indexing...is a means of expressing relations on a basis of the mechanisms of thought, to be converted directly into indexing notation. The subject concepts involved may be very different from case to case, but the framework of possible relations is limited and invariant, and thus bypasses the subject field and linguistic problems, and can be applied at any level of detail or complexity (1980, p. 267ff)." Farradane then describes the system of nine relators which he believes describe the universal mechanisms whereby humans relate concepts in their minds, rejecting completely any direct comparison between his relations and similar relations in linguistic descriptions. However, in his discussion of his system, Farradane must give examples which are language expressions and his examples will be extended to Fillmore's case grammar relations to determine analogies between the two systems.

The preliminary rules given below provide a description of the assumptions and decisions about language expression which Farradane considered to underlie the assignment of relations and the indexing of key terms and phrases. The central assumptions in this system are:
1. "A relation exists between two concepts when some meaning is implied between them in the mind (1980, p. 269)."

2. "...nine relations (together with their possible negation) have been found to be sufficient to express meaning in all subject fields. In terms of the basic relations, the mechanisms of our thought are not different in different subject fields (1980, p. 269)."

3. "If word meanings are clear, the expression of relations will also be clear (1980, p. 313)."

4. "The ease of indexing any complex subject will depend very much on the clarity of expression, in ordinary language, of the subject statement to be indexed. It will also depend on the accuracy of knowledge expressed in the subject (1980, p. 273)."

From these statements we can assume that the relations between concepts described by Farradane must be expressed in 'ordinary language' before indexing can be accomplished and that his examples can be used to draw possible analogies between his relations and deep case relations, which are also abstract relations. Point 4 is clearly a recognition by Farradane that world knowledge is required for interpretation of the relations between keyterms, a point which will be discussed under the section on Courrier's study.
The definitions of the relations given below were taken directly from Farradane (1980). The punctuation marks used by him to indicate the relator assigned to a relation will be used in the examples since many examples contain more than one relation and it is not always clear which relation is being discussed when the comparison is made between the FRIS relations and the case relations without the indicators. The relations in FRIS which are analogous to the deep case relations in Fillmore's grammar include:

1. Relation 4: Self-activity /*. This relation is used to express the intransitive verb situation... this relation must not be used with a transitive verb... An example of this relation is:

   (1) man/*walking
   Agent  Action

This relations is used, then, to specify the relation between an Agent and the Action instigated by the Agent. This relation is also used 'to express the dative case of 'to' (or the ablative 'from') as in:

   (2) children/*French/*teaching

   Teaching  French  to children
   Action  Object  Goal

As is discussed fully in Chapter 5, there are two ways of interpreting the relation between 'teaching' and 'French'.

Note: The punctuation mark following the '/' is the relational indicator.
the person(s) being taught. Although the deep case role assignments given above are in disagreement with the assignments given in the later example, they are just as valid since the later decision was an arbitrary one.

And, finally, 'the relation can also mean 'through'... e.g. 'flowing through a pipe'

(3) flowing/*pipe

        flowing through a pipe
         Action     Path

This relation is used, therefore, for the relation between an Action and the Path along which the Action proceeds.

2. Relation 5. Dimension /+. This relation 'expresses position in space or time. In other words, this relation is equivalent to the Location and Time case roles.

3. Relation 6. Action/--. This relation 'is used for any thing or operation acting on, or affecting, another thing or action...' as in

(4) clothes/--detergent

which corresponds to the case relation between the Instrument (detergent) and the Object (clothes);
(5) water/-purifying
which corresponds to the deep structure case
relation between an Action-Process verb
(purifying) and the Object of the verb (water);
4. Relation 7. Association /:/ 'This expresses
various forms of association...the relation of an
agent of, or tool for, an operation as in:
(6) etching/;acid
which corresponds to the case grammar relation between
an Action-Process verb (etching) and the Instrument
which is the immediate cause of the action described
by the verb.
5. Relation 9. Functional dependence /:- "This
expresses the relation of one thing causing or
producing something, which is also describable as a
product arising out of an initial thing as in:
(7) wheat#: bread
which corresponds to the relations Source and Goal,
or earlier and later states where wheat is the
Source nominal and bread assumes the Goal role.
An interesting example provided by Faradane which demon-
strates the kinds of distinctions his system is designed
to provide can be compared to the distinctions which case
grammar roles and verb types also provide. The phrase
a milk jug can be interpreted in several different ways,
each of which describes a different deep structure relation between milk and jug. The following compares the two systems handling of these distinctions.

1. use of a jug to contain milk. This phrase would come under the Association /; relation in FRIS - i.e., a Tool, the jug, and its use, containing milk. This is analogous to the case grammar relations:

   (8) (Someone uses) - a jug - to contain -
   Agent Instrument Action/Process
   milk.
   Object

2. a jug of (containing) milk. This phrase is considered by Farradane to constitute a whole/part relation which is a semantic, not a functional, relation and is not included in case grammar. The case grammar equivalent for the relations specified in this phrase could be:

   (9) The jug - contains - milk.
   Location State Object

That is, in this instance, the phrase can be considered to constitute a single concept - i.e., a jug of milk - which describes a State relation between the two nouns.

3. milk in a jug, which Farradane says is a Dimension relation denoting the Location of the milk. The case grammar analysis would be:

   The milk - is (in) - the jug.
   Object State Location
4. milk jug, which Farradane does not note, is simply a phrase describing what kind of jug is being discussed.

The ambiguity of the phrase a milk jug noted by Farradane is the result of the fact that this phrase simply does not provide enough contextual information upon which to resolve the ambiguity - i.e., to determine exactly which relationship between jug and milk is intended. The important point here is that Farradane's system can be used to specify the relationship between the two nouns and that the specifications provided by the relations have parallels with those provided by the use of case grammar.

Farradane's system incorporates the following case relations in Fillmore's grammar:

- Fillmore: Instrument, Agent, Source, Goal, Action, Dimension, Dimension, Self-activity

In addition, Farradane distinguishes explicitly between the three verb types Action, Process, and Action-Process.
It must be stated, however, that there is no real one-to-one correspondence between the case roles and Farradane's relational indicators, but that the relational indicator listed covers one or more aspects of the corresponding case role. For example, no one relation covers the appearance of an Agent nominal with an Action-Process verb because Farradane restricts the use of the relation between an agent-type noun and a verb to intransitive verbs. The result is that there is no relation that covers the deep case relation between an Agent type noun and an Action-Process verb. Similarly, with the Goal case, Farradane's relations only cover the earlier state and later state aspects of the Goal role. These qualifications and restrictions should be kept in mind when looking at the list of correspondences between the two systems.
In summary, these three indexing systems are most interesting from the perspective of their differences and the viewpoints taken by their creators of the principles underlying the relations between keyterms and the context in which they are used. All three systems contain abstract relations which correspond to one degree or another with the abstract relations in Fillmore's case grammar. But the notions and assumptions underlying them are quite different.

The Western Reserve University system, while it covers all of the aspects of all of the case roles in case grammar, uses a large number of role indicators but does not provide any means for indicating the relations holding between specific keyterms. PRECIS does not cover as many aspects of the case roles and does not indicate the relations between keyterms directly but does preserve these relations through use of language devices such as prepositions and punctuation, or syntactic devices. It also is the first attempt to incorporate case grammar roles into an indexing system. Farradane's relational system is the only one which indicates directly the nature of the relationship between keyterms, but the relations preserved do not include many of the aspects of the case roles in Fillmore's grammar.
While the strengths and difficulties inherent in each of these systems are not discussed here, each of these systems provides an interesting example of the thinking about language and language patterns over the years in traditional indexing in information science. When the WRU system was created, the abstract grammars of case had not yet been developed for English. Farradane and Austin both reached outside the field of information science to disciplines and theories which they felt had something to contribute to information science. Each has made an important contribution to the thinking about the nature of language and of the use of language to describe topics of interest in the bibliographic retrieval environment. The most interesting conclusion for the study here is the correspondence between the notions underlying the relations or roles in these systems and case grammar. It would seem that there is a small and possibly finite set of abstract relations which could be used to improve the performance of bibliographic retrieval systems. But which relations and the manner in which they should be indicated must still be determined.
2.2. A Study of Case Grammar in Information Science.

Yves Courrier (1978) tested the consistency with which subjects were able to assign case roles and verb types to noun and verb phrases in sentences in abstracts. Courrier used a 4X4 matrix model of verb types developed by Cook (1979) which provides a cross-classification for the four verb types — i.e., Action, State, Process, and Action-Process — and four classes of case frames — i.e., Basic, Experiencer, Benefactive and Locative. This model divides the list of case roles into two categories:

1. essential cases — i.e., Agent and Object — and
2. nonessential cases — i.e., Experiencer, Benefactive and Locative.

The essential cases are said to be directly related to the selectional features of the verb. For example, Action and Action-Process verbs require an Agent; State, Process, and Action-Process verbs require an Object. According to this model, the cases Instrument, Source, Goal and Time are omitted from the model because they are peripheral to the case frame. However, in the analyses of the sentences, the subjects were allowed to include these cases if they so desired.

According to Courrier, the most serious problems encountered by the subjects which were directly related to the matrix model included:
1. differentiation between State and Process verbs;
2. differentiation between the Benefactive and Experiencer case roles; and,
3. assignment of the Locative and Instrument case roles.

These difficulties arose because of problems in the interpretation of the definitions of the verb types and case roles and, because of difficulties in the interpretation of the proper 'sense' of individual verbs and nouns. For example, there was disagreement among the subjects about whether knowing is a State or a Process verb. Another example provided by Courrier involves the Benefactive and Experiencer roles and the interpretation of the intended scope of these roles, and the definition of a verb. For the sentence The cliff dominates the city, Courrier assigned the following relations:

(1) The cliff — dominates — the city.
    Agent       Action       Experiencer

That is, Courrier decided that dominates is a psychological verb and requires an Experiencer. Many of his subjects, on the other hand, noted that the Experiencer role is confined to the class of nouns which are capable of experiencing a mental state or psychological event. The city did not, in their opinion, fall into the class of nouns which are capable of experiencing a psychological event; at least not in this instance. The disagreement
also concerned the interpretation of the verb *dominates* since the sentence seems to be referring to a metaphorical use of the word - i.e., passive physical domination as opposed to active psychological or physical domination. Courrier suggests that the problem arises from the nature of the French language and its use of the same word in different contexts and with slightly altered meanings. The argument centered, in any case, on whether a verb has one true 'sense' or whether specification of the intended meaning of the verb is provided by the surrounding context. Acceptance of the latter means that the assignment of a verb type to the same verb in different sentences is context dependent and, therefore, the same verb can assume different relations with different nominal elements.

While the above two examples do not cover all of the problems Courrier encountered, they point up the two central issues for the application of any case grammar model to real-life language patterns - i.e., the interpretation of the meaning and scope of the verb types and case roles themselves, and the role of lexical features of words in the determination of the case role or verb type they may assume. In addition, the following potential sources of difficulty can also be identified in Courrier's study:

1. the choice of case roles and verb types;
2. the definitions of the case roles and verb types;
3. the extension of the definitions of relations to complex text; and

4. consistent interpretation of case roles and verb types during the analysis of statements in texts.

Unfortunately, not much direct information about the definitions which should be used can be drawn from Courrier's study because the study undertaken here uses a different set of roles and verb types. However, the conclusion can be made that instructions and definitions must be very consistent across language analysts so that the source of any difficulty can be clearly identified.

The study by Courrier does suggest that explication of functional relations, most specifically in Farradane's system and the WRU system, may be more difficult than was assumed by the creators of these systems. The problems encountered by Courrier and his subjects involved world knowledge and the interpretation of the meaning of language expressions at an abstract level which was not explained by the structure of the sentences, by the context surrounding the sentences, or by the meaning of the words appearing in the sentences. This problem must be addressed before any system using functional relations indicators can be developed which can be used with confidence by indexers and searchers.
CHAPTER 3

METHODOLOGY
(General)

In most aspects the methodology used in this study followed standard procedures for bibliographic retrieval system experiments and performance evaluations, although the searches were conducted solely to provide a valid sample upon which to perform the analysis required by the language behavior being investigated. General procedures, common assumptions, and problems in the control of variables in this environment have been widely discussed, so it was considered that bibliographic retrieval system evaluation guidelines would provide a suitable framework for data collection. In addition, the language behavior investigation conducted here was motivated by a desire to improve the understanding of underlying language structures and behavior involved in this environment in the hope that this understanding may be used to improve retrieval system performance.

The descriptions of the test environment and considerations involved in any research using this environment are drawn from general conclusions in major works on bibliographic retrieval system research including classical studies (Cleverdon, 1966; Aitcheson, 1963; Lancaster, 1974; Keen, 1975; Western Reserve University, 1968), general texts (Lancaster, 1978; Salton, 1983) and
general reviews of research (Spark-Jones, 1981) and of research methodology (Tague, 1981). The approach used in construction of the operational definitions of the elements - i.e., parameters and variables - in the test environment and the characteristics of these elements and of the relationship among them was consistent with general conclusions drawn by these authors.
3.0. The Hypotheses

The following hypotheses were tested in this study:

1. that there is a significant relationship between
   a) the functional relations between keyterms in query statements and in abstracts, and
   b) the aboutness decisions of subject specialists, secondary judges, the researcher (acting as first language analyst) and a second language analyst;

2. that there is a significant relationship between the level of subject knowledge and the consistency of aboutness decisions in relation to the functional relations model; and

3. that there would be a significant relation between the functional relations assigned to keyterms in queries and in abstracts by different language analysts.

The first hypothesis is the main hypothesis for this study and tests directly the consistency of language behavior in aboutness decision-making and the ability of the functional relations model to describe the consistencies which occur. While it is assumed that there will be some variation in the level of agreement between the
different aboutness judges and the functional relations models, a high level of consistency in language behavior in aboutness decision-making for each group of judges and both language analysts is required to support the hypothesis.

The second hypothesis was designed to test for interaction between subject knowledge and consistency of language behavior in aboutness decisions, and is designed to determine whether or not persons other than the inquirer use language patterns - i.e., functional relations - when trying to interpret the intentions and anticipate the decisions of the inquirers based on textual information alone - i.e., the transcripts of the query negotiations and the texts of the abstracts. If there is strong evidence of such language behavior patterns in the decisions of the secondary judges, there is the possibility that some means of exploiting this natural language behavior could be devised for use in query negotiation and search strategy formation.

The third hypothesis addresses the consistency with which the functional relations can be assigned to keyterms in queries and abstracts by different persons using the definitions provided. Evidence of internal consistency in assignments would indicate that functional relations can be identified and specified and are, therefore, viable
as a description of language patterns in texts. Disagreement between the two analysts as to which functional relations definitions should be applied to which nominal or verbal elements would indicate that the definitions employed are inadequate or otherwise faulty.

The methodology employed, operational definitions of parameters and variables, and procedures for each of the three hypotheses will be discussed in order. Section 1 describes the test environment and the methodology employed in the collection of the data - i.e., the queries and abstracts - used in testing the hypotheses. Section 2 describes the test of the hypothesis involving the subject specialists, Section 3 describes the tests of the hypotheses involving the secondary judges, and Section 4 describes the tests of the hypothèses involving the second language analyst.
3.1. The Environment for Testing the Hypotheses.

Any complete test involving a bibliographic retrieval system involves three components:

1. a collection of documents or document representatives - i.e., the data base system and the data bases,
2. a set of search requests, or queries, and
3. relevance assessments.

The operational definitions of each of these components for each experiment are influenced by the researcher's assumptions about the effect of each definition on the validity, reliability, and generalizability of the conclusions of the study and by constraints imposed by the experimental environment.

The parameters, variables, and procedures involved in the collection of data for this study are discussed in detail in the following sections. Briefly, they are:

1. Parameters and variables:
   a) the data base system,
   b) the data bases used for searching,
   c) the subject specialists (the first group of aboutness judges),
   d) the searchers,
   e) the queries,
f) the format of printed results,
g) the second group of aboutness judges, and
h) the second language analyst.

2. Procedures:
a) obtaining queries and determining which queries were suitable for inclusion in the final test set,
b) assignment of queries to searchers,
c) query negotiation and search strategy formation,
d) construction of the test sample of abstracts to be given to the subject specialists for decision,
e) subject specialists' decision-making procedure,
f) construction of the test sample for testing the consistency of functional relations assignments by the second linguist,
g) the second linguist's decision-making procedures,
h) procedure for obtaining the aboutness decisions of the second group of aboutness judges, and
i) the data analysis procedures for all the tests.
A special concern throughout the construction of the test designs was consideration of the sources of potential bias due to language factors. For example, it is assumed generally that there is a difference in precision of meaning and use of terminology across various disciplines. That is, it is widely assumed that use of language and terminology in the hard sciences is in some way more consistent, precise, and specific than use of language and terminology in the social sciences and the humanities. While this assumption was not tested directly in this project, the implications of the assumption were potentially important for generalizability of any results from this study. Therefore, the sample contains representatives from various disciplines within the sciences, the social sciences, and the humanities. Other questions of bias are discussed under the procedure where they are most relevant.
3.1.1. Parameters and Variables.

3.1.1.1. The Data Base System.

The first decision that must be made in a test involving bibliographic retrieval systems is what kind of system will be used. A commercial on-line machine-readable data base was used in this study for several reasons:

1. the results of searches could be obtained immediately;
2. use of Boolean operators made it possible to link two, three, or more groups of keyterms and their equivalents so that each record retrieved was known to contain at least one member from each keyterm group;
3. the abstracts required for the aboutness decisions could be obtained immediately and at relatively reasonable cost; and
4. the various formats available for printing the records retrieved simplified data organization.

Another very important capability of the on-line bibliographic retrieval system is the ability to use keyterm searching of all fields - i.e., free-text searching. For example, one keyterm presented in a query, Computer Literacy, appeared in one form or another in
over fifty abstracts before it was included in the list of descriptors. Had the searcher used the descriptor only, s/he would have retrieved only the abstracts received after the descriptor had been included in the index.

Another important consideration is whether to use one data base system or more than one, unless systems themselves are being compared, of course. It was decided in this study to use only one data base system so that the experience of the searchers and the effect of their experience on their knowledge of the command language of the system would not affect their performance and thus introduce an unwanted source of variation. In general, however, the choice of the data base system should not affect validity, reliability or generalizability of results so long as the system is large and contains a very general and widely used set of data bases.

The DIALOG data base system was chosen from among the several systems available for the following reasons:

1. DIALOG is a very large system and contains data bases for all major subject disciplines, many special subjects and many areas of interdisciplinary research.

2. The searchers were most familiar with the command language of the DIALOG system.
3. DIALOG includes on-line thesauri for some data bases and this feature was considered desirable by the searchers.

4. Special education rates on most of the data bases in the system reduced the overall cost of the project considerably.

The research question required that abstracts be included in the document representatives used for the aboutness judgements. The DIALOG data system contains data bases which meet this requirement for all major disciplines. Therefore, it was decided that only one system need be used and that use of one system would eliminate any question of the effect of differences among systems or among searcher's abilities to search various systems.
3.1.2. The Data Bases.

Considerations involved in choosing and using the data bases are:

1. the characteristics of the collection of document representatives such as subject coverage and contents - e.g., abstracts, subject descriptors, index phrases, etc. - and
2. the characteristics of the retrieval mechanism - e.g., search aids provided by specific data bases such as on-line or printed thesauri or other search aids, instructional materials, index language structures, retrieval limitation devices, etc.

In general, use of a wide range of subjects and data bases improves generalizability of results and randomization of selection procedures is essential. The searchers in this study, therefore, were encouraged to use all relevant data bases which provided abstracts or references to sources of abstracts - e.g., Dissertation Abstracts International.
3.1.1.3. The Subject Specialists (Inquirers).

In most tests the persons who submit queries to be searched also make the necessary relevance decisions. The characteristics of the inquirers who submit queries are determined in large part by the research environment and the corresponding population of potential participants. A researcher may choose to accept anyone who submits a suitable query or may set some minimum level of academic or professional achievement. In general, fairly expert participants are preferred where the inquirers will also serve as the relevance judges, since it has been established that the level of subject knowledge affects relevance judgements (Saracevic, 1975, p. 341). At the same time, the inquirers who are accepted as subjects for a test should be representative of the population of potential participants, or of the subset of the population meeting specific criteria.

The best test of the hypothesis for this study required subject specialists who were highly cognizant of the terminology of their disciplines and who, therefore, would best recognize the senses and referents of terms appearing in the records retrieved in response to their queries. Consequently, faculty members and advanced graduate students were considered to be the most appropriate subjects available in this situation.
The subject specialists who participated in this study were drawn from the faculty, doctoral students, researchers, and advanced Masters student populations at the University of Western Ontario (27 subjects), the University of Guelph (2 subjects), and the University of Montreal (1 subject).
3.1.1.4 Searchers.

Considerations when choosing the searchers for a study using the bibliographic retrieval system environment include experience, training, subject expertise, and attitudinal variables. In tests where none of these characteristics constitute an independent variable, the researcher should establish some mean standard for each of the characteristics listed.

The searchers who participated in this study were selected from among the students at the School of Library and Information Science, the University of Western Ontario, who had completed the on-line reference section in an advanced reference course, had performed on-line searches in two or more courses, and wished to participate in the study. Criteria for selection of the searchers were:

1. subject background;
2. professors' evaluations of demonstrated or perceived ability in query negotiation, search strategy formation, and on-line searching; and
3. the number and variety of searches already performed by the searcher prior to the test.

Subject background was an important consideration when choosing searchers because a wide variety of subjects were to be searched. The four searchers chosen each had
knowledge of one or more general subject areas - i.e., sciences, social sciences, history, literature and education - and searches were assigned to them strictly on the basis of subject background. This was done, of course, to eliminate variation in searcher performance across subject areas.

The training and experience in on-line searching each searcher had received had been fairly consistent, but still varied enough to eliminate any bias which might have resulted if their training and experience had been too idiosyncratic. And, finally, the searchers had all indicated their eagerness to participate in the study, their main motive being the real-life experience their participation would provide.

To ensure that the searchers would be competent in query negotiation and search strategy formation, they were given further instruction. The instruction in query negotiation was provided by an expert in the area and not by the researcher, so the instruction was based on general principles and was not directed toward the purposes of this study. The searchers were also given some further instruction in search strategy formation and in the characteristics of the database system, the data bases, and the search aids which were available for their use. Because searcher performance was not a test variable, all searchers received the same information prior to searching.
3.1.1.5. Queries.

In general, tests from which inferences will be made about queries require the collection of a set of real queries presented by real users to an operational system. Some randomization process may be included in the query collection design to eliminate potential bias. All queries presented to a system within a certain period of time or across various shorter periods of time, for example, constitute a block sample which may or may not be random. But with such a sample, one can introduce some element of chance so as to satisfy the conditions of inference. On the other hand, where the research question requires some special restrictions on query type, as in this study, the sample may be unbiased if every query of the type required has an equal chance of being included in the sample. In this study all queries which met the requirements described below were searched and efforts were made to inform as many people as possible of the availability of the searches to any one meeting the requirements for level of academic achievement and query characteristics. The required query characteristics included:

1. Each query had to contain at least two key phrases which assumed different semantic roles and were in a direct syntactic relationship which could be expressed in a sentence or proposition.
2. All queries had to be in the English language.
3. Queries could be about any subject and could be searched on any relevant data base(s) which provided abstracts. However, certain types of queries did not provide a sufficient test of the hypothesis and, therefore, were rejected. For example, queries which referenced a particular author and subject were not considered to be appropriate even though it could be argued that they met the requirements for selection. However, a query about an author as a concept - e.g., Henry James - in relation to some other concept - e.g., some influence or a particular technique or style - was considered suitable.

Queries had to be broad enough to retrieve enough citations to provide a sufficient test of the hypothesis. The number required was a minimum of six retrieved records containing abstracts.

This list of criteria was used in the initial acceptance of a query as a potential member of the final test set. The procedures for obtaining queries and for selecting those queries which would comprise the final test set are described in the next section.
The first consideration when setting conditions and procedures for query negotiations is the medium of communication. Options are:

1. personal face-to-face interviews between the searcher and the inquirer;
2. telephone interviews between the searcher and the inquirer; and
3. queries received by mail with no interaction between the searcher and the inquirer.

Lancaster (1974) found that the medium in which the query was conveyed had a significant effect on retrieval performance evaluations, but he did not explore reasons for the variation. The researcher may choose to use one or more mediums - e.g., both personal interviews and telephone interviews consist of verbal communication, so both types might be used. If more than one medium is used, randomization of mediums across the set of queries is required. In addition, the researcher should decide at this point whether some form describing the negotiation is to be filled out or whether a more informal approach will be used.

The two mediums allowing direct verbal communication between the searcher and the inquirer were considered preferable for this study, of course, because the interviews were taped.
3.1.1.6. The Format of the Printed Results.

The searchers were instructed to print the final retrieved set of document representatives in two formats. One format included accession number, title, and abstract only. This format was used for all aboutness judgements so that considerations of author, source, language (in a very few cases where the abstract was in English but the article was in another language), age, or assigned descriptors would not influence the decisions of the subject specialists. The other format included all other information provided in the record except the abstract. The subject specialists were sent copies of the total record for all document representatives retrieved for their query after their aboutness decisions had been received. The total record was also used in the analysis of the data.
3.1.2. Procedural Decisions.

The researcher must make decisions about how the data collection procedures will be carried out. Among the most important procedures in this study were:

a) the call for queries,
b) the assignment of queries to searchers,
c) query negotiation, search strategy formation, and searching procedures,
d) output formats, and
e) relevance judgements.

Each of these decisions can affect test results and determine the validity and reliability of any conclusions drawn (Tagae, 1981).

The method by which queries are elicited should ensure an unbiased sample. Therefore, all potentially qualified inquirers should have an equal opportunity to learn of the study and to respond if they are interested. Public notices were used in this study to inform potential inquirers. These notices were placed in all libraries on the campus and in the graduate student center. This method informed qualified persons from many areas of subject knowledge as demonstrated by the wide variety of subject areas represented among the queries.

The assignment of queries to searchers was affected, of course, by subject content. But because there was some
overlap in the subject background of the searchers, the
assignments were divided fairly evenly among all the
searchers. This was done as a safeguard against the
learning affect that might have occurred had one searcher
performed a significantly larger proportion of the
searches, a real possibility given the relatively little
real life experience in on-line searching these searchers
had obtained prior to the test.

The following discussion describes the details of
the procedures that were carried out to obtain the titles
and abstracts to be judged and the judgements themselves.
General considerations were discussed in the previous
section.
3.1.2.1. Obtaining Queries.

The queries used in this experiment were obtained from faculty and graduate students (see 3.1.1.4.). Notices were sent to a cross-section of subject libraries including:

a) the social sciences and humanities library,
b) the natural sciences library,
c) the health sciences library,
d) the business library, and
e) the education library.

In addition, a notice was posted in the graduate students' social center. (The law library was not included because legal data bases are not included in the system used in this study.) The notices listed the educational level required and interested parties were requested to contact the researcher directly.
3.1.2.2. Assignment of Queries to Searchers.

After the initial contact between the researcher and a subject specialist, the query was assigned to the searcher who had the appropriate subject background. It was then the responsibility of the searcher to contact the subject specialist, conduct the query negotiation, construct the search strategy, select the appropriate data bases, and conduct the actual search and print the resulting retrieved set of document representatives in the required formats.
3.1.2.3. Query Negotiation and Search Strategy Formation.

The major departures from real-life on-line reference practice occurred at the query negotiation and search strategy formation stages. First, the environment in which the query negotiations were performed was not the normal library setting, but the offices of the subject specialists or an office reserved for this purpose at S.L.I.S. In addition, the interviews were tape-recorded which is not a normal practice in libraries. Second, the exhaustivity and/or specificity of the search strategies was to be determined in part by the requirements of the study, a condition which a formal professional reference service could not accommodate. (In fact, search strategy formation was directed toward high recall and the searches included as many quasi-synonyms as the subject specialists wanted.) Because of these special conditions and because of the additional effort required of the persons who acted as subject specialists, the searches were provided free of charge.

The tape recording of the query negotiations was a necessary condition for participation of the subject specialists for this study. The query negotiations were tape-recorded so that the semantic factors involved in the query which were not obvious in the final written statement or in the search strategy could be determined, if necessary
during the analysis of the data. In addition, a further subtest required that other persons be able to read the entire transcript of the query negotiation. To protect the privacy of the subject specialists, the tapes were identified by a code meaningful only to the researcher and the original list of subject specialists' names and the codes assigned to them was destroyed when the research was completed. Also, because the queries often represented highly original research interests, the full results of the searches were given only to the subject specialists and search strategies were erased.

The subject specialists were not aware of the nature of the semantic relations being tested because such knowledge might have biased their evaluation of the search results. However, at the same time, the query negotiations had to provide as much information as possible about the nature of the functional relations between the key-terms in the query. Therefore, the searchers were instructed to obtain as much background information as possible and to encourage the subject specialists to explain terms and talk about the relations between concepts. The recording of each query negotiation also provided a check for any indication that the subject specialists had been inadvertently provided with information which might bias their relevance decisions. (No such
problems were detected and were, in fact, unlikely since the searchers also were not aware of the nature of the study."

(NOTE: As required by the University of Western Ontario Committee on Research Involving Human Subjects, all potential subject specialists were required to sign a release stating that they understood the potential restrictions on the services they would receive and that they knew that the query negotiation was being tape recorded and others would hear the recording.)

(See APPENDIX I for an example of the release form.)
Before the interviews were conducted, the subject specialists were informed of test requirements and of restrictions on searches and output from the searches to ensure that s/he would still want to participate. A description of the method by which their privacy was to be protected was also fully explained at that time.

Another important consideration at this stage of the data collection is who will control the query negotiation and search processes. It was considered to be essential in this study that the researcher should step in only in case of emergency and then only to provide information which would not bias the test results. For example, the searchers were helped with an unfamiliar data base as far as commands and the general structure of indexes or access files were concerned, but at no time were terms to be entered in the search strategy suggested.

The conditions and procedures for search strategy formation and data base selection were carefully explicated for the searchers. Each searcher received the same instructions to ensure consistency. The searchers were also informed that abstracts were required when printing out the retrieved document representatives so no data base which did not provide abstracts should be searched (with the exception of Dissertation Abstracts International which provides citations to abstracts.)
3.1.2.4. Construction of the Test Sets of Abstracts.

After the searches were completed and the abstracts retrieved had been received from the searchers, the following procedures were observed in the construction of each query/abstracts set:

1. All document representatives which did not contain complete abstracts were removed from the test set (unless they provided a reference to an abstract, as noted before).

2. Where a language limit had not been possible during the search, all abstracts which were not in the English language were removed from the test set.

3. The set was further limited to those abstracts which contained words which were considered to be related in sense or reference to the terms in the search strategy - i.e., broader terms, narrower terms or related terms - and had been so indicated by the indexing language, and homonyms of key terms.

4. Where two different abstracts for the same title were retrieved on two different data bases, both of the abstracts were included in the test as a matter of interest.

5. If the number of document representatives remaining was fewer than forty, but no fewer than six, the entire set was included in the test set to be presented to the subject specialist for aboutness
judgements. If more than forty abstracts remained, a test set of thirty-five was extracted. To ensure that the sample chosen was valid, the abstracts were divided into groups and examples from each group were chosen according to the following criteria:

a) Abstracts which contained keyterms which were the same as the keyterms entered in the search strategy and which contained the same functional relations between the keyterms as those in the query;

b) abstracts which contained keyterms which were the same as the keyterms entered in the search strategy and where the keyterms appeared to carry a different sense or reference - i.e., homonyms - and/or where the functional relations between the keyterms did not match the functional relations in the query;

c) abstracts which contained terms which were considered to be related to the terms in the search strategy in meaning and in the same functional relations as between keyterms in the query; and

d) abstracts which contained keyterms which were considered to be related in meaning to the terms in the query but not in the same functional relations as in the query.
Examples from all four groups were included in the final test set if all four types were represented. Special care was taken to ensure that the total variety of functional relations between keyterms was included in the final set as well. That is, multiple examples of each type of variation in sense, reference, and functional relations were desirable in order to obtain the best test of internal consistency in aboutness judgements.

6. Each abstract in the test set was assigned a code number and the numbers were entered on the form to be used by the subject specialist to indicate his/her decision for each member of the test set. (See APPENDIX II for an example of this form.)

7. The evaluation form and the test set of titles and abstracts were sent to the subject specialist for his/her decisions.
3.1.2.5. Relevance Judgements.

'Relevance' judgements are the criteria by which the results of most bibliographic retrieval experiments are measured. Problems in the definition of relevance are discussed in the Introduction, so only the decisions about procedure will be described here. The decisions involved in the relevance evaluation procedure are:

1. Who shall evaluate retrieval results?
2. How will relevance be defined?
3. What scale of relevance will be used?

These decisions depend largely on the research question and the purpose of the test. Since the test here was for language behavior patterns in the determination of match or not match between the query and abstracts, the inquirer was considered to be the best judge of this match. Two other groups of judges were also used to test the effect of subject expertise on the relevance decisions and on language behavior patterns in relevance decisions.

The definition of relevance in this context refers to the way in which relevance is defined in the instructions to the persons who will be making relevance assessments - e.g. 'utility', 'pertinence' or 'aboutness' (See Section 0.2.1.). For this test it was necessary to make it very clear to the relevance judges that 'aboutness' was the criterion on which the decisions were to be based. (For details see APPENDIX III.)
And, finally, the nature of the research question is also the most important factor in deciding upon the scale of relevance which will be used by the judges in making their decisions. Common scales are:

1. binary relevance - i.e., a simple 'yes' or 'no';
2. multi-valued relevance - i.e., three or more levels of assessment based on criteria such as 'not relevant'/somewhat relevant'/highly relevant', or 'not useful'/somewhat useful'/highly useful', etc.;
3. ranked relevance - i.e., records are ranked in order from most relevant or useful to least relevant or useful; and
4. weighted relevance - i.e., each member of the retrieved set is assigned a number, or weight, which represents the sum of the weights given to the concepts it contains which match the concepts contained in the query.

Simple binary relevance decisions were used in this study. Although the trend at present is to use degrees of relevance or match - i.e., alternatives 2, 3, and 4 above (Tague, 1981) - simple binary relevance was used in this study to maximize the difference between judgements and to provide the most rigorous test of the descriptive power of the functional relations model. That is, the
object was to measure a simple match or non-match between query statements and statements in abstracts. To introduce degrees of match at this point would have introduced extraneous language behaviors which could have obscured the information being sought.

All persons assessing relevance, of course, must receive the same instructions. These instructions must clearly specify the notion of relevance which is to be applied when making the decisions, the meaning of the scale, and the method of indicating the decisions. (See APPENDIX III for the instructions to the relevance judges.)

After the test set of abstracts retrieved for a query had been constructed, the abstracts, the instructions to be followed in aboutness decision-making and the form on which the subject specialist was to record his/her decisions were sent to the subject specialist. No other communication between the researcher and the subject specialist occurred until after the decisions had been returned.

The instructions to the subject specialist (APPENDIX III) were designed to define clearly the criteria by which the aboutness decisions were to be made without biasing the decisions in favor of the hypothesis. Therefore, the instructions do not indicate in any way the purpose of the research, nor do they introduce restrictions which would force the subject specialists to
make decisions favorable to the hypothesis. The decision form on which the subject specialists were asked to record their decisions (APPENDIX II) required only a simple 'yes' or 'no' answer for each abstract.

Upon receipt of the subject specialists' responses to the test set of abstracts, the subject specialist was given all records retrieved in response to his/her query. All subject specialists were given the opportunity to have their searches revised and re-searched if they were dissatisfied with the final results. None requested further searching and many expressed satisfaction with the results they had received. Part of this satisfaction was attributed to the fact that literature they had known of had been retrieved, so they were inclined to trust the results of the entire search. Another reason given by several subject specialists for this satisfaction was the lengthy and thorough nature of the query negotiations conducted by the searchers. The subject specialists felt that their queries had been well understood and this appeared to increase their confidence in the search results.
3.1.2.6. The Analysis Procedure.

A semantic analysis of each query and of each abstract was performed to explicate the functional relations between:

a) keyterms in the query, and

b) keyterms in each abstract.

The analysis and explication of functional relations in statements — i.e., propositions — in the queries and abstracts required the development of a descriptive format which could be applied consistently to all data. As pointed out earlier, this test assumes that abstracts which are judged to be about a query tend to contain a statement which resembles a statement in the query. Conversely, abstracts which are judged to be not about the query contain the same keyterms, or homonyms or quasi-synonyms, as the query but differ in some significant way from both the query and the abstracts judged to be about the query. In order to adequately express the difference between about and not about abstracts, the format had to provide for clear delineation of:

1. semantic categories such as generic relationships, use of a phrase which was considered to be equivalent to a keyterm, and related terms; and

2. the functional relation assumed by each nominal element and the verbal element in statements in queries and abstracts.
The descriptive format developed for this test is a semantic network diagram of functional relations (FIGURE 1).

Each nominal element in a query statement (and the quasi-synonymous terms for each element which, by definition, can assume the same role as the main keyterm) was listed under the label of the role it assumed in relation to the verbal element. The verb and its synonyms were listed under the Verbal Element label and the verb type was also indicated.
3.1.2.6.2. The Analysis of the Queries.

**Step 1.** The keyterms in the query were classed according to semantic relations. That is, synonymous, related, broader and/or narrower terms which the subject specialist had approved as equivalent expressions for a key concept were grouped. Any of the terms in a group could assume the same functional relations in a query statement - e.g., *parents* and *caretakers* were considered to be equivalent since they both assume the same function in the example which follows.

**Step 2.** The functional relations between the keyterms in the query statement were indicated using the semantic network diagram. Verbal elements often had not been specified as keyterms, but implied the nature of the desired verbal frame - e.g., caused, affected, influenced, etc. All denote an Action-Process verb.

The example of a query analysis which follows was taken from a earlier experiment and consists of a single query statement for purposes of clarity. As will be seen in the next section, the actual analysis was much more complicated.

The analysis proceeded as follows:
EXAMPLE 1. Description of a Query.

Query: What effect does the caretaking figure and/or the living environment have on language development in the mentally retarded?

Agent
(CARETAKER)
(PARENTS)
(TEACHERS)

Verb
(Action/ Process)
(AFFECTS)
(INHIBITS)
(AIDED)
ETC.

Instrument
(LIVING ENVIRONMENT)
(HOME)
(INSTITUTION)

Experiencer
(MENTALLY RETARDED)
(MENTALLY HANDICAPPED)

Object
(LANGUAGE DEVELOPMENT)
(VERBAL SKILLS)

Note: For purposes of clarity, the typefaces used to represent the various components of the diagram will be:

**Xxxxxxxxx** indicates the propositional category - i.e., the verb of the case role.

(CAPITALS) indicates the keyterm in the query and synonymous or related terms indicated by the subject specialist.
As can be seen in the semantic diagram:

a) AFFECTS, the verbal element, is an Action/Process verb - i.e., it describes a change in the Object nominal element and requires an Agent and/or an Instrument nominal element as the instigator of the action or the direct cause of the effect;

b) CARETAKER (or PARENTS or TEACHER) is the Agent;

c) LIVING ENVIRONMENT (or HOME or INSTITUTION) is the Instrument;

d) LANGUAGE DEVELOPMENT (or VERBAL SKILLS) is the Object, or thing affected; and

e) MENTALLY RETARDED (persons) is the Experiencer.
3.1.2.6.3. The Analysis of the Abstracts.

The analysis of the abstracts followed essentially the same procedures as the analysis of the query. However, as will be seen in the examples, an important difference was that, while all of the abstracts contained the same keywords or related terms as the query statements, the fact that some abstracts were 'not about' the query indicated that something in the meaning of the keyterms or the functional relations between keyterms must be different from those in the query in some significant way. Determination of the nature of these differences was, of course, one of the most important objectives of the analysis.

**Step 1.** The abstracts were examined and keyterms, equivalent expressions, or homonyms which matched those in the query were underlined and listed;

**Step 2.** The propositions containing the keyterms, equivalent phrases, or homonyms were extracted from the abstracts;

**Step 3.** Where necessary, the functional relations between keyterms, equivalent phrases, or homonyms were clarified by reduction of several statements into one statement which preserved the context in which the word had been used;
Step 4. The functional relations between keyterms, equivalent phrases or homonyms were explicated using the semantic network diagram. Where there was no direct functional relationship between keyterms or homonyms, usually in an abstract judged to be not about the query, further explanation was provided.

EXAMPLE 2a describes the analysis of an abstract judged to be about the query described in EXAMPLE 1. The procedure used for reduction of several sentences to one statement which preserved the meaning relations between the keyterms is also presented.
EXAMPLE 2a. (About)

Abstract.

Surrogate and natural parent comparisons between institutional and non-institutional children.

Rated the surrogate parents of 20-8-13 yr old institutionalized educable mental retardates and the natural parents of a matched sample of 20 noninstitutionalized EMRs on Hollingshead's Two-Factor Index of Social Position. Results indicate that the institutional surrogate parents had significantly higher socioeconomic status. The present and previous findings suggest that this higher social position may have a positive influence on language habilitation among institutional EMRs.

Reduction:

Rated surrogate parents of institutionalized mental retardates;
natural parents of noninstitutionalized mental retardates.

(higher social status) influences language habilitation among institutionalized mental retardates.

parents influence language habilitation among mental retardates.

Clearly, the reduced statement, and even various statements in the abstract, are semantically and even syntactically very similar to the query statement. The functional and meaning relations can be expressed as:

(See EXAMPLE 2b.)
EXAMPLE 2b. Description of About Abstract.

Agent
(PARENTS)
"parents"

Verbal Element
(Action/Process)
(AFFECTS)
"influenced"

Experiencer
(MENTALLY RETARDED)
"mental retardates"

Object
(LANGUAGE DEVELOPMENT)
"language habilitation"

Although the statement in the abstract does not identically match the statement in the query, the functional relations between keyterms are identical to those in the query. Also, the meanings of the keyterms in the abstract are sufficiently similar for the aboutness judge to accept them as equivalent. That is:

a) INFLUENCE, the verbal element, is an Action/Process verb and is equivalent to AFFECTS;

b) PARENTS is the Agent;
c) LANGUAGE HABILITATION is the Object and is equivalent to LANGUAGE DEVELOPMENT; and
d) MENTAL RETARDATES is the Experiencer.

The effect of the INSTITUTION is not discussed, so the Instrument role remains unfilled. However, the abstract was judged to be about the query because either the Agent or the Instrument role was required, not both.

The next abstract, EXAMPLE 3, was also retrieved in response to the query. It contains the same, or equivalent, keyterms or phrases as both the query and the previously described about abstract. However, it was judged to be not about the topic named in the query.
EXAMPLE 3a. (Not About)

Abstract.

Increasing mentally retarded adolescents' verbalizations about current events.

Examined the effects of antecedent and consequent events on the verbal behavior of 3 institutionalized mentally retarded male adolescents. Verbal statements related to current national and international events were recorded after exposure to TV news programs. The accuracy of verbalizations was examined as a function of: (a) exposures to TV news presentations in massed form (i.e., viewing the entire news program before an opportunity to describe it) vs distributed form (i.e., viewing each news item separately with each followed by an opportunity to describe it), and (b) contingent tokens and social praise for correct verbal responses (i.e., statements corresponding to news items presented). Both the temporal distribution of news presentations and the reinforcement procedures improved the accuracy of verbal statements by the subjects.

Reduction:

- events (have effect on) verbal behavior of institutionalized mental retardates;
- distribution of news presentations improved accuracy of verbal statements by the subjects;
- news presentations improved verbal behavior of mentally retarded.

The reduced statement in this abstract is not equivalent to the query statement even though it contains the same, or equivalent, keyterms as those in the query statement. The description of the functional relations is as follows:
EXAMPLE 3b. Description of a Not About Abstract.

It is clear in this diagram that an important function has been changed. The summary of the analysis is:

a) IMPROVES, the verbal element, is an Action/Process verb and is equivalent to AFFECTS;

b) MENTAL RETARDATES assumes the Experiencer role; and
c) VERBAL RESPONSE assumes the Object relation and is a quasi-synonym for LANGUAGE DEVELOPMENT.

These three functional relations are the same as those for both the query and the about abstract. The significant difference between this abstract and the abstract judged to be about the query is:

d) TV NEWS PRESENTATION (distribution of) is the Instrument in relation to the verb - i.e., is the thing which directly affects VERBAL RESPONSE: and

e) INSTITUTION has assumed the Location relation.

Therefore, this abstract does not contain a statement which is equivalent to the statement in the query in functional relations even though it contains all of the concepts named in the query.

An analysis such as the one described above was performed for each of the thirty query/abstracts sets which comprised the data for this study. As will be seen in the more detailed description which follows in the next section, the general pattern of the analysis described above is consistent across the data - i.e., the abstracts judged to be not about the query by the subject specialists had a significant difference in one or more functional relations from those specified in the query.
3.2. The Test of the Secondary Judges' Aboutness Decisions.

The purpose of this test, as stated in the introduction to this chapter, was to determine the effect of subject knowledge on language patterns in aboutness decision-making. The results of this second test of the ability of functional relations to explain aboutness decision-making are important for determining whether or not functional relations might be useful in systems where someone other than the inquirer will perform system searches and the initial evaluation of search results. This test was designed to determine whether or not the language-behavior pattern evidenced by the subject specialists is also evidenced by other persons. If so, there is a strong indication that this language-behavior pattern should be further examined for means to exploit this natural language behavior in the bibliographic retrieval system environment.
3.2.1. The Secondary Judges.

The persons who acted as secondary aboutness judges were recruited from a beginning Reference class at the School of Library and Information Science, The University of Western Ontario. This class was chosen because the students had just completed a series of classes on query negotiation and on-line searching of bibliographic retrieval systems, and it was considered that this encounter with the test environment provided a background which would help them understand the task they were required to perform. Because the subjects were not to be told the purpose of the exercise until after they had completed their decisions, the exercise was easily represented as a reasonable part of their course work.

There were two measures of the performance of the secondary judges involved in this part of the test of the functional relations model. The first measure simply tested the decisions of the secondary judges against the functional relations based on the query statements used by the subject specialists. This first test, then, measured the ability of the secondary judges to interpret the intended query statements of the subject specialist on the basis of the text of the query negotiations. The second measure tested the internal consistency of the language behavior of the secondary judges' aboutness
decisions and was based on functional relations models which reflected their own interpretations of the query statements they had assumed had been intended by the subject specialists.

In addition to the primary test involving functional relations, a secondary test to measure the affect of subject knowledge on language behavior was introduced. Each secondary judge was given a query for which s/he had some formal education background and a query for which s/he had no formal education background or even informal personal interest. The two measures of language behavior were applied to both sets of judgements.
3.2.2. The Test Procedure.

Each student in the S.L.I.S. class was asked to fill out a subject background profile (See APPENDIX IV). From the original set of 47 students, 30 were chosen who had as much experience as possible in one subject represented in the queries on the one hand, and as little experience as possible in another subject represented in the queries on the other. For example, a student with a B.Sc. in Biology and no courses or real interest in Literature would make aboutness decisions for a query/abstracts set in Biological Science as an Experienced judge and for a query/abstracts set in Literature as a Not Experienced judge. None of the students were aware that a test was being conducted and none were aware of whether or not they had been included in the test even if they had learned of the experiment.

The test was performed during class time to ensure that each student would spend approximately the same amount of time making their assessments and to ensure that they followed the instructions. The students were asked to refrain from discussing the queries or the instructions for making aboutness assessments among themselves. Questions about the instructions were dealt with before the query/abstracts sets were distributed. The students were asked to perform the following procedure:
1. read each query negotiation carefully and determine the query statement or query statements which they felt the subject specialist considered to be necessary for a match in the abstracts;
2. read the letter of instruction given to the subject specialists and use these same instructions in making aboutness decisions (see APPENDIX III); and
3. indicate on the decisions form (see APPENDIX II) whether they thought each abstract was about or not about the query.

The decisions of the secondary judges were then analyzed for agreement with the subject specialists' decisions and with decisions based on functional relations.
3.3. The Test of the Reliability of the Functional Relations Model.

The purpose of the test of the reliability of the functional relations model was twofold. The main purpose was to test the reliability of the assignment of functional relations to nominal elements and verbal elements in the query statements and abstracts. The test also served to test the reliability of the researcher's interpretation of the contents of abstracts. The linguist who participated in this part of the study had a background in linguistics and had studied under Fillmore during the period the case grammar model was being developed. His familiarity with the principles underlying the model was considered to be highly desirable.
3.3.1. Construction of the Test Sets.

The construction of the test sets the second language analyst would make decisions for proceeded as follows:

1. Query statements. The query statements used by the subject specialists in making their relevance decisions were listed. Except where reduction of several statements had been necessary, the query statements were given in the form in which they had appeared in the query. Where reductions had been necessary, the reduced form only was included in the test list.

2. Abstracts. A random sample of abstracts was chosen using random numbers generated by a computer program. The whole text of each abstract was included in the test set.
3.3.2. Instructions to the Language Analyst.

The language analyst was given a description of the case roles and verb types and an example of the semantic network diagram. (See Chapter 3, p. 133 for an example of the diagram.) The descriptions of the case roles and verb types were taken almost directly from the text of the definitions provided in Chapter 1 - i.e., Fillmore's definitions - so the second language analyst would be using the same information when making his decisions as had been used by the researcher in the initial approach to the analysis of the data. Where alterations occurred, they were for purposes of textual cohesion and did not describe or attempt to remedy any problems encountered by the researcher in the analysis.

The second language analyst was instructed to:

1. read each query statement and describe the functional relations between the nominal element(s) and the verbal element in each abstract;
2. read each abstract and describe the functional relations between the keyterms found in the abstract; and
3. to indicate whether any of the query statements were matched by statements in the abstract.

(See Appendix VI for the full text of the instructions and the full set of query statements.)
CHAPTER 4
METHODOLOGY
(Analysis)

The purpose of this chapter is to describe the procedures actually followed in the analysis of the data. Section 1 of this chapter provides the definitions which were actually used in the assignment of functional relations. These definitions are consistent with the definitions provided by Fillmore, but the extension of Fillmore's examples to the key-terms in queries and abstracts required some further development of his definitions. The examples given in this section are drawn from the data to more clearly demonstrate how the functional relations were assigned to the vocabulary and language structures which occurred in the data.

Section 2 describes the analysis of one query and set of abstracts in some detail and of another set of abstracts more briefly. The purpose of this section is to demonstrate how the analysis was performed and how the aboutness decisions of the subject specialists were used to determine which query statements contained in the query were used to determine which query statements had been used by the subject specialists when making their aboutness decisions.
4.1.0. The Definition of the Functional Relations.

The definitions of the verb types and case roles given in Chapter 2 were drawn entirely from Fillmore. While his definitions have not been altered in any significant way in this section, the works of Chafe (1970) and Cook (1979) are drawn upon in the description of the verb types because Fillmore emphasized the case roles and his definitions of the verb types are very brief. Chafe and Cook, on the other hand, emphasized the verbal frame in their discussions of case grammar. However, all three of these authors agreed on four basic verb types and on the basic characteristics of each verb type.

The definitions of the case roles are based entirely on the considerations raised by Fillmore's discussion of the characteristics of these roles in his 1971 article, "Some problems for case grammar". The application of these roles to the data required that certain decisions be made about the characteristics of the nouns which could assume a particular case role.
4.1.1. The Verb Types.

The most important extension of Fillmore's original definitions was the inclusion of required case roles for each of the verb types. The definitions of the verb types are determined by the kind of state or activity they describe and the kind of relationships that exist between the verbal element and nominal elements in a statement. Verbs describe either a state or an activity. An activity can include an Action, a Process, or an Action and a Process. The definitions of the four verb types are as follows:

1. **State** verbs describe the condition or state of a nominal element as in:

   (1) Populist movements - existed - Object State
during the Depression.

   **Time**

   A State type verb requires an Object which is in the state or condition described by the verb.

2. **Action** verbs describe an activity which does not result in a change in the condition of a nominal element as in:

   (2) Henry James - used - Agent Action

   *Impressionistic techniques.*

   **Object**

   An Action verb requires an Agent who instigates the action.
3. Process verbs describe an activity which results in a change in the state or condition of a nominal element, an Object, but not by the direct action of another nominal element as in:

(3) Populist movements - arise - Object during economic Depressions.
                  Process Time

4. Action-Process verbs describe both an Action and a Process as in:

(4) Corona virus - causes - Instrument Action/Process
demyelinating disease - in mice.
                  Goal Object

Action-Process verbs require both an Agent who instigates the action and/or an Instrument which causes the change, and an Object which is changed or otherwise affected by the Process described by the verb and/or a Goal which results from the Process described by the verb.

The verb type assignment is determined by the lexical features of the verb and by the world knowledge of the person assigning the verb type. Clues as to the nature of the activity or state described by the verb are also provided by the perceived relationship between the nominal elements in the sentence and the verbal element. This
point will be of central importance in the discussion of the kinds of relationship between nominal elements and the role of the verbal element in defining these relationships in Chapter 5.
4.1.2. The Case Roles.

The assignment of case roles to nominal elements in query and abstract statements also required clarification of the definitions of the roles used in this model. Fillmore's definitions of the cases and his discussion of the nature of the nouns which could assume particular case roles provided strong guidelines, but the nouns which appeared in many of the query statements and abstracts posed problems in interpretation of the nature and scope of the definitions. These problems are discussed more fully in Chapter 5. The definitions used during the analysis are given below.

1. **Agent.** The basic definition of an Agent is 'the typically animate perceived instigator of the Action described by the verb' as in:

   (5) Henry James - used - 
   Agent                Action

   Impressionistic techniques. 
   Object

   This case was also considered to include:

   a) aggregates of animate beings which were considered to constitute an entity which acted as an Agent, as in:

   (6) The literary establishment - excludes 
   Agent.                Action

   women authors. 
   Object
b) institutions or formal organizations which were considered to constitute an entity which could act as an Agent as in:

(7) A community information center — Agent
automated — social information files.
Action/Process  Object

c) organisms which instigated an Action as in:

(8) Nematodes — migrate — Agent  Action
in animal tissues: Path

d) inanimate objects which act in ways associated with human capabilities as in:

(9) The computer — chooses — Agent  Action
the most appropriate programs... Object

2. Instrument. The basic definition of the Instrument role was 'the case of the inanimate force or object causally involved in the Action or State identified by the verb' as in:

(10) A university [uses] — a computer — Agent  Instrument
to handle — student records. Action  Object
Fillmore extended the definition of the Instrument case to include natural force nouns as in:

(11) Leaf characteristics - hinder Instrument Action/Process
      the growth of rust. Object

3. Experiencer. According to Fillmore, 'where there is a genuine psychological event or a mental state verb we have the Experiencer as in:

(12) Social class - affects Instrument Action/process Object
      the attitudes of People. Experiencer

4. Object. The Object role is assigned to the nominal element which is affected or changed by the Action or Process identified by the verb, or which is in the State described by the verb. In other words, the Object case can occur with any of the four verb types as in:

(13) Leaf characteristics - are associated with- Object State
disease resistance. Object

(14) A Liberal/Conservative Scale - measures Instrument Action
      political attitudes. Object

(15) Populist movements - arise Object Process
during economic Depressions. Time
(16) Superovulation affects fertility. 
   Instrument Action/Process Object

5. **Source** and **Goal**. Depending on the type of verb, the Source and Goal cases are assigned to:
   a) earlier and later locations with verbs of motion;
   b) earlier and later points in time with verbs of motion, Process verbs, or Action-Process verbs; and
   c) earlier and later states with Process verbs or Action-Process verbs.

   Examples of these related case roles were not common in the data so only one can be offered as follows:

   (17) Nematodes migrate in rabbits
   Agent Action Location

   from the point of entry to the ankle.
   Source Goal

6. The **Goal** case alone can be assigned as follows:
   a) where there is a transfer or movement of something to a person, the receiver as destination assumes the Goal role;
   b) where there is an end result or later state with Action-Process or Process verbs, the result or state assumes the Goal role; and
   c) with verbs of impingement the thing which moves is the Object and the thing to which it moves, or upon which it impinges, assumes the Goal role.
Examples of these three uses of the Goal role are given in Chapter 5 because there are problems inherent in the definition of this role which require some explanation.

7. Path. The Path case role occurs with verbs of motion and is assigned to the nominal elements which describe the course upon which the movement occurs as in:

(18) Nematodes — migrate — through the hepatic portal system — in crows.

8. Location. The Location role was assigned not only to places but also to animate entities when these entities were the environment in which an event occurred, as in Sentence (18) above.
4.1.3. The Assignment of Functional Relations.

The rules given below describe the procedure followed when determining the functional relations between keyterms required by the subject specialists (or secondary judges) for an abstract to be judged to be about the query.

**Rule 1.** A relationship between two or more keyterms (or their quasi-synonyms) had to be stated in the query for the relationship to be considered to be a query statement. Alternative relations were accepted under the conditions of the next rule.

**Rule 2.** The following relationship between keyterms were accepted if the subject specialists' decisions indicated that these alternatives were considered to be about the query as in:

a) If a causal relation between Term A and Term B was stated but a simple direct association was accepted consistently, the statement of simple association was accepted as a query statement. Statements of simple association usually contained a state type verb.

b) If a simple association between Term A and Term B was stated but a causal relationship was accepted consistently as about in the abstracts, the causal relation was accepted as a query statement.
Rule 3. In order for a statement in a query - or an alternative statement as described in Rule 2 - to be considered to be a query statement, the majority of abstracts in which it appeared had to be accepted as about by the subject specialist or the secondary judge.

Rule 4. If a query statement or alternative statement appeared in only one abstract, the decision of the subject specialist or the secondary judge was accepted as given.

Rule 5. If a functional relation was consistently accepted as about except where a particular quasi-synonym appeared as a keyterm in the relation, the rejection was considered to be due to word meaning and not to inconsistency in judgement by the judge.

Rule 6. Where a negative statement of relation - i.e., a statement saying Term A does not affect Term B or Term A is not associated with Term B - appeared in an abstract, the decision of the subject specialist or the secondary judge was accepted as given even though, according to Putnam (1958), such a statement is also about the topic stated in the query.
4.2.0. Analysis of Functional Relations in Aboutness Decisions.

The purpose of this section is to describe in detail the analysis of two representative sets of queries, abstracts and aboutness decisions for language behavior patterns involving functional relations. The actual procedures followed during the analysis of the data were:

1. Extraction of all potential query statements* from the query** and assignment of functional relations to the keyterms in each statement;
2. analysis of each abstract to determine whether or not it contained one or more statements which matched one or more query statements;
3. comparison between the aboutness decisions made by the subject specialist and the statements in the query and abstracts to determine:
   a) which query statements had actually been used in making the decisions, and
   b) how consistent the subject specialists were in their decisions - i.e., were.

* A query statement is defined here as a single statement which indicates desired relationships between two or more keyterms.

** A query is defined here as the total description of the topic of interest provided by an inquirer.
abstracts which contained certain statements accepted consistently and, conversely, were abstracts which did not contain these statements rejected consistently.

The first task in the analysis of aboutness decision-making involved determination of which of the many potential query statements in a query were actually used by the subject specialist in making decision. The previous example of a single sentence query statement was pretty straightforward because the single statement clearly indicated the relations between the keyterms. The queries in the data for this study, on the other hand, consisted of lengthy explanations of intended meanings of keyterms and/or relations between keyterms. Therefore, it was first necessary to extract from the query all of the statements containing keyterms which the subject specialist had indicated should be used in the search strategy. This list of potential query statements was then compared with the retrieved abstracts to determine which statements also appeared in the abstracts. The reduced list of query statements was then used in the determination of which quasi-synonyms and functional relations between keyterms were required for an abstract to be considered to be about the query.
4.2.1. Analysis of the Queries.

The key portions of a query (Query 13) from which query statements were extracted are given below along with a description of how query statements were identified. This query was chosen because it contains some clear and fairly simple examples of considerations involved in determining query statements and because it was typical of the entire set of queries. (For the full text of this query see APPENDIX IV.) The other example, Query 28, includes only very brief sections because the purpose of the discussion involving this query is clarification of some of the problems encountered in the analysis.

The analysis proceeded as follows:

QUERY 13:

"I'm interested in Type A behavior. Type A behavior is a kind of behavior related to coronary heart disease. And the components of Type A behavior are things like hostility, speed and impatience—stress type things. I am particularly interested in a test of Type A behavior. The name of that test is the Jenkins Activity Survey. And, indeed, the whole focus of my study is the construction and validation of the Jenkins Activity Survey."

So far we have three potential topic statements:

(1a) Type A behavior is related to coronary heart disease;

(1b) the Jenkins Activity Survey tests Type A behavior; and
(1c) the construction and validation of the Jenkins Activity Survey.

"The validation tests on the Jenkins Activity Survey have been suggesting that another measure, the structured interview, is much better than the Jenkins Activity Survey. I think that if we go back and look at the construction of this measure that that is where the problem lies."

A fourth, and fairly complex potential query in this section is:

(1d) validation tests on the Jenkins Activity Survey suggest that the structured interview is much better than the Jenkins Activity Survey.

Also, "construction of (the Jenkins Activity Survey) is restated." There is also the suggestion that validation tests compare the structured interview and the Jenkins Activity Survey.

The next sentence does not provide any new query statements or concepts, but it does clarify what is meant by construction of a test.

"The test construction refers to the method that you use to come up with your items and that you use to examine the quality of those items."

The next section clarifies what is meant by validation of the test and further definition of the intended purpose of the test.
"The second and closely related aspect of test construction is test validation. To validate a test you simply engage in a study to determine whether the test measures what it's supposed to measure. When you look at the validity studies on Type A behavior in general, when they use the structured interview to predict coronary heart disease, they have a certain hit rate. When they use the Jenkins Activity Survey, which is supposed to measure Type A behavior, they have a much lower hit rate when they attempt to predict coronary heart disease. And I think the problem is that when they constructed the Jenkins Activity Survey they didn't construct it to predict heart disease. They constructed it to predict the structured interview - and the structured interview was originally constructed to predict heart disease."

This section implies again that validation studies compare the Jenkins Activity Survey and the structured interview, but there is no direct statement of this fact. However, there are several other potential query statements in this section:

(1e) the Jenkins Activity Survey measures Type A behavior; (Same as (1b))
(1f) the Jenkins Activity Survey predicts coronary heart disease; and
(1g) the Jenkins Activity Survey predicts the structured interview.

And, finally,

"...they take the Jenkins Activity Survey and the structured interview and say which is better at predicting coronary heart disease."

As can be seen in the above examples, often the subject specialist sums up the desired topics throughout the
interview with clear statements or sentences interspersed with explanations of concepts.

So far the following potential queries have been identified in Query 13:

1. a) Type A behavior is related to coronary heart disease;
   b) the Jenkins Activity Survey tests Type A behavior;
   c) the construction of the Jenkins Activity Survey;
   d) validation of the Jenkins Activity Survey;
   e) validation tests suggest that the structured interview is better than the Jenkins Activity Survey for.....
   f) the Jenkins Activity Survey predicts coronary heart disease;
   g) the Jenkins Activity Survey predicts the structured interview;
   h) they compare the ability of the Jenkins Activity Survey and the structured interview to predict heart disease.
4.2.1.1. Analysis of Functional Relations in the Queries.

Analysis of the linguistic characteristics of the query statements revealed a variety of structures. Some query statements contained only one verbal element and one nominal element and a match or not match in functional relations between the query and abstracts was determined by the type of verbal element and/or by the various functional relations the nominal element assumed in relation to the verbal element in an abstract. For example, the following query statements contained only one nominal element:

a) construction of the Jenkins Activity Survey.

Verbal Element
(Action/Process)
"construction"

Goal
"Jenkins Activity Survey"

b) validation of the Jenkins Activity Survey.

Verbal Element
(Action)
"validation"

Object
"Jenkins Activity Survey"

*denote functional relations based decisions as opposed to about and not about which denote the decisions of the subject specialists or the secondary judges.*
Other nominal elements required to complete the propositions given above such as Agent - e.g., a researcher - and/or an Instrument - e.g.; a method - were not required for a match. However, the Jenkins Activity Survey had to assume the functional relation with each of the verbal elements described in the examples for an abstract to be considered to be about the query by the subject specialist.

Other query statements were more complicated and involved several nominal elements:

c) Type A behavior is related to coronary heart disease. (The verb phrase here is more properly stated as 'related to (the development of)').

The implication intended by the inquirer who submitted this query statement was that Type A behavior in some way contributed to the development of coronary heart disease. But the statement itself simply describes an association with no causal relation indicated. Therefore, in this statement Type A behavior and coronary heart disease assume the same case role - Object, i.e.,

Verbal element

(State)

"related to"

Object

"Type A behavior"

"coronary heart disease"
The next query statement, while stated in a single sentence, contains two separate propositions:

d) The Jenkins Activity Survey was constructed to predict the structured interview.

i) constructed the Jenkins Activity Survey.

ii) the Jenkins Activity Survey predicts the structured interview.

Statement (di) is the same as statement (a), and statement (dii) is:

**Instrument**

"Jenkins Activity Survey"

**Verbal Element**

(Action)

"predicts"

**Object**

"structured interview"

The procedures just described, then, were followed in the analysis of all of the queries. The list of potential query statements was then used during the analysis of the abstracts as a guide to the relations between key terms that should appear in any abstracts judged to be about the query. The exact list of query statements actually used was determined when the decisions of the subject specialist were analyzed.
4.2.2. Analysis of the Abstracts.

The procedure for the analysis of the abstracts was essentially the same as that described for the analysis of the queries, with the exception that the analysis of the abstracts was guided by the query statements already identified and by the fact that it was known that some abstracts were about the query and some were not. To reiterate briefly, the procedure followed was:

1. Keyterms in each abstract were located and underlined.
2. The sentences containing keyterms, or quasi-synonyms for the keyterms, were extracted and, where necessary and/or possible, were reduced to a single sentence.
3. Functional relations were assigned to the keyterms in each statement.
4. The abstract statements were compared to the query statements to determine whether or not there was a match in functional relations.

The examples which follow were taken from the abstracts retrieved for Query 13.
The following abstract was included in the abstract set retrieved for Query 13 and was judged to be about the query by the subject specialist:

Title: Progress toward validation of a computer-scored test for the Type A coronary-prone behavior pattern.

Abstract: The Jenkins Activity Survey, a self-administered computer-scored questionnaire, was constructed to detect the coronary-prone behavior pattern. Item weighting was done to maximize its agreement with an interview criterion. 83 men who had had a heart attack scored significantly higher than 468 controls. They were more hard-driving, competitive, and responsible.

While the title says validation of a computer-scored test, the first sentence in the abstract names the Jenkins Activity Survey as the test validated. Also, the abstract includes the Jenkins Activity Survey detects coronary-prone behavior pattern (synonymous with Type A behavior) and item weighting is part of the construction of the Survey. Therefore, we have the following phrases which match query statements:

a) validation of the Jenkins Activity Survey;

b) construction of the Jenkins Activity Survey;

c) the Jenkins Activity Survey detects Type A coronary-prone behavior pattern.

A comparison of the list of propositions extracted from the abstract with the list of query statements shows that the three statements correspond to query statements
(d), (c) and (b) respectively. The abstract also contains a reference to agreement between the Jenkins Activity Survey and the structured interview, but there is no direct indication that the level of agreement was tested.

The next abstract was also included in the set of abstracts for Query 13:

Title: The structured interview and questionnaire methods of assessing coronary-prone behavior in male and female college students.

Abstract: Several self-report measures of type A behavior were compared with the Rosenman and Friedman structured interview method of assessment in male and female college students. The student version of the Jenkins Activity Survey was found to correlate weakly with the interview for both males and females, while scales derived from the Gough and Thurstone inventories showed moderate correlations with interview typing for both sexes. A brief scale derived from a recent reanalysis of the Framingham study correlated with the interview appreciably greater for females than for males. In all, these data suggest the need for caution in using only paper-and-pencil tests to assess type A behavior.

This abstract contains only one query statement match — i.e., measures (Jenkins Activity Survey) were compared with structured interview method of assessment and ... the Jenkins Activity Survey ... correlates weakly with the interview. The other abstracts judged to be about the query which contained references to a comparison between the Jenkins Activity Survey and the structured interview also contained verbs synonymous with comparison such as correlated with or agreed with.
The next abstract was also judged to be about the query:

**Title:** Assessment of conceptual tempo in the Type A (coronary prone) behavior pattern.

**Abstract:** Previous research has documented an association between a behavioral complex termed Type A, and an increased incidence of coronary heart disease. To further define some of the behavioral characteristics that distinguish Type A Ss from their noncoronary prone (Type B) counterparts, 65 female and 61 male undergraduates were administered the Matching Familiar Figures Test. While there were clear sex differences in performance, the results offer only qualified support for the notion that Type A's are more impulsive than Type B's. One component of Type A, Speed and Impatience, was related to increased impulsivity in females but not in males. Findings suggest the importance of considering the interaction of gender with behavioral subcomponents of the Type A behavior pattern.

This abstract contains the statement **association between** Type A behavioral complex and coronary heart disease, or:

- **Verbal Element**
  - (State)

  "is associated with"
  - (increased incidence of)

- **Object**
  - "Type A behavior"
  - "coronary heart disease"

which matches the query statement:
  - a) Type A behavior is related to coronary heart disease.
The abstract does not contain any other statement which matches a query statement, so we know that this query statement was used in making aboutness decisions.

The next abstract retrieved for Query 13 was judged to be not about the query:

Title: Habitual catecholamine excretion and its relation to habitual distress.

Abstract: To study individual differences in habitual subjective and physiological arousal a scale for measuring subjective arousal in the course of daily living was constructed and evaluated (the day-to-day stress scale, DTDSS). 30 male and 22 female medical and university students (aged 18-35 yrs) completed the DTDSS. Urine samples were obtained on 17 occasions from 20 of the male Ss over a 48 hr period to provide an estimate of free catecholamines. These Ss also completed the Eysenck Personality Inventory, Rotter's Internal-External Locus of Control Scale, and an abbreviated version of the Jenkins Activity Survey. Samples were collected at home during daily activities and night rest, as well as under laboratory conditions with alternated periods of activity and inactivity. Results indicate that (1) scores on the DTDSS correlated significantly with average adrenaline level, and (2) low physiological reactivity was indicative of neuroticism.

The following statements were extracted from this abstract:

a) ...a scale (DTDSS) was constructed;

b) ...a scale (DTDSS) was evaluated;

c) ...students...completed...the Jenkins Activity Survey.

This example contains no single proposition (even where reduced) which contains two or more keyterms from the
in the query statements. Thus there is no functional relationship between the verbs **constructed** and **validated** and the *Jenkins Activity Survey*, as shown in the following diagrams:

1) **Verbal Element**
   (Action/Process)
   "constructed" [Goal]
   "scale (DtDSS)"

2) **Verbal Element**
   (Action)
   "evaluated" [Object]
   "scale (DtDSS)"

In both of the above statements, the functional relations are the same as for the relevant abstracts and the query statements, but the **scale constructed** and **evaluated** is not the *Jenkins Activity Survey*. Therefore there is no match.

The next abstract, which was judged to be **not about** the query, is more interesting and provides some information about the subject specialist's aboutness decisions which was not obtained from the **about** abstracts concerning the topic intended by the subject specialist.
Title: Heritability of type A behavior.

Abstract: Ninety-three pairs of monozygotic and 97 pairs of dizygotic middle-aged American, male twins were studied to estimate the heritability of Type A behavior -- a coronary heart disease-prone behavior pattern. Participants were given the interview assessment of Type A behavior. In addition, they completed four psychological test batteries: the Thurstone Temperament Schedule, the Jenkins Activity Survey, the California Psychological Inventory, and the Gough Adjective Check List. Type A behavior measured by interview was found to be nonheritable. Psychological test scales which significantly correlated with Type A behavior, however, generally had significant heritability estimates. Only the Adjective Check List scales, which significantly correlated with Type A, showed nonsignificant heritability estimates. Selected items from these Adjective Check List scales may provide a method for assessment of Type A behavior possibly free from genetic influence.

The keyterms of interest in this abstract are scattered across many sentences and the relations between them really must be inferred from the surrounding sentences -- e.g., we do not know for certain whether the Jenkins Activity Survey was one of the tests that correlated with Type A behavior. But two sets of relations between keyterms are clearly stated and the decisions regarding these statements by the subject specialist and by the secondary judges -- i.e., that they were not about the query. Obviously, the estimation of both the Jenkins Activity Survey and the interview were compared, but the stated purpose of the study was to estimate the heritability of Type A behavior and the influence of this
variable on test results and not a direct comparison between the two tests. In addition, this abstract was rejected even though it clearly refers to use of the Jenkins Activity Survey to assess (test or measure) Type A behavior. Therefore, we can conclude that, since no abstract in the set which contained only this statement was accepted as about, the statement is not sufficient to be accepted as an intended query statement and can be stricken from the list of query statements used in making aboutness decisions. This conclusion is supported further by the decision on the next abstract (only a portion will quoted here).

Title: Life styles of myocardial infarct patients and of control groups: various similarities and differences.

Abstract: In a population sample of 'healthy' males the classifications on the Jenkins Activity Survey for assessing the Type A/B coronary-prone behavior pattern, and on the Maastricht Questionnaire, measuring emotional drain and subsequent depression, were investigated in relation to retrospective reporting of life changes in a newly developed Questionnaire...Type A subjects did report significantly more life changes on their work environment and family situation...

Obviously, the Jenkins Activity Survey is not the main subject of this abstract. Rather, this abstract is about life style measures and effects. However, the Jenkins Survey was used to assess Type A behavior. Now we can
say with some certainty that the query statement, the Jenkins Activity Survey tests Type A behavior was not accepted by the subject specialist as a sufficient statement for acceptance of an abstract as about the query.

The final analysis of all abstracts for Query 13 showed that three query statements were used consistently by the subject specialist for discriminating between about and not about abstracts as follows:

1. Validation of the Jenkins Activity Survey;
2. the Jenkins Activity Survey was compared to the structured interview; and
3. Type A behavior is related to heart disease.
A fourth query statement, construction of the Jenkins Activity Survey, appeared in two abstracts which were judged to be about the query, but both of these abstracts also included one of the three statements listed above. Therefore, there was no way of determining whether or not this statement would have been sufficient on its own for acceptance of an abstract as about by the subject specialist.

The subject specialist's decisions and the functional relations based decisions were in complete agreement for Query 13. This level of consistency was not demonstrated by all other subject specialists, however. The following description of the analysis of selected abstracts for
another query shows how inconsistencies in a subject specialist's decisions were determined using functional relations.

Two abstracts retrieved for Query 28 were judged to be not about by the subject specialist, but were judged to be a match using functional relations. The examination of these abstracts below shows that, while the statements which match query statements are rather weak examples of agreement, they seem to contain statements which closely resemble topic statements in the query which were otherwise considered to be sufficient for acceptance of all other abstracts in which they appeared as about by the subject specialist. Two abstracts were also accepted as about by the subject specialist which were rejected using functional relations. (The query statements and excerpts from the abstracts are given below.)

The query statements for Query 28 which are of interest here were:

1. Women's or girls' - search for - identity.
   Agent Action Object
2. Women's or girls' - (psychological) growth.
   Object Process
3. Feminist criticism of - children's or women's literature.
   Action Object

The following abstract was judged to be not about by the subject specialist, but a match using functional relations.
Title: Women's fantasies and feminist utopias.

Abstract: Eight feminist utopian works of fiction are examined; despite apparent dissimilarities, their utopias contain common features...Some common areas of attention include...growth rather than duality as the essential design of life.

Because the works discussed in this abstract are works of fiction (by literature the subject specialist intended to mean fiction) and the idea of growth in these works is one of the topics under discussion, this abstract was judged to be a match for Query 2 using functional relations.

The next abstract was also judged to be not about by the subject specialist, but a match using functional relations.

Abstract: "The Yellow Wallpaper," by Charlotte Perkins Gilman, is a remarkable piece of history and sociology, as well as a feminist story concerning the search for self.

This abstract also contains a match for a query statement - i.e., Query 1.

The next two abstracts were judged to be about the query by the subject specialist, but were judged to be not a match using functional relations.

Abstract: ...presents units for helping students explore the search for identity...the hero/heroine concept...

Abstract: ...Topics covered under fiction include books on...growing-up female...The nonfiction section lists books about...personal growth and development.

In the first abstract, the search for identity is not related to women or girls per se, and the hero/heroine
concept is not related to criticism of any sort. Therefore, the statements containing the keyterms in this abstract do not contain them in the relations specified in the query. The second abstract was rejected as not a match by the researcher because growing up female does not specify the kind of search for identity or idea of a quest that the subject specialist described. The second statement, personal growth and development is not functionally related to fiction or to girls or women.

Another example of disagreement between the subject specialist and the functional relations decisions occurred for Query 17. The query statement involved was:

   Agent  Action  Object  Time

The following abstract was judged to be about this query statement by the subject specialist, but was rejected as not a match using functional relations.

Abstract: Enrollments in agricultural education have risen sharply in recent years...After discussing southern black and white land grant colleges (founded in 1862 and 1890 respectively), the authors present the results of a large survey of agricultural students at those colleges. They found that students attracted to agriculture education include substantial proportions of females.

The functional relations between the keyterms in this abstract were:

Females - [attend] - /colleges - in recent years:
   Agent  Action  Object  Time
Colleges - were founded - in 1890 and 1862.
Goal Action/Process Time

That is, in this abstract 1862 and 1890 are functionally related to colleges were founded and not to females attend colleges. Three other abstracts contained almost identical statements, including the following:

Women - in 1830 and 1862 Land Grant Universities.
Location

This statement is obviously ambiguous. We cannot determine from this statement the functional relation assumed by women or the functional relation between the times given and Land Grant Universities. However, as in the previous example, the 19th century dates describe the time when the colleges were founded, not the time the women were attending the colleges being discussed. A more careful reading of the abstracts clarifies the ambiguity. The error made by this subject specialist is a little surprising because the following abstract statement was rejected by the subject specialist even though it contains the same functional relations as those above:

Abstract: ...In Quebec, 1765 to 1840 was the heroic age for the foundation of classical colleges. 1840 to 1919 saw new establishments and the next two decades saw the expansion of the system to include institutions for women.

The remaining disagreements between decisions based on functional relations and the subject specialists
decisions (10 in total) were unremarkable in that they followed no discernible pattern. Most of them seemed to be due to some degree of ambiguity of the kind described for Query 28 above or to simple human error.
4.2.3. Tabulation of the Results of the Analysis.

The results of the analysis of each query/abstracts set were recorded in tables as follows:

**EXAMPLE 1. Query 13. (Jenkins Activity Survey)**

<table>
<thead>
<tr>
<th>Abstract no.</th>
<th>Subject Specialists' Decision</th>
<th>Functional Relations Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>2</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>3</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>4</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>5</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>6</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>7</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>8</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>9</td>
<td>&quot;ABOUT&quot;</td>
<td>MATCH</td>
</tr>
<tr>
<td>10</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>11</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>12</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>13</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>14</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>15</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
</tbody>
</table>
**EXAMPLE 2. Query 28. (Women's and Children's Fiction)**

<table>
<thead>
<tr>
<th>Abstract no.</th>
<th>Subject Specialists' Decision</th>
<th>Functional Relations Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>2</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>3</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>4</td>
<td>NOT ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>5</td>
<td>NOT ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>6</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>7</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>8</td>
<td>ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>9</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>10</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>11</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>12</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>13</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>14</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>15</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>16</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>17</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>18</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>19</td>
<td>ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>20</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>21</td>
<td>NOT ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>22</td>
<td>ABOUT</td>
<td>MATCH</td>
</tr>
<tr>
<td>23</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
<tr>
<td>24</td>
<td>NOT ABOUT</td>
<td>NOT MATCH</td>
</tr>
</tbody>
</table>
The results for each query/abstracts set were then summarized in contingency tables where the columns were the subject specialists' decisions and the rows were the functional relations decisions. The four categories of match represented in these tables can be described as:

1. **ABOUT/MATCH** includes the total number of abstracts judged to be ABOUT by the subject specialist which were also judged to be a MATCH using functional relations. The abstracts counted under this category, then, contained a statement which matched a query statement that the subject specialist had indicated was sufficient for an abstract to be judged to be ABOUT the query.

2. **ABOUT/NOT MATCH** includes those abstracts judged to be ABOUT by the subject specialist which were judged to be NOT a MATCH using functional relations. The abstracts counted under this category, then, did not contain a statement which matched a query statement that the subject specialist had indicated was sufficient for an abstract to be judged to be ABOUT the query.

3. **NOT ABOUT/MATCH** included those abstracts judged to be ABOUT the query by the subject specialist which were judged to be a MATCH using functional relations. The abstracts counted under this category, then, did contain a statement which MATCHED a query statement which the subject specialist
had otherwise indicated was sufficient for an abstract to be judged to be ABOUT.

4. NOT ABOUT/NOT MATCH included those abstracts judged to be NOT ABOUT by the subject specialist and NOT A MATCH using functional relations. The abstracts counted under this category, when, did not contain a statement which MATCHED a query statement.

Categories 1 and 4, then, represented agreement between the subject specialists' decisions and the functional relations decisions, and Categories 2 and 3 represented disagreement between the subject specialists' decisions and the functional relations decisions.
4.3. The Analysis of the Secondary Judges' Decisions.

There were two tests for the secondary judge's decisions. The first test compared the decisions of the secondary judges with the decisions of the subject specialists. The second tested the internal consistency of the secondary judge's decisions using functional relations.

**MEASURE 1.** The aboutness decisions of the secondary judges were compared with the functional relations decisions based on the subject specialists' aboutness decisions. The results of this comparison for the Experienced judgements and for the Inexperienced judgements were represented in 2X2 contingency tables where the rows were the decisions based on functional relations (MATCH) and the columns were the decisions of the secondary judges (ABOUT).

**MEASURE 2.** Measure 2 required an analysis of each judge's decisions which closely followed the analysis of the subject specialists' decisions. A quick scan of the data showed that many of the disagreements between the secondary judges and the functional relations model used for MEASURE 1 were due to disagreements about quasi-synonyms. That is, the secondary judges often rejected - i.e., judged to be not about - abstracts which contained quasi-synonyms which had been mentioned in the query and had been accepted as equivalent by the subject specialist. Therefore, the first step in the analysis of a secondary judge's decisions
was the determination of which quasi-synonyms had been accepted and which had been rejected consistently. If the secondary judge had been inconsistent in judgements involving quasi-synonyms, the decision for the majority of abstracts containing the term was considered to be the decision on which MATCH/NOT MATCH would be decided.

The number of abstracts which were judged to be not about the query because they contained a term which had been consistently rejected by the secondary judge were entered under the NOT ABOUT/NOT MATCH category. The remaining abstracts were analyzed using the same procedures used for analyzing the agreement between the researcher and the subject specialist — i.e., the decisions of each secondary judge were compared to the functional relations decisions based on their own choice of query statements.

The results of the analyses were recorded in the same way as for the subject specialists' decisions. Each judge's results were represented in a 2X2 contingency table where the rows represented the decisions based on functional relations and the columns represented the decisions of the secondary judge. There were thirty contingency tables for the Experienced decisions and thirty tables for the Inexperienced decisions.

The analysis of the second linguist's decisions involved two sets of decisions:

1. the MATCH/NOT MATCH decisions based on whether or not there was a MATCH in functional relations between the keyterms in the abstracts and the corresponding queries; and

2. the case role and verb type assignments for keyterms in the query statements and in the abstracts.

The first test of agreement between the researcher and the second language analyst for MATCH or NOT MATCH of statements in abstracts and query statements involved a simple comparison of the two sets of decisions. The results of this test were expressed in percent of agreement.

The second test for agreement in functional relations assignments was analyzed to determine the level of agreement for each case role and verb type so that the consistency of application of the definitions for the two analysts could be determined as well as the overall agreement. The results for each functional relations and for the total set of functional relations was obtained. The results were represented as a simple percentage.
CHAPTER 5

STATISTICAL TESTS

Fienberg (1980) has postulated a number of log-linear statistical models for the analysis of cross-classified data. These models permit one to test for various degrees of relationship among the cross-classified variables. Four such models are applied in this chapter to the data from the functional relations/aboutness test. The tests and the results of the tests are presented by statistical MODEL used to obtain expected values so that the comparisons between the performances of each of the groups of judges will be clearer. The contingency tables and a summary of the chi square values for each test under each statistical MODEL will include the subject specialists' decisions, the experienced decisions and inexperienced decisions compared with the subject specialists' query statements decisions, and the experienced and inexperienced decisions compared with their own choice of query statements. The results of the tests involving comparisons between the researcher and the second language analyst are presented in Section 2 of this chapter. Section 3 provides a brief summary of the conclusions reached for all of the tests.
5.1. Relationship of Aboutness, Match and Query.

The results of the analyses of the various judges' decisions were represented in 2x2x30 contingency tables with the dimensions:

1. About/Not About decisions of the judges;
2. Match/Not Match decisions based on functional relations; and
3. the 30 queries.

There were three bases for determining whether there was a match to the functional relations of a query statement:

1. the query statements selected by the subject specialists;
2. the query statements selected by the secondary judge when making experienced judgements; and
3. the query statements selected by the secondary judges when making inexperienced judgements.

The functional relations model based on the subject specialists' query decisions was also used for assessing the relationships between Match and Aboutness for the experienced and inexperienced judgements. There were, then, five relationships presented in the five contingency tables as follows:

CONTINGENCY TABLE 1 represents the subject specialists' aboutness decisions compared with the functional relations decisions based on their query statements;
CONTINGENCY TABLE 2 represents the secondary judges' experienced aboutness decisions compared with the functional relations based on the subject specialists' query statements;

CONTINGENCY TABLE 3 represents the secondary judges' inexperienced aboutness decisions compared with the functional relations decisions based on the subject specialists' query statements;

CONTINGENCY TABLE 4 represents the secondary judges' experienced aboutness decisions compared with the functional relations decisions based on their own query statements; and

CONTINGENCY TABLE 5 represents the secondary judges' inexperienced decisions compared with the functional relations decision based on their own query statements.

In summary, then, the agreements between the aboutness decisions of the three groups of judges and the functional relations matching based on the subject specialists' decisions are represented in Contingency Tables 1-3; and the agreement between the secondary judges' experienced aboutness decisions and inexperienced aboutness decisions and the functional relations matching based on their own decisions are represented in Contingency Tables 4 and 5.
Four statistical models describing the various degrees of relationships among the three variables About/Not About (the judges' decisions), Match/Not Match (the functional relations decisions), and Queries were tested for each of the contingency tables. The Models were:

MODEL I tested for three-way independence of Aboutness, match and queries;
MODEL II TESTED independence of the queries from the aboutness/match interactions;
MODEL III tested whether aboutness is independent of the query given fixed relevance or nonrelevance (match or not match in functional relations); and
MODEL IV tested independence of About and Match.

The chi square statistic applied to each model was:

\[ \chi^2 = \frac{(X_{ijk} - \hat{M}_{ijk})^2}{\hat{M}_{ijk}} \]

where

\( X_{ijk} \) = the observed value in each cell; and
\( \hat{M}_{ijk} \) = the expected value for each cell.

The description of the statistical tests and of the results will be given in the order in which the models were applied, and will include:

1. the description of the model;
2. the contingency tables for each of the four models.
3. the results of the Chi square tests and the conclusions reached on the basis of the results of the test.

The columns of the tables represent the cross-categories:

1. ABOUT and MATCH,
2. NOT ABOUT and MATCH,
3. ABOUT and NOT MATCH,
4. NOT ABOUT and NOT MATCH.

where ABOUT represents the aboutness judgements of the judge and MATCH the functional relations decisions. The observed values given for each cell represent the total number of documents in the cross-category for each query. The expected values for each cell were obtained using the model being discussed.
MODEL I. This model tested three-way independence of aboutness, match, and queries:

\[ H_0 : \lambda_{ijk} = \frac{(X_{i++})(X_{j++})(X_{++k})}{N^2} \]

where

\[ \lambda_{ijk} \] is the expected value for each cell;

\[ X_{i++} \] is the total number of about documents;

\[ X_{2++} \] is the total number of not about documents;

\[ X_{+1+} \] is the total number of matching documents;

\[ X_{+2+} \] is the total number of not matching documents;

\[ X_{++k} \] is the total number of documents in the \( k \)th query... \( k=1,...,k \)

\[ N = X_{+++} \]

and

Degrees of Freedom = IJK-I-J-K+2

\[ \chi^2 = (2)(2)(30)-2-2-30+2 = 88 \]

The critical value at \( \alpha = .05 \) is 110.

The five tables which follow show the observed and expected values for each cell and the Chi square for each query.
## MODEL I. Subject Specialists' Relevance and Query Judgements

### NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>$X^\hat{m}$</td>
<td>$X^\hat{n}$</td>
</tr>
<tr>
<td>1. 6</td>
<td>3.9</td>
</tr>
<tr>
<td>2. 5</td>
<td>2.6</td>
</tr>
<tr>
<td>3. 21</td>
<td>11.2</td>
</tr>
<tr>
<td>4. 31</td>
<td>14.2</td>
</tr>
<tr>
<td>5. 12</td>
<td>10.3</td>
</tr>
<tr>
<td>6. 23</td>
<td>12.0</td>
</tr>
<tr>
<td>7. 20</td>
<td>14.6</td>
</tr>
<tr>
<td>8. 10</td>
<td>13.8</td>
</tr>
<tr>
<td>9. 14</td>
<td>12.9</td>
</tr>
<tr>
<td>10. 16</td>
<td>7.3</td>
</tr>
<tr>
<td>11. 11</td>
<td>12.0</td>
</tr>
<tr>
<td>12. 22</td>
<td>11.6</td>
</tr>
<tr>
<td>13. 12</td>
<td>6.5</td>
</tr>
<tr>
<td>14. 16</td>
<td>10.8</td>
</tr>
<tr>
<td>15. 9</td>
<td>11.2</td>
</tr>
<tr>
<td>16. 15</td>
<td>12.5</td>
</tr>
<tr>
<td>17. 19</td>
<td>12.5</td>
</tr>
<tr>
<td>18. 22</td>
<td>12.0</td>
</tr>
<tr>
<td>19. 8</td>
<td>4.7</td>
</tr>
<tr>
<td>20. 16</td>
<td>14.2</td>
</tr>
<tr>
<td>21. 13</td>
<td>7.7</td>
</tr>
<tr>
<td>22. 13</td>
<td>15.1</td>
</tr>
<tr>
<td>23. 25</td>
<td>13.3</td>
</tr>
<tr>
<td>24. 23</td>
<td>11.6</td>
</tr>
<tr>
<td>25. 7</td>
<td>3.9</td>
</tr>
<tr>
<td>26. 31</td>
<td>13.8</td>
</tr>
<tr>
<td>27. 16</td>
<td>14.2</td>
</tr>
<tr>
<td>28. 11</td>
<td>10.3</td>
</tr>
<tr>
<td>29. 13</td>
<td>6.5</td>
</tr>
<tr>
<td>30. 14</td>
<td>8.6</td>
</tr>
</tbody>
</table>

$\chi^2 = 907.7$ Critical value $= 110$ $H_o$ rejected.
MODEL I. Experienced Relevance and Subject Specialists’ Query Judgements.

**NUMBER OF DOCUMENTS**

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
<th>MATCH</th>
<th>NOT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>$X$</td>
<td>$M$</td>
<td>$X$</td>
<td>$M$</td>
</tr>
<tr>
<td>1. 4</td>
<td>3.3</td>
<td>1.8</td>
<td>2.5</td>
</tr>
<tr>
<td>2. 3</td>
<td>2.2</td>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td>3. 13</td>
<td>9.6</td>
<td>5.2</td>
<td>8.7</td>
</tr>
<tr>
<td>4. 17</td>
<td>12.2</td>
<td>6.6</td>
<td>9.2</td>
</tr>
<tr>
<td>5. 7</td>
<td>8.9</td>
<td>4.8</td>
<td>6.7</td>
</tr>
<tr>
<td>6. 8</td>
<td>10.4</td>
<td>5.6</td>
<td>7.8</td>
</tr>
<tr>
<td>7. 11</td>
<td>12.6</td>
<td>6.8</td>
<td>9.5</td>
</tr>
<tr>
<td>8. 6</td>
<td>11.8</td>
<td>6.4</td>
<td>9.0</td>
</tr>
<tr>
<td>9. 7</td>
<td>11.1</td>
<td>6.0</td>
<td>8.4</td>
</tr>
<tr>
<td>10: 6</td>
<td>6.3</td>
<td>3.4</td>
<td>10.4</td>
</tr>
<tr>
<td>11: 8</td>
<td>10.4</td>
<td>5.6</td>
<td>7.8</td>
</tr>
<tr>
<td>12: 10</td>
<td>10.0</td>
<td>5.4</td>
<td>7.6</td>
</tr>
<tr>
<td>13: 7</td>
<td>5.6</td>
<td>3.0</td>
<td>5.4</td>
</tr>
<tr>
<td>14: 12</td>
<td>9.3</td>
<td>0</td>
<td>5.0</td>
</tr>
<tr>
<td>15: 10</td>
<td>9.6</td>
<td>5.2</td>
<td>7.3</td>
</tr>
<tr>
<td>16: 11</td>
<td>10.7</td>
<td>5.8</td>
<td>8.1</td>
</tr>
<tr>
<td>17: 21</td>
<td>10.7</td>
<td>5.8</td>
<td>8.1</td>
</tr>
<tr>
<td>18: 17</td>
<td>10.4</td>
<td>5.6</td>
<td>7.8</td>
</tr>
<tr>
<td>19: 4</td>
<td>4.1</td>
<td>2.2</td>
<td>3.1</td>
</tr>
<tr>
<td>20: 16</td>
<td>12.2</td>
<td>6.6</td>
<td>9.2</td>
</tr>
<tr>
<td>21: 10</td>
<td>6.7</td>
<td>3.6</td>
<td>5.0</td>
</tr>
<tr>
<td>22: 11</td>
<td>13.0</td>
<td>7.0</td>
<td>9.8</td>
</tr>
<tr>
<td>23: 17</td>
<td>11.5</td>
<td>6.2</td>
<td>8.7</td>
</tr>
<tr>
<td>24: 23</td>
<td>10.0</td>
<td>5.4</td>
<td>7.6</td>
</tr>
<tr>
<td>25: 5</td>
<td>3.3</td>
<td>1.8</td>
<td>2.5</td>
</tr>
<tr>
<td>26: 16</td>
<td>11.8</td>
<td>6.4</td>
<td>9.0</td>
</tr>
<tr>
<td>27: 11</td>
<td>12.2</td>
<td>6.6</td>
<td>9.2</td>
</tr>
<tr>
<td>28: 12</td>
<td>8.9</td>
<td>4.8</td>
<td>6.7</td>
</tr>
<tr>
<td>29: 11</td>
<td>6.7</td>
<td>3.6</td>
<td>5.0</td>
</tr>
<tr>
<td>30: 8</td>
<td>7.4</td>
<td>4.0</td>
<td>5.6</td>
</tr>
</tbody>
</table>

$\chi^2 = 339.8$  Critical value = 110  $H_0$ : rejected
MODEL I. Inexperienced. Relevance and Subject Specialists' Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>NOT ABOUT</th>
<th>MATCH ABOUT</th>
<th>NOT MATCH ABOUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1. 3</td>
<td>3.2</td>
<td>1.7</td>
<td>2.1</td>
</tr>
<tr>
<td>2. 4</td>
<td>2.2</td>
<td>0.0</td>
<td>1.1</td>
</tr>
<tr>
<td>3. 16</td>
<td>9.4</td>
<td>5.1</td>
<td>3.9</td>
</tr>
<tr>
<td>4. 28</td>
<td>11.9</td>
<td>2.6</td>
<td>0.5</td>
</tr>
<tr>
<td>5. 6</td>
<td>8.6</td>
<td>9.3</td>
<td>3.6</td>
</tr>
<tr>
<td>6. 5</td>
<td>10.1</td>
<td>0.3</td>
<td>1.4</td>
</tr>
<tr>
<td>7. 20</td>
<td>12.2</td>
<td>0.6</td>
<td>3.1</td>
</tr>
<tr>
<td>8. 5</td>
<td>11.5</td>
<td>4.8</td>
<td>2.2</td>
</tr>
<tr>
<td>9. 13</td>
<td>10.8</td>
<td>5.4</td>
<td>12.5</td>
</tr>
<tr>
<td>10. 14</td>
<td>6.1</td>
<td>3.2</td>
<td>18.6</td>
</tr>
<tr>
<td>11. 9</td>
<td>10.1</td>
<td>5.3</td>
<td>9.0</td>
</tr>
<tr>
<td>12. 9</td>
<td>9.7</td>
<td>4.1</td>
<td>8.4</td>
</tr>
<tr>
<td>13. 7</td>
<td>5.4</td>
<td>2.9</td>
<td>3.7</td>
</tr>
<tr>
<td>14. 11</td>
<td>9.0</td>
<td>4.8</td>
<td>8.2</td>
</tr>
<tr>
<td>15. 4</td>
<td>9.4</td>
<td>4.9</td>
<td>16.4</td>
</tr>
<tr>
<td>16. 12</td>
<td>10.4</td>
<td>5.5</td>
<td>10.9</td>
</tr>
<tr>
<td>17. 13</td>
<td>10.4</td>
<td>8.7</td>
<td>6.0</td>
</tr>
<tr>
<td>18. 13</td>
<td>10.1</td>
<td>8.4</td>
<td>5.0</td>
</tr>
<tr>
<td>19. 8</td>
<td>4.0</td>
<td>2.1</td>
<td>10.4</td>
</tr>
<tr>
<td>20. 11</td>
<td>11.9</td>
<td>5.6</td>
<td>11.6</td>
</tr>
<tr>
<td>21. 10</td>
<td>6.5</td>
<td>3.4</td>
<td>8.4</td>
</tr>
<tr>
<td>22. 6</td>
<td>12.6</td>
<td>5.3</td>
<td>31.1</td>
</tr>
<tr>
<td>23. 13</td>
<td>11.2</td>
<td>5.9</td>
<td>3.1</td>
</tr>
<tr>
<td>24. 20</td>
<td>9.7</td>
<td>5.1</td>
<td>17.4</td>
</tr>
<tr>
<td>25. 1</td>
<td>3.2</td>
<td>0.0</td>
<td>7.5</td>
</tr>
<tr>
<td>26. 20</td>
<td>11.5</td>
<td>0.1</td>
<td>15.6</td>
</tr>
<tr>
<td>27. 14</td>
<td>11.9</td>
<td>6.3</td>
<td>35.4</td>
</tr>
<tr>
<td>28. 11</td>
<td>8.6</td>
<td>2.4</td>
<td>14.1</td>
</tr>
<tr>
<td>29. 11</td>
<td>6.5</td>
<td>2.3</td>
<td>5.2</td>
</tr>
<tr>
<td>30. 10</td>
<td>7.2</td>
<td>1.3</td>
<td>5.2</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 376.2 \quad \text{Critical value } = 100 \quad H_0 \text{ rejected} \]
# MODEL I. Experienced Relevance and Query Judgements.

## NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>MATCH NOT ABOUT</th>
<th>NOT MATCH ABOUT</th>
<th>NOT MATCH NOT ABOUT</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>$\hat{X}$</td>
<td>$X$</td>
<td>$\hat{X}$</td>
<td>$X$</td>
</tr>
<tr>
<td>1. 4</td>
<td>3.1</td>
<td>2</td>
<td>2.4</td>
<td>1</td>
</tr>
<tr>
<td>2. 3</td>
<td>2.0</td>
<td>1</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>3. 15</td>
<td>8.8</td>
<td>4</td>
<td>7.0</td>
<td>1</td>
</tr>
<tr>
<td>4. 17</td>
<td>11.2</td>
<td>14</td>
<td>8.9</td>
<td>1</td>
</tr>
<tr>
<td>5. 7</td>
<td>8.2</td>
<td>0</td>
<td>6.5</td>
<td>3</td>
</tr>
<tr>
<td>6. 5</td>
<td>9.5</td>
<td>3</td>
<td>7.6</td>
<td>3</td>
</tr>
<tr>
<td>7. 10</td>
<td>11.6</td>
<td>6</td>
<td>9.2</td>
<td>3</td>
</tr>
<tr>
<td>8. 14</td>
<td>10.9</td>
<td>2</td>
<td>8.6</td>
<td>0</td>
</tr>
<tr>
<td>9. 9</td>
<td>10.2</td>
<td>3</td>
<td>8.1</td>
<td>3</td>
</tr>
<tr>
<td>10. 6</td>
<td>5.8</td>
<td>3</td>
<td>4.6</td>
<td>1</td>
</tr>
<tr>
<td>11. 17</td>
<td>9.5</td>
<td>0</td>
<td>7.6</td>
<td>3</td>
</tr>
<tr>
<td>12. 10</td>
<td>9.2</td>
<td>0</td>
<td>7.3</td>
<td>0</td>
</tr>
<tr>
<td>13. 6</td>
<td>5.1</td>
<td>2</td>
<td>4.1</td>
<td>1</td>
</tr>
<tr>
<td>14. 9</td>
<td>8.5</td>
<td>0</td>
<td>6.8</td>
<td>3</td>
</tr>
<tr>
<td>15. 24</td>
<td>8.8</td>
<td>0</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td>16. 12</td>
<td>9.9</td>
<td>2</td>
<td>7.8</td>
<td>0</td>
</tr>
<tr>
<td>17. 21</td>
<td>9.9</td>
<td>0</td>
<td>7.8</td>
<td>3</td>
</tr>
<tr>
<td>18. 21</td>
<td>9.5</td>
<td>3</td>
<td>7.6</td>
<td>0</td>
</tr>
<tr>
<td>19. 4</td>
<td>3.7</td>
<td>3</td>
<td>3.0</td>
<td>0</td>
</tr>
<tr>
<td>20. 32</td>
<td>11.2</td>
<td>0</td>
<td>8.9</td>
<td>0</td>
</tr>
<tr>
<td>21. 10</td>
<td>6.1</td>
<td>0</td>
<td>4.9</td>
<td>2</td>
</tr>
<tr>
<td>22. 21</td>
<td>11.9</td>
<td>5</td>
<td>9.5</td>
<td>1</td>
</tr>
<tr>
<td>23. 17</td>
<td>10.5</td>
<td>6</td>
<td>8.4</td>
<td>1</td>
</tr>
<tr>
<td>24. 20</td>
<td>9.2</td>
<td>0</td>
<td>7.3</td>
<td>2</td>
</tr>
<tr>
<td>25. 5</td>
<td>3.1</td>
<td>1</td>
<td>2.4</td>
<td>0</td>
</tr>
<tr>
<td>26. 10</td>
<td>10.9</td>
<td>5</td>
<td>8.6</td>
<td>0</td>
</tr>
<tr>
<td>27. 11</td>
<td>11.2</td>
<td>5</td>
<td>8.9</td>
<td>3</td>
</tr>
<tr>
<td>28. 12</td>
<td>8.2</td>
<td>1</td>
<td>6.5</td>
<td>0</td>
</tr>
<tr>
<td>29. 11</td>
<td>6.1</td>
<td>4</td>
<td>4.9</td>
<td>2</td>
</tr>
<tr>
<td>30. 8</td>
<td>6.8</td>
<td>2</td>
<td>5.4</td>
<td>0</td>
</tr>
</tbody>
</table>

$\chi^2 = 596.2$  
Critical value = 11D  
$H_0$ rejected
MODEL 1. Inexperienced Relevance and Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>NOT ABOUT ABOUT</th>
<th>NOT ABOUT ABOUT</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>1. 3 3.1</td>
<td>3 2.4</td>
<td>1 2.0</td>
<td>2 1.5</td>
</tr>
<tr>
<td>2. 4 2.0</td>
<td>0 1.6</td>
<td>0 1.3</td>
<td>2 1.0</td>
</tr>
<tr>
<td>3. 19 8.8</td>
<td>0 7.0</td>
<td>2 5.7</td>
<td>5 4.4</td>
</tr>
<tr>
<td>4. 28 11.2</td>
<td>3 8.9</td>
<td>2 7.3</td>
<td>0 5.6</td>
</tr>
<tr>
<td>5. 6 8.2</td>
<td>1 6.5</td>
<td>4 5.3</td>
<td>13 4.1</td>
</tr>
<tr>
<td>6. 4 9.5</td>
<td>3 7.6</td>
<td>1 6.2</td>
<td>20 4.8</td>
</tr>
<tr>
<td>7. 26 11.6</td>
<td>0 9.2</td>
<td>0 7.5</td>
<td>18 5.8</td>
</tr>
<tr>
<td>8. 20 10.9</td>
<td>8 8.6</td>
<td>0 7.0</td>
<td>4 5.4</td>
</tr>
<tr>
<td>9. 23 10.2</td>
<td>1 8.1</td>
<td>2 6.6</td>
<td>4 5.1</td>
</tr>
<tr>
<td>10. 14 5.8</td>
<td>0 4.6</td>
<td>0 3.7</td>
<td>3 2.9</td>
</tr>
<tr>
<td>11. 13 9.5</td>
<td>5 7.6</td>
<td>0 6.2</td>
<td>10 4.8</td>
</tr>
<tr>
<td>12. 9 9.2</td>
<td>0 7.3</td>
<td>0 5.9</td>
<td>18 4.6</td>
</tr>
<tr>
<td>13. 7 5.1</td>
<td>4 4.1</td>
<td>0 3.3</td>
<td>4 2.6</td>
</tr>
<tr>
<td>14. 11 8.5</td>
<td>1 6.8</td>
<td>2 5.5</td>
<td>11 4.3</td>
</tr>
<tr>
<td>15. 4 8.8</td>
<td>3 7.0</td>
<td>5 5.7</td>
<td>14 4.4</td>
</tr>
<tr>
<td>16. 15 9.9</td>
<td>1 7.8</td>
<td>0 6.4</td>
<td>13 4.9</td>
</tr>
<tr>
<td>17. 13 9.9</td>
<td>3 7.8</td>
<td>1 6.4</td>
<td>12 4.9</td>
</tr>
<tr>
<td>18. 17 9.5</td>
<td>4 7.6</td>
<td>2 6.2</td>
<td>5 4.8</td>
</tr>
<tr>
<td>19. 8 3.7</td>
<td>0 3.0</td>
<td>0 2.4</td>
<td>3 1.9</td>
</tr>
<tr>
<td>20. 12 11.2</td>
<td>7 8.9</td>
<td>3 7.3</td>
<td>11 5.6</td>
</tr>
<tr>
<td>21. 10 6.1</td>
<td>1 4.9</td>
<td>0 4.0</td>
<td>7 3.1</td>
</tr>
<tr>
<td>22. 9 11.9</td>
<td>2 9.5</td>
<td>3 7.7</td>
<td>21 6.0</td>
</tr>
<tr>
<td>23. 14 10.5</td>
<td>3 8.4</td>
<td>2 6.8</td>
<td>12 5.3</td>
</tr>
<tr>
<td>24. 23 9.2</td>
<td>0 7.3</td>
<td>1 5.9</td>
<td>3 4.6</td>
</tr>
<tr>
<td>25. 1 3.1</td>
<td>6 2.4</td>
<td>0 2.0</td>
<td>2 1.5</td>
</tr>
<tr>
<td>26. 17 10.9</td>
<td>6 8.6</td>
<td>2 7.0</td>
<td>7 5.4</td>
</tr>
<tr>
<td>27. 14 11.2</td>
<td>2 8.9</td>
<td>1 7.3</td>
<td>16 5.6</td>
</tr>
<tr>
<td>28. 11 8.2</td>
<td>2 6.5</td>
<td>2 5.3</td>
<td>9 4.1</td>
</tr>
<tr>
<td>29. 11 6.1</td>
<td>4 4.9</td>
<td>2 4.0</td>
<td>1 3.1</td>
</tr>
<tr>
<td>30. 0 6.8</td>
<td>1 5.4</td>
<td>2 4.4</td>
<td>7 3.4</td>
</tr>
</tbody>
</table>

\( \chi^2 = 655.7 \quad \text{Critical value} = 110 \quad H_0: \text{rejected} \)
The results of the chi square tests for each of the five categories under MODEL I can be summarized as:

1. Subject specialist. Category 1. 907.7
2. Experienced. Category 2. 339.8
3. Inexperienced. Category 3 376.2
4. Experienced. Category 4 596.2
5. Inexperienced. Category 5 655.7

With a critical value of 110 at $\alpha = .05$, it is concluded that the null hypothesis of no relationship between ABOUT, MATCH, and QUERIES is rejected. However, this model does not tell us which variables are related. Since the relationship between ABOUT and MATCH is the relationship of interest, the next model was used to determine whether there was a significant effect on this relationship from variation among QUERIES.
MODEL II. This model tested independence of the aboutness/match interactions from the queries:

\[ H_0 : \hat{m}_{ijk} = \frac{(X_{ij+})(X_{++k})}{N} \]

where

- \( m_{ijk} \) = the expected value for each cell;
- \( X_{ij+} \) = the total number of about and matching documents;
- \( X_{++k} \) = the total number of not about and matching documents;
- \( X_{12+} \) = the total number of about and not matching documents;
- \( N = X_{+++} \)

and

- Degrees of Freedom = \((I-1)(J-1)K\)
- \( \chi^2 = (1)(1)30 = 30 \)

The critical value at \( \alpha = .05 \) is 42.6

The five tables which follow show the Observed and Expected values for each cell.
MODEL II. Subject Specialists’ Relevance and Query Judgements.

**NUMBER OF DOCUMENTS**

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>X</td>
<td>M</td>
</tr>
<tr>
<td>1.</td>
<td>6 5.8</td>
</tr>
<tr>
<td>2.</td>
<td>5 3.8</td>
</tr>
<tr>
<td>3.</td>
<td>21 16.6</td>
</tr>
<tr>
<td>4.</td>
<td>31 21.1</td>
</tr>
<tr>
<td>5.</td>
<td>12 15.4</td>
</tr>
<tr>
<td>6.</td>
<td>23 17.9</td>
</tr>
<tr>
<td>7.</td>
<td>20 21.8</td>
</tr>
<tr>
<td>8.</td>
<td>10 20.5</td>
</tr>
<tr>
<td>9.</td>
<td>14 19.2</td>
</tr>
<tr>
<td>10.</td>
<td>16 10.9</td>
</tr>
<tr>
<td>11.</td>
<td>22 17.3</td>
</tr>
<tr>
<td>12.</td>
<td>16 16.0</td>
</tr>
<tr>
<td>13.</td>
<td>9 16.6</td>
</tr>
<tr>
<td>14.</td>
<td>15 18.6</td>
</tr>
<tr>
<td>15.</td>
<td>19 18.6</td>
</tr>
<tr>
<td>16.</td>
<td>22 17.9</td>
</tr>
<tr>
<td>17.</td>
<td>8 7.0</td>
</tr>
<tr>
<td>18.</td>
<td>16 21.1</td>
</tr>
<tr>
<td>19.</td>
<td>13 11.5</td>
</tr>
<tr>
<td>20.</td>
<td>13 22.4</td>
</tr>
<tr>
<td>21.</td>
<td>25 19.8</td>
</tr>
<tr>
<td>22.</td>
<td>23 17.3</td>
</tr>
<tr>
<td>23.</td>
<td>7 5.8</td>
</tr>
<tr>
<td>24.</td>
<td>31 20.5</td>
</tr>
<tr>
<td>25.</td>
<td>16 21.1</td>
</tr>
<tr>
<td>26.</td>
<td>11 15.4</td>
</tr>
<tr>
<td>27.</td>
<td>13 11.5</td>
</tr>
<tr>
<td>28.</td>
<td>14 12.8</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 199.7 \]  
Critical value = 42.6  
\[ H_0 : \text{ rejects} \]
MODEL II. Experienced Relevance and Subject Specialists’ Query Judgements.

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>MATCH NOT ABOUT</th>
<th>NOT MATCH ABOUT</th>
<th>NOT MATCH NOT ABOUT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X   M</td>
<td>X   M</td>
<td>X   M</td>
</tr>
<tr>
<td>1.</td>
<td>4   4.0</td>
<td>2   2.0</td>
<td>1   1.2</td>
</tr>
<tr>
<td>2.</td>
<td>3   2.6</td>
<td>2   1.3</td>
<td>0   7.8</td>
</tr>
<tr>
<td>3.</td>
<td>13  11.4</td>
<td>8   5.7</td>
<td>3   3.4</td>
</tr>
<tr>
<td>4.</td>
<td>17  14.5</td>
<td>14  7.3</td>
<td>1   4.3</td>
</tr>
<tr>
<td>5.</td>
<td>7   10.6</td>
<td>4   5.3</td>
<td>3   3.1</td>
</tr>
<tr>
<td>6.</td>
<td>8   12.3</td>
<td>15  6.2</td>
<td>0   3.6</td>
</tr>
<tr>
<td>7.</td>
<td>11  15.0</td>
<td>9   7.5</td>
<td>2   4.4</td>
</tr>
<tr>
<td>8.</td>
<td>6   14.1</td>
<td>14  7.0</td>
<td>10  4.2</td>
</tr>
<tr>
<td>9.</td>
<td>7   13.2</td>
<td>7   6.6</td>
<td>5   3.9</td>
</tr>
<tr>
<td>10.</td>
<td>6   7.5</td>
<td>10  3.7</td>
<td>1   2.2</td>
</tr>
<tr>
<td>11.</td>
<td>8   12.3</td>
<td>4   6.2</td>
<td>9   3.6</td>
</tr>
<tr>
<td>12.</td>
<td>10  7.5</td>
<td>12  5.9</td>
<td>0   3.5</td>
</tr>
<tr>
<td>13.</td>
<td>7   6.6</td>
<td>5   3.3</td>
<td>0   2.0</td>
</tr>
<tr>
<td>14.</td>
<td>12  11.0</td>
<td>5   5.5</td>
<td>0   3.3</td>
</tr>
<tr>
<td>15.</td>
<td>10  11.4</td>
<td>0   5.7</td>
<td>14  3.4</td>
</tr>
<tr>
<td>16.</td>
<td>11  12.8</td>
<td>5   6.4</td>
<td>1   3.8</td>
</tr>
<tr>
<td>17.</td>
<td>21  12.8</td>
<td>0   6.4</td>
<td>3   3.8</td>
</tr>
<tr>
<td>18.</td>
<td>17  12.3</td>
<td>5   6.2</td>
<td>4   3.6</td>
</tr>
<tr>
<td>19.</td>
<td>4   4.8</td>
<td>4   2.4</td>
<td>0   1.4</td>
</tr>
<tr>
<td>20.</td>
<td>16  14.5</td>
<td>0   7.3</td>
<td>16  4.3</td>
</tr>
<tr>
<td>21.</td>
<td>10  7.9</td>
<td>3   4.0</td>
<td>2   2.3</td>
</tr>
<tr>
<td>22.</td>
<td>11  15.4</td>
<td>1   7.7</td>
<td>11  4.6</td>
</tr>
<tr>
<td>23.</td>
<td>17  13.6</td>
<td>8   6.8</td>
<td>3   4.0</td>
</tr>
<tr>
<td>24.</td>
<td>23  7.5</td>
<td>0   5.9</td>
<td>1   3.5</td>
</tr>
<tr>
<td>25.</td>
<td>5   4.0</td>
<td>2   2.0</td>
<td>0   1.2</td>
</tr>
<tr>
<td>26.</td>
<td>16  14.1</td>
<td>15  7.0</td>
<td>0   4.2</td>
</tr>
<tr>
<td>27.</td>
<td>11  14.5</td>
<td>5   7.3</td>
<td>3   4.3</td>
</tr>
<tr>
<td>28.</td>
<td>12  10.6</td>
<td>12  5.3</td>
<td>0   3.1</td>
</tr>
<tr>
<td>29.</td>
<td>11  7.9</td>
<td>4   4.0</td>
<td>2   2.3</td>
</tr>
<tr>
<td>30.</td>
<td>8   8.8</td>
<td>6   4.4</td>
<td>0   2.6</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 352.0 \quad \text{Critical value} = 42.6 \quad H_0: \text{rejected} \]
MODEL II. Inexperienced Relevance and Subject Specialists' Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>X</td>
<td>M</td>
</tr>
<tr>
<td>1.</td>
<td>3 4.0</td>
</tr>
<tr>
<td>2.</td>
<td>4 2.6</td>
</tr>
<tr>
<td>3.</td>
<td>16 11.4</td>
</tr>
<tr>
<td>4.</td>
<td>28 14.5</td>
</tr>
<tr>
<td>5.</td>
<td>6 10.6</td>
</tr>
<tr>
<td>6.</td>
<td>5 12.3</td>
</tr>
<tr>
<td>7.</td>
<td>20 15.0</td>
</tr>
<tr>
<td>8.</td>
<td>5 14.1</td>
</tr>
<tr>
<td>9.</td>
<td>13 13.2</td>
</tr>
<tr>
<td>10.</td>
<td>14 7.5</td>
</tr>
<tr>
<td>11.</td>
<td>9 12.3</td>
</tr>
<tr>
<td>12.</td>
<td>10 12.3</td>
</tr>
<tr>
<td>13.</td>
<td>7 6.6</td>
</tr>
<tr>
<td>14.</td>
<td>11 11.0</td>
</tr>
<tr>
<td>15.</td>
<td>4 11.4</td>
</tr>
<tr>
<td>16.</td>
<td>12 12.8</td>
</tr>
<tr>
<td>17.</td>
<td>13 12.8</td>
</tr>
<tr>
<td>18.</td>
<td>13 12.3</td>
</tr>
<tr>
<td>19.</td>
<td>8 4.8</td>
</tr>
<tr>
<td>20.</td>
<td>9 14.5</td>
</tr>
<tr>
<td>21.</td>
<td>10 7.9</td>
</tr>
<tr>
<td>22.</td>
<td>6 15.4</td>
</tr>
<tr>
<td>23.</td>
<td>13 13.6</td>
</tr>
<tr>
<td>24.</td>
<td>20 11.9</td>
</tr>
<tr>
<td>25.</td>
<td>1 4.0</td>
</tr>
<tr>
<td>26.</td>
<td>20 14.1</td>
</tr>
<tr>
<td>27.</td>
<td>14 14.5</td>
</tr>
<tr>
<td>28.</td>
<td>11 10.6</td>
</tr>
<tr>
<td>29.</td>
<td>11 7.9</td>
</tr>
<tr>
<td>30.</td>
<td>10 8.8</td>
</tr>
</tbody>
</table>

\(\chi^2 = 306.2\) Critical value = 42.6 \(H_0\) rejected
MODEL II. Experienced Relevance and Query Judgements.

## NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>NOT MATCH ABOUT</th>
<th>MATCH NOT ABOUT</th>
<th>NOT MATCH NOT ABOUT</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>X ( \hat{M} )</td>
<td>X ( \hat{M} )</td>
<td>X ( \hat{M} )</td>
<td>X ( \hat{M} )</td>
<td>X ( \hat{M} )</td>
</tr>
<tr>
<td>1.  4 4.5</td>
<td>2 .9</td>
<td>1 .5</td>
<td>2 3.1</td>
<td>2.3</td>
</tr>
<tr>
<td>2.  3 3.0</td>
<td>1 .6</td>
<td>0 .3</td>
<td>2 2.0</td>
<td>.6</td>
</tr>
<tr>
<td>3. 15 13.0</td>
<td>4 2.6</td>
<td>1 1.3</td>
<td>6 8.8</td>
<td>2.1</td>
</tr>
<tr>
<td>4. 17 16.5</td>
<td>14 3.3</td>
<td>1 1.7</td>
<td>1 11.2</td>
<td>44.3</td>
</tr>
<tr>
<td>5. 7 12.0</td>
<td>0 2.4</td>
<td>3 1.2</td>
<td>14 8.2</td>
<td>11.3</td>
</tr>
<tr>
<td>6. 5 14.0</td>
<td>3 2.8</td>
<td>3 1.4</td>
<td>17 9.5</td>
<td>13.5</td>
</tr>
<tr>
<td>7. 10 17.0</td>
<td>6 3.4</td>
<td>3 1.7</td>
<td>15 11.6</td>
<td>6.9</td>
</tr>
<tr>
<td>8. 14 16.0</td>
<td>2 3.2</td>
<td>0 1.6</td>
<td>16 10.9</td>
<td>4.8</td>
</tr>
<tr>
<td>9. 9 15.0</td>
<td>3 3.0</td>
<td>3 1.5</td>
<td>15 10.2</td>
<td>6.2</td>
</tr>
<tr>
<td>10. 6 8.5</td>
<td>3 1.7</td>
<td>1 .9</td>
<td>7 5.8</td>
<td>1.9</td>
</tr>
<tr>
<td>11. 17 14.0</td>
<td>0 2.8</td>
<td>3 1.4</td>
<td>8 9.5</td>
<td>5.4</td>
</tr>
<tr>
<td>12. 10 13.5</td>
<td>0 2.7</td>
<td>0 1.4</td>
<td>17 9.2</td>
<td>11.6</td>
</tr>
<tr>
<td>13. 6 7.5</td>
<td>2 1.5</td>
<td>1 .8</td>
<td>6 5.1</td>
<td>.7</td>
</tr>
<tr>
<td>14. 9 12.5</td>
<td>0 2.5</td>
<td>3 1.3</td>
<td>13 8.5</td>
<td>8.1</td>
</tr>
<tr>
<td>15. 24 13.0</td>
<td>0 2.6</td>
<td>0 1.3</td>
<td>2 8.8</td>
<td>18.5</td>
</tr>
<tr>
<td>16. 12 14.5</td>
<td>2 2.9</td>
<td>0 1.5</td>
<td>15 9.9</td>
<td>4.8</td>
</tr>
<tr>
<td>17. 21 14.5</td>
<td>0 2.9</td>
<td>3 1.5</td>
<td>5 9.9</td>
<td>9.7</td>
</tr>
<tr>
<td>18. 21 14.0</td>
<td>3 2.8</td>
<td>0 1.4</td>
<td>4 9.5</td>
<td>8.1</td>
</tr>
<tr>
<td>19. 4 6.5</td>
<td>3 1.1</td>
<td>0 .6</td>
<td>4 3.7</td>
<td>4.9</td>
</tr>
<tr>
<td>20. 32 16.5</td>
<td>0 3.3</td>
<td>0 1.7</td>
<td>1 11.2</td>
<td>28.9</td>
</tr>
<tr>
<td>21. 10 9.0</td>
<td>0 1.8</td>
<td>2 .9</td>
<td>6 6.1</td>
<td>3.2</td>
</tr>
<tr>
<td>22. 21 17.5</td>
<td>5 3.5</td>
<td>1 1.8</td>
<td>8 11.9</td>
<td>3.0</td>
</tr>
<tr>
<td>23. 17 15.5</td>
<td>6 3.1</td>
<td>1 1.6</td>
<td>7 10.5</td>
<td>4.2</td>
</tr>
<tr>
<td>24. 20 13.5</td>
<td>0 2.7</td>
<td>2 1.4</td>
<td>5 9.2</td>
<td>8.0</td>
</tr>
<tr>
<td>25. 5 4.5</td>
<td>1 .9</td>
<td>0 .5</td>
<td>3 3.1</td>
<td>.6</td>
</tr>
<tr>
<td>26. 10 16.0</td>
<td>5 3.2</td>
<td>0 1.6</td>
<td>17 10.9</td>
<td>8.3</td>
</tr>
<tr>
<td>27. 11 16.5</td>
<td>5 3.3</td>
<td>3 1.7</td>
<td>14 11.2</td>
<td>4.4</td>
</tr>
<tr>
<td>28. 12 12.0</td>
<td>1 2.4</td>
<td>0 1.2</td>
<td>11 8.2</td>
<td>3.0</td>
</tr>
<tr>
<td>29. 11 9.0</td>
<td>4 1.8</td>
<td>2 .9</td>
<td>1 6.1</td>
<td>8.7</td>
</tr>
<tr>
<td>30. 8 10.0</td>
<td>2 2.0</td>
<td>0 1.0</td>
<td>10 6.8</td>
<td>2.9</td>
</tr>
</tbody>
</table>

\( \chi^2 = 240.9 \)  
Critical value = 42.6  
\( H_0 \) : rejected
MODEL II. Inexperienced Relevance and Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>MATCH NOT ABOUT</th>
<th>NOT MATCH ABOUT</th>
<th>NOT MATCH NOT ABOUT</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>$\hat{n}$</td>
<td>$x$</td>
<td>$\hat{m}$</td>
<td>$x$</td>
</tr>
<tr>
<td>1.</td>
<td>3</td>
<td>4.5</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>2.</td>
<td>4</td>
<td>3.0</td>
<td>0</td>
<td>.7</td>
</tr>
<tr>
<td>3.</td>
<td>19</td>
<td>13.0</td>
<td>0</td>
<td>2.9</td>
</tr>
<tr>
<td>4.</td>
<td>28</td>
<td>16.5</td>
<td>3</td>
<td>3.6</td>
</tr>
<tr>
<td>5.</td>
<td>6</td>
<td>12.0</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>6.</td>
<td>4</td>
<td>14.0</td>
<td>3</td>
<td>3.1</td>
</tr>
<tr>
<td>7.</td>
<td>26</td>
<td>17.0</td>
<td>0</td>
<td>3.7</td>
</tr>
<tr>
<td>8.</td>
<td>20</td>
<td>16.0</td>
<td>.8</td>
<td>3.5</td>
</tr>
<tr>
<td>9.</td>
<td>23</td>
<td>15.0</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>10.</td>
<td>14</td>
<td>8.5</td>
<td>0</td>
<td>1.9</td>
</tr>
<tr>
<td>11.</td>
<td>13</td>
<td>14.0</td>
<td>5</td>
<td>3.1</td>
</tr>
<tr>
<td>12.</td>
<td>9</td>
<td>13.5</td>
<td>0</td>
<td>3.0</td>
</tr>
<tr>
<td>13.</td>
<td>7</td>
<td>7.5</td>
<td>4</td>
<td>1.7</td>
</tr>
<tr>
<td>14.</td>
<td>11</td>
<td>12.5</td>
<td>1</td>
<td>2.8</td>
</tr>
<tr>
<td>15.</td>
<td>4</td>
<td>13.0</td>
<td>3</td>
<td>2.9</td>
</tr>
<tr>
<td>16.</td>
<td>15</td>
<td>14.5</td>
<td>1</td>
<td>3.2</td>
</tr>
<tr>
<td>17.</td>
<td>13</td>
<td>14.5</td>
<td>3</td>
<td>3.2</td>
</tr>
<tr>
<td>18.</td>
<td>17</td>
<td>14.0</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td>19.</td>
<td>8</td>
<td>5.5</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td>20.</td>
<td>12</td>
<td>16.5</td>
<td>7</td>
<td>3.6</td>
</tr>
<tr>
<td>21.</td>
<td>10</td>
<td>9.0</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>22.</td>
<td>9</td>
<td>17.5</td>
<td>2</td>
<td>3.9</td>
</tr>
<tr>
<td>23.</td>
<td>14</td>
<td>15.5</td>
<td>3</td>
<td>3.4</td>
</tr>
<tr>
<td>24.</td>
<td>23</td>
<td>13.5</td>
<td>0</td>
<td>3.0</td>
</tr>
<tr>
<td>25.</td>
<td>1</td>
<td>4.5</td>
<td>6</td>
<td>1.0</td>
</tr>
<tr>
<td>26.</td>
<td>17</td>
<td>16.0</td>
<td>6</td>
<td>3.5</td>
</tr>
<tr>
<td>27.</td>
<td>14</td>
<td>16.5</td>
<td>2</td>
<td>3.6</td>
</tr>
<tr>
<td>28.</td>
<td>11</td>
<td>12.0</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>29.</td>
<td>11</td>
<td>9.0</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>30.</td>
<td>10</td>
<td>10.0</td>
<td>1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

$\chi^2 = 242.5$ Critical value = 42.6 $H_0$ : rejected
The results of the chi square tests for each of the five categories under MODEL II can be summarized as:

1. Subject specialists. Category 1. 199.7
2. Experienced. Category 2. 352.0
3. Inexperienced. Category 3. 306.2
4. Experienced. Category 4. 240.9
5. Inexperienced. Category 5 242.5

With a critical value of 42.6 at $\alpha = .05$, it is concluded that the null hypothesis of independence of ABOUT and MATCH from the QUERIES is rejected. However, this model assumes an even distribution of ABOUT and MATCH across QUERIES. The contingency tables show that there is wide variation across the QUERIES in proportions of ABOUT/MATCH and NOT ABOUT/NOT MATCH - i.e., Cells 1 and 4. The next model, MODEL III, takes this uneven distribution into account.
MODEL III. This model tests whether aboutness is independent of the query given fixed match or not match:

\[ H_0: \frac{(X_{ij+})(X_{jk+})}{X_{+j+}} \]

where

- \( M_{ik+|j} \) is the expected value for each cell;
- \( X_{11+} \) is the total number of matching and about documents;
- \( X_{12+} \) is the total number of not about and matching documents;
- \( X_{21+} \) is the total number of about and not matching documents;
- \( X_{22+} \) is the total number of not about and not matching documents;
- \( X_{+1k} \) is the total number of matching documents in the \( k^{th} \) query, \( k = 1, \ldots, n \);
- \( X_{+2k} \) is the total number of not matching documents in the \( k^{th} \) query, \( k = 1, \ldots, n \);
- \( X_{+1+} \) is the total number of matching documents;
- \( X_{+2+} \) is the total number of not matching documents;

and

Degrees of Freedom = \((K-1)(I-L)(J-1)\)

= \((29)(1)(1)\) = 58

The critical value at \( \alpha = .05 \) is 77.

The five tables which follow show the observed and expected values for each cell.
MODEL III. Subject Specialists' Relevance and Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>X</td>
<td>M</td>
</tr>
<tr>
<td>1.</td>
<td>6 5.9</td>
</tr>
<tr>
<td>2.</td>
<td>5 5.0</td>
</tr>
<tr>
<td>3.</td>
<td>21 20.8</td>
</tr>
<tr>
<td>4.</td>
<td>31 30.7</td>
</tr>
<tr>
<td>5.</td>
<td>12 11.9</td>
</tr>
<tr>
<td>6.</td>
<td>23 22.8</td>
</tr>
<tr>
<td>7.</td>
<td>20 19.8</td>
</tr>
<tr>
<td>8.</td>
<td>10 9.9</td>
</tr>
<tr>
<td>9.</td>
<td>14 13.9</td>
</tr>
<tr>
<td>10.</td>
<td>16 15.8</td>
</tr>
<tr>
<td>11.</td>
<td>11 10.9</td>
</tr>
<tr>
<td>12.</td>
<td>22 21.8</td>
</tr>
<tr>
<td>13.</td>
<td>12 11.9</td>
</tr>
<tr>
<td>14.</td>
<td>16 15.8</td>
</tr>
<tr>
<td>15.</td>
<td>9 8.9</td>
</tr>
<tr>
<td>16.</td>
<td>15 15.8</td>
</tr>
<tr>
<td>17.</td>
<td>19 19.8</td>
</tr>
<tr>
<td>18.</td>
<td>22 21.8</td>
</tr>
<tr>
<td>19.</td>
<td>8 7.9</td>
</tr>
<tr>
<td>20.</td>
<td>16 15.8</td>
</tr>
<tr>
<td>21.</td>
<td>13 12.9</td>
</tr>
<tr>
<td>22.</td>
<td>13 12.9</td>
</tr>
<tr>
<td>23.</td>
<td>25 24.8</td>
</tr>
<tr>
<td>24.</td>
<td>23 22.8</td>
</tr>
<tr>
<td>25.</td>
<td>7 6.9</td>
</tr>
<tr>
<td>26.</td>
<td>31 30.7</td>
</tr>
<tr>
<td>27.</td>
<td>16 15.8</td>
</tr>
<tr>
<td>28.</td>
<td>11 12.9</td>
</tr>
<tr>
<td>29.</td>
<td>13 14.9</td>
</tr>
<tr>
<td>30.</td>
<td>14 13.9</td>
</tr>
</tbody>
</table>

\( \chi^2 = 57.1 \)  \quad \text{Critical value} = 77 \quad H_0: \text{Accepted}
MODEL III. Experienced Relevance and Subject Specialists’ Query Judgements.

**NUMBER OF DOCUMENTS**

<table>
<thead>
<tr>
<th>MATCH</th>
<th>ABOUT</th>
<th>NOT ABOUT</th>
<th>ABOUT</th>
<th>NOT ABOUT</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>M</td>
<td>X</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>4</td>
<td>4.0</td>
<td>2</td>
<td>2.0</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>3</td>
<td>3.4</td>
<td>2</td>
<td>1.7</td>
<td>0</td>
</tr>
<tr>
<td>3.</td>
<td>13</td>
<td>14.1</td>
<td>8</td>
<td>6.9</td>
<td>3</td>
</tr>
<tr>
<td>4.</td>
<td>17</td>
<td>20.8</td>
<td>14</td>
<td>10.2</td>
<td>1</td>
</tr>
<tr>
<td>5.</td>
<td>7</td>
<td>7.4</td>
<td>4</td>
<td>3.6</td>
<td>3</td>
</tr>
<tr>
<td>6.</td>
<td>8</td>
<td>15.4</td>
<td>15</td>
<td>7.6</td>
<td>0</td>
</tr>
<tr>
<td>7.</td>
<td>11</td>
<td>13.4</td>
<td>9</td>
<td>6.6</td>
<td>2</td>
</tr>
<tr>
<td>8.</td>
<td>6</td>
<td>6.7</td>
<td>4</td>
<td>3.3</td>
<td>10</td>
</tr>
<tr>
<td>9.</td>
<td>7</td>
<td>9.4</td>
<td>7</td>
<td>4.6</td>
<td>5</td>
</tr>
<tr>
<td>10.</td>
<td>6</td>
<td>10.7</td>
<td>10</td>
<td>5.3</td>
<td>1</td>
</tr>
<tr>
<td>11.</td>
<td>8</td>
<td>8.0</td>
<td>4</td>
<td>4.0</td>
<td>9</td>
</tr>
<tr>
<td>12.</td>
<td>10</td>
<td>14.7</td>
<td>12</td>
<td>7.3</td>
<td>0</td>
</tr>
<tr>
<td>13.</td>
<td>7</td>
<td>8.0</td>
<td>5</td>
<td>4.0</td>
<td>0</td>
</tr>
<tr>
<td>14.</td>
<td>12</td>
<td>11.4</td>
<td>5</td>
<td>5.6</td>
<td>0</td>
</tr>
<tr>
<td>15.</td>
<td>10</td>
<td>6.7</td>
<td>0</td>
<td>3.3</td>
<td>14</td>
</tr>
<tr>
<td>16.</td>
<td>11</td>
<td>10.7</td>
<td>5</td>
<td>5.3</td>
<td>1</td>
</tr>
<tr>
<td>17.</td>
<td>21</td>
<td>14.1</td>
<td>0</td>
<td>6.9</td>
<td>3</td>
</tr>
<tr>
<td>18.</td>
<td>17</td>
<td>14.7</td>
<td>5</td>
<td>7.3</td>
<td>4</td>
</tr>
<tr>
<td>19.</td>
<td>4</td>
<td>5.4</td>
<td>4</td>
<td>2.6</td>
<td>0</td>
</tr>
<tr>
<td>20.</td>
<td>16</td>
<td>10.7</td>
<td>0</td>
<td>5.3</td>
<td>16</td>
</tr>
<tr>
<td>21.</td>
<td>10</td>
<td>8.7</td>
<td>3</td>
<td>4.3</td>
<td>2</td>
</tr>
<tr>
<td>22.</td>
<td>11</td>
<td>8.0</td>
<td>1</td>
<td>4.0</td>
<td>11</td>
</tr>
<tr>
<td>23.</td>
<td>17</td>
<td>16.8</td>
<td>8</td>
<td>8.3</td>
<td>3</td>
</tr>
<tr>
<td>24.</td>
<td>23</td>
<td>15.4</td>
<td>0</td>
<td>7.6</td>
<td>1</td>
</tr>
<tr>
<td>25.</td>
<td>5</td>
<td>4.7</td>
<td>2</td>
<td>2.3</td>
<td>0</td>
</tr>
<tr>
<td>26.</td>
<td>16</td>
<td>20.8</td>
<td>15</td>
<td>10.2</td>
<td>0</td>
</tr>
<tr>
<td>27.</td>
<td>11</td>
<td>10.7</td>
<td>5</td>
<td>5.3</td>
<td>3</td>
</tr>
<tr>
<td>28.</td>
<td>12</td>
<td>9.4</td>
<td>2</td>
<td>3.3</td>
<td>3</td>
</tr>
<tr>
<td>29.</td>
<td>11</td>
<td>10.1</td>
<td>4</td>
<td>4.0</td>
<td>2</td>
</tr>
<tr>
<td>30.</td>
<td>8</td>
<td>9.4</td>
<td>6</td>
<td>3.3</td>
<td>0</td>
</tr>
</tbody>
</table>

\( \chi^2 = 167.1 \) \hspace{1cm} \text{Critical value} = 77 \hspace{1cm} H_o: \text{rejected}
MODEL III. Inexperienced Relevance and Subject Specialists' Query Judgements.

### NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
<th>MATCH</th>
<th>NOT MATCH</th>
<th>MATCH</th>
<th>NOT MATCH</th>
<th>MATCH</th>
<th>NOT MATCH</th>
<th>MATCH</th>
<th>NOT MATCH</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 3</td>
<td>4.0</td>
<td>2 2</td>
<td>0 3.3</td>
<td>1 2</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.8</td>
</tr>
<tr>
<td>2. 4</td>
<td>3.3</td>
<td>1 1.7</td>
<td>0 3.3</td>
<td>1 1.7</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.7</td>
</tr>
<tr>
<td>3. 16</td>
<td>13.8</td>
<td>0 3.4</td>
<td>5 1.7</td>
<td>5 7.1</td>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. 28</td>
<td>20.5</td>
<td>0 1.3</td>
<td>2 7.3</td>
<td>3 10.5</td>
<td>11.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. 6</td>
<td>7.3</td>
<td>9 8.7</td>
<td>4 4.3</td>
<td>5 3.7</td>
<td>7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. 5</td>
<td>15.2</td>
<td>5 3.4</td>
<td>6 7.3</td>
<td>18 7.8</td>
<td>22.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. 20</td>
<td>17.2</td>
<td>8 5.4</td>
<td>0 2.6</td>
<td>6 8.8</td>
<td>5.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. 5</td>
<td>6.6</td>
<td>6 14.7</td>
<td>16 7.3</td>
<td>5 3.4</td>
<td>16.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. 13</td>
<td>9.2</td>
<td>5 10.7</td>
<td>11 5.3</td>
<td>1 4.8</td>
<td>13.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. 14</td>
<td>9.2</td>
<td>3 2.0</td>
<td>0 1.0</td>
<td>0 4.8</td>
<td>8.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. 9</td>
<td>8.6</td>
<td>9 10.1</td>
<td>6 5.0</td>
<td>4 4.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.3</td>
</tr>
<tr>
<td>12. 9</td>
<td>14.5</td>
<td>5 3.4</td>
<td>0 1.7</td>
<td>13 7.5</td>
<td>8.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. 7</td>
<td>7.9</td>
<td>3 2.0</td>
<td>0 1.0</td>
<td>5 4.1</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. 11</td>
<td>9.9</td>
<td>8 6.7</td>
<td>2 3.3</td>
<td>4 5.1</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. 4</td>
<td>6.6</td>
<td>11 10.7</td>
<td>5 5.3</td>
<td>6 3.4</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. 12</td>
<td>10.6</td>
<td>10 8.7</td>
<td>3 4.3</td>
<td>4 5.4</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. 13</td>
<td>13.9</td>
<td>7 5.3</td>
<td>1 2.6</td>
<td>8 7.1</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. 13</td>
<td>14.5</td>
<td>0 4.0</td>
<td>6 2.0</td>
<td>9 7.5</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. 8</td>
<td>5.3</td>
<td>3 2.0</td>
<td>0 1.0</td>
<td>0 2.7</td>
<td>5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. 9</td>
<td>10.6</td>
<td>12 11.4</td>
<td>5 5.6</td>
<td>7 5.4</td>
<td>.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. 10</td>
<td>8.6</td>
<td>5 3.4</td>
<td>0 1.7</td>
<td>3 4.4</td>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. 6</td>
<td>7.9</td>
<td>17 15.4</td>
<td>6 7.6</td>
<td>6 4.1</td>
<td>1.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. 13</td>
<td>16.5</td>
<td>3 4.0</td>
<td>3 2.0</td>
<td>12 8.5</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. 20</td>
<td>15.2</td>
<td>2 2.7</td>
<td>2 1.3</td>
<td>3 7.8</td>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. 1</td>
<td>4.6</td>
<td>2 1.3</td>
<td>0 7</td>
<td>6 2.4</td>
<td>9.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. 20</td>
<td>20.5</td>
<td>1 1.7</td>
<td>0 3</td>
<td>11 10.5</td>
<td>.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. 14</td>
<td>10.6</td>
<td>16 11.4</td>
<td>1 5.6</td>
<td>2 5.4</td>
<td>8.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. 11</td>
<td>8.6</td>
<td>9 7.4</td>
<td>2 3.6</td>
<td>2 4.4</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. 11</td>
<td>9.9</td>
<td>1 2.0</td>
<td>2 1.0</td>
<td>4 5.1</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. 10</td>
<td>9.2</td>
<td>5 9.4</td>
<td>1 2.0</td>
<td>5 4.8</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \chi^2 = 170.5 \)  Critical value = .77  \( H_0 \) : rejected
MODEL III. Experienced Relevance and Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH ABOUT</th>
<th>MATCH NOT ABOUT</th>
<th>NOT MATCH ABOUT</th>
<th>NOT MATCH NOT ABOUT</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>$\hat{X}$</td>
<td>X</td>
<td>$\hat{X}$</td>
<td>X</td>
</tr>
<tr>
<td>1.</td>
<td>4.0</td>
<td>2.0</td>
<td>1.0</td>
<td>0.4</td>
</tr>
<tr>
<td>2.</td>
<td>3.3</td>
<td>1.0</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>3.</td>
<td>15.8</td>
<td>4.0</td>
<td>3.2</td>
<td>1.0</td>
</tr>
<tr>
<td>4.</td>
<td>17.25.7</td>
<td>14.0</td>
<td>5.3</td>
<td>1.0</td>
</tr>
<tr>
<td>5.</td>
<td>7.5</td>
<td>0.0</td>
<td>1.2</td>
<td>3.0</td>
</tr>
<tr>
<td>6.</td>
<td>5.6</td>
<td>3.0</td>
<td>1.4</td>
<td>3.0</td>
</tr>
<tr>
<td>7.</td>
<td>10.13.3</td>
<td>6.0</td>
<td>2.7</td>
<td>3.0</td>
</tr>
<tr>
<td>8.</td>
<td>14.13.3</td>
<td>2.0</td>
<td>2.7</td>
<td>0.0</td>
</tr>
<tr>
<td>9.</td>
<td>9.10.0</td>
<td>3.0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>10.</td>
<td>6.7.5</td>
<td>3.0</td>
<td>1.5</td>
<td>1.0</td>
</tr>
<tr>
<td>11.</td>
<td>17.14.1</td>
<td>0.0</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>12.</td>
<td>10.8.3</td>
<td>0.0</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>13.</td>
<td>6.6.6</td>
<td>2.0</td>
<td>1.4</td>
<td>1.0</td>
</tr>
<tr>
<td>14.</td>
<td>9.7.5</td>
<td>0.0</td>
<td>1.5</td>
<td>3.0</td>
</tr>
<tr>
<td>15.</td>
<td>24.19.9</td>
<td>0.0</td>
<td>4.1</td>
<td>0.0</td>
</tr>
<tr>
<td>16.</td>
<td>12.11.6</td>
<td>2.0</td>
<td>2.4</td>
<td>0.0</td>
</tr>
<tr>
<td>17.</td>
<td>21.17.4</td>
<td>0.0</td>
<td>3.6</td>
<td>3.0</td>
</tr>
<tr>
<td>18.</td>
<td>21.19.9</td>
<td>3.0</td>
<td>4.1</td>
<td>0.0</td>
</tr>
<tr>
<td>19.</td>
<td>4.5.8</td>
<td>3.0</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>20.</td>
<td>32.26.6</td>
<td>0.0</td>
<td>5.4</td>
<td>0.0</td>
</tr>
<tr>
<td>21.</td>
<td>10.8.3</td>
<td>0.0</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>22.</td>
<td>17.21.6</td>
<td>6.0</td>
<td>4.4</td>
<td>1.0</td>
</tr>
<tr>
<td>23.</td>
<td>20.19.1</td>
<td>0.0</td>
<td>3.9</td>
<td>2.0</td>
</tr>
<tr>
<td>24.</td>
<td>21.16.6</td>
<td>5.0</td>
<td>3.4</td>
<td>1.0</td>
</tr>
<tr>
<td>25.</td>
<td>5.5.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.4</td>
</tr>
<tr>
<td>26.</td>
<td>10.12.5</td>
<td>5.0</td>
<td>2.6</td>
<td>0.0</td>
</tr>
<tr>
<td>27.</td>
<td>11.13.3</td>
<td>5.0</td>
<td>2.7</td>
<td>3.0</td>
</tr>
<tr>
<td>28.</td>
<td>12.10.8</td>
<td>1.0</td>
<td>2.2</td>
<td>0.0</td>
</tr>
<tr>
<td>29.</td>
<td>11.12.5</td>
<td>4.0</td>
<td>2.6</td>
<td>2.0</td>
</tr>
<tr>
<td>30.</td>
<td>8.8.3</td>
<td>2.0</td>
<td>1.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>

$\chi^2 = 109.3$ Critical value = 7.7 $H_o$ rejected
MODEL III. Inexperienced Relevance and Query Judgements.

NUMBER OF DOCUMENTS

<table>
<thead>
<tr>
<th>MATCH</th>
<th>NOT MATCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOUT</td>
<td>NOT ABOUT</td>
</tr>
<tr>
<td>( X )</td>
<td>( \hat{X} )</td>
</tr>
<tr>
<td>1.</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>4</td>
</tr>
<tr>
<td>3.</td>
<td>19</td>
</tr>
<tr>
<td>4.</td>
<td>28</td>
</tr>
<tr>
<td>5.</td>
<td>6</td>
</tr>
<tr>
<td>6.</td>
<td>4</td>
</tr>
<tr>
<td>7.</td>
<td>26</td>
</tr>
<tr>
<td>8.</td>
<td>20</td>
</tr>
<tr>
<td>9.</td>
<td>23</td>
</tr>
<tr>
<td>10.</td>
<td>14</td>
</tr>
<tr>
<td>11.</td>
<td>13</td>
</tr>
<tr>
<td>12.</td>
<td>9</td>
</tr>
<tr>
<td>13.</td>
<td>7</td>
</tr>
<tr>
<td>14.</td>
<td>11</td>
</tr>
<tr>
<td>15.</td>
<td>4</td>
</tr>
<tr>
<td>16.</td>
<td>15</td>
</tr>
<tr>
<td>17.</td>
<td>13</td>
</tr>
<tr>
<td>18.</td>
<td>17</td>
</tr>
<tr>
<td>19.</td>
<td>8</td>
</tr>
<tr>
<td>20.</td>
<td>12</td>
</tr>
<tr>
<td>21.</td>
<td>10</td>
</tr>
<tr>
<td>22.</td>
<td>9</td>
</tr>
<tr>
<td>23.</td>
<td>14</td>
</tr>
<tr>
<td>24.</td>
<td>23</td>
</tr>
<tr>
<td>25.</td>
<td>1</td>
</tr>
<tr>
<td>26.</td>
<td>17</td>
</tr>
<tr>
<td>27.</td>
<td>14</td>
</tr>
<tr>
<td>28.</td>
<td>11</td>
</tr>
<tr>
<td>29.</td>
<td>11</td>
</tr>
<tr>
<td>30.</td>
<td>10</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 120.4 \]  \quad \text{Critical value} = 77  \quad H_0: \text{rejected}
The results of the chi square tests for each of the five categories under MODEL III can be summarized as:

1. Subject Specialists. Category 1. 57.1
2. Experienced. Category 2. 167.1
3. Inexperienced. Category 3. 170.5
5. Inexperienced. Category 5. 120.4

With a critical value of 77 at $\alpha = .05$, it is concluded that:

1. for the Subject Specialists, the ABOUT/MATCH interaction is independent of the query; and
2. for the Experienced and Inexperienced judgements, there is significant variation among the queries.

Therefore, it is concluded that the null hypothesis of independence of aboutness from the query is accepted for the Subject Specialists and rejected for the Experienced and Inexperienced judges.

All of the models used so far have included all three variables. The next model tests the independence of ABOUT and MATCH alone. That is, MODEL IV measures the association between the judges decisions and the functional relations decisions which are summed over all the queries for each of the four cells.
MODEL IV. This model tests independence of About and Match:

\[ H_0 : \hat{M}_{ij} = \frac{(X_{i++})(X_{+j+})}{N} \]

where

\( \hat{M}_{ij} \) = the expected value for each cell;

\( X_{i++} \) = the total number of about documents;

\( X_{2++} \) = the total number of not about documents;

\( X_{+1+} \) = the total number of matching documents;

\( X_{+2+} \) = the total number of not matching documents;

\( X_{+++} \) = the total number of documents in the \( k^{th} \) query, \( k = 1, ..., n \);

\( N = X_{+++} \)

and

Degrees of Freedom = \((I-1)(J-1) = (1)(1) = 1.\)

At \( \alpha = .05 \) the critical value for this model is 3.8.
MODEL IV.

The tables for the five contingency tables are given as follows:

1. Subject Specialists' Relevance and Query Judgements;
2. Experienced Relevance and Subject Specialists' Query Judgements;
3. Inexperienced Relevance and Subject Specialists' Query Judgements;
4. Experienced Relevance and Query Judgements; and
5. Inexperienced Relevance and Query Judgements.

<table>
<thead>
<tr>
<th>NUMBER OF DOCUMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATCH ABOUT</td>
</tr>
<tr>
<td>X</td>
</tr>
<tr>
<td>1. 474</td>
</tr>
<tr>
<td>2. 322</td>
</tr>
<tr>
<td>3. 321</td>
</tr>
<tr>
<td>4. 371</td>
</tr>
<tr>
<td>5. 372</td>
</tr>
</tbody>
</table>

With a critical value of 3.8 at α = .05, the null hypothesis is rejected for all five categories. That is, there is a significant relationship between the decisions of the judges and the functional relations decisions.
5.2. The Functional Relations Tests.

The agreement between the researcher, acting as first language analyst, and the second language analyst was measured using simple ratios. The table below summarizes the statistics for the overall agreement for Queries and Abstracts.

<table>
<thead>
<tr>
<th>Case Roles</th>
<th>Queries</th>
<th>Abstracts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AGREE</td>
<td>DISAGREE</td>
</tr>
<tr>
<td>Agent</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Instrument</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Source</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Goal</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Object</td>
<td>42</td>
<td>24</td>
</tr>
<tr>
<td>Location</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verb Types</th>
<th>Queries</th>
<th>Abstracts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AGREE</td>
<td>DISAGREE</td>
</tr>
<tr>
<td>State</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Action</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>Process</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Action/Process</td>
<td>8</td>
<td>23</td>
</tr>
</tbody>
</table>

The overall agreement for both queries and abstracts was 65.6%. A look at the table shows that much of the disagreement between the two analysts is over the case roles
Object and Goal, and the verb type Action-Process, where there was only 24% agreement for Goal and Object and 21.5% agreement for Action-Process verbs. With these three functional relations removed from the total, the agreement between the language analysts is 84%.

The test of the agreement on Match/Not Match of statements in abstracts and queries was measured using a simple ratio. The agreement between the researcher and the language analyst for this test was 87%.

The general conclusion for the test of agreement in functional relations assignments is that, with the exception of the above noted case roles and verb type, the degree of agreement between the two analysts indicates that functional relations can be recognized in complex text and that most of the definitions were fairly clear, although the disagreements indicated that some clarification of these definitions for extension to complex text is required. This issue is discussed more fully in Chapter 6.

The general conclusion for the test of agreement on Match/Not Match of abstracts and queries is that the ability of a person other than the subject specialist to recognize about and not about abstracts is improved when the second person is given the query statements used by the subject specialist in making decisions, rather than the
whole text of the query. Many of the abstracts over which the second language analyst and the researcher disagreed were ambiguous - i.e., the relationships between the key terms were implied but not stated. This kind of ambiguity can probably only be resolved by a user of a system under any circumstances. The important point here is that functional relations alone could not be used to resolve all aboutness decisions, but were useful in most other decisions.
CHAPTER 6

EVALUATION OF THE FUNCTIONAL RELATIONS MODEL.

The general purpose of the analysis of functional relations in the queries and abstracts was to determine the strength of the relationship between functional relations between keyterms and the aboutness decisions of the various judges. The purpose of the discussion in this chapter is to describe the language patterns which emerged from the analysis of the data and from the analysis procedure itself.

In general, as stated previously, the case roles and verb types used here were adequate for describing the relations between keyterms in query statements and in abstracts and for discerning similarities or differences between statements. The bibliographic retrieval system provided a particularly interesting environment for this test because of the wide variety of subject environments included among the queries - e.g., psychology, microbiology, literature, social sciences, etc. Not only was the functional relations model usually adequate for describing the functional relations between keyterms in all of these subject areas, but the model also provided a means of explaining the different ways in which subjects are described and the factors which influence these descriptions.
However, as noted in Chapter 4, while the functional relations proved to be adequate, problems arose during the analysis which required reconsideration of the definitions of the case roles and verb types. The greatest single problem was extension of the definitions of the functional relations definitions provided by Fillmore to the vocabulary and language structures found in the queries and abstracts.

The definitions of the case roles and verb types used by the researcher when analyzing the functional relations between keyterms in query and abstract statements are described in Chapter 4. The purpose of this chapter is to discuss the issues raised concerning some of these definitions during the analysis. The problems encountered were due not only to the definitions themselves but to the application of the definitions to the wide diversity of subjects and kinds of nouns and verbs which appeared in the queries and abstracts, the importance of world knowledge for interpretation of the nature of the relations between keyterms, and the relationship between the functional relations between keyterms implied by the natural language expressions which occurred.

The first, and most important, conclusion drawn from the analysis was that the case roles and verb types were adequate for describing all of the relationships between keyterms in query and abstract statements with one exception
which will be discussed. That is, no additions to the list of case roles and verb types is required. However, the consistent application of the model was more difficult. Several problems in the interpretation of the definitions arose which have serious implications of the use of such a model or development of a useful model for a bibliographic retrieval system.

The discussion in this chapter will be limited to the problems inherent in the model which were perceived during the analysis, to considerations which must be addressed in any attempt to solve these problems, and to possible alternative solutions. Otherwise, modification of the model itself is not attempted here. Because the model was so highly descriptive of a language behavior pattern in aboutness recognition and because some method of using functional relations indicators has long been of interest in bibliographic retrieval system research, the nature of the successes and failures of the model are important. These successes and failures clearly indicated issues in language expression in the bibliographic retrieval system environment which must be solved if some method of indicating functional relations is to be used successfully in the environment.
Section 1 of this chapter discusses the definitions and how various decisions were made during the analysis. This section also discusses some of the issues raised in the attempt to apply Fillmore's definitions to keyterms in queries and abstracts. As will be seen, there are language patterns which must be accounted for in any attempt to apply a functional relations model in the bibliographic retrieval environment.

Section 2 describes the language structures in the texts of the queries and abstracts which described certain logical relations holding between nominal elements or between events and how these relations were indicated in complex language expressions. The relations of special interest in the data were causative and affective relations and the ways in which the degrees of strength of these relations were indicated in the language structures expressing them.
6.1. The Definitions of the Case Roles and Verb Types.

The discussion in this section will begin with a description of problems involving the definition of case roles and verb types. Many of these problems were discussed by Fillmore (1971), but the nature of these problems is illustrated even more clearly by examples from the data in this study. The most interesting problems involved the definition of the Agent role and the interpretation of the functional relations between keyterms in statements containing Action-Process verbs which described causative or affective relations.
6.1.1. The Agent and Instrument roles.

The most interesting problem involving definitions of the Agent and Instrument roles occurred when attempting to assign these roles to entities which did not meet the requirements for Agency set down by Fillmore. The problem was further complicated by the fact that Agency in sentences containing Action verbs and Agency in sentences containing Action-Process verbs describing causation seemed to differ. This problem was most clearly demonstrated in statements involving viruses, plants, microbes and parasites.

Fillmore states in "Some problems for case grammar", "...I no longer confuse animacy with Agency...natural force nouns - i.e., phenomena which are not subject to anybody's control but which cause things to happen (e.g., lightening, tuberculosis or erosion) - should be absorbed into the Instrument case (1971, p. 43)," and, "...it is possible for the Instrument case to be occupied by a sentence but not the Agent (1971, p. 44)." The implication of these statements is that Agency involved intentionality. The problem for the analysis involved, then, sentences containing Action-process verbs where the instigator of the Action was an entity which was more akin to a natural force than to an entity which is capable of intentionality. (An interesting note is that not all cultures would consider this to be a problem.) The questions raised were:
1. Where does one draw the line between concepts which must be considered to be natural forces and concepts which have the ability to form intentions.

2. How does one deal with the fact that, while it is possible to assign the Instrument role to natural force nouns in sentences containing Action-Process verbs, it is not possible to assign the Instrument role to the same noun in a sentence which contains an Action verb since an Action verb does not accept an Instrument.

The following sentences drawn from one of the actual queries in this study illustrate how this problem might be dealt with in one instance. That is, they illustrate that the way in which a description is stated affects the assignment of cases as follows:

(1) Trypanosomes affect the immune response of the host.
(2) The antigenic variability of Trypanosomes affects the immune response of the host.
(3) Trypanosomes affect the immune response of the host by varying their antigenic properties.

We could say that all three of these sentences convey essentially the same general information - i.e., that the immune response of a host is affected in some way by Trypanosomes. However, the ways in which these statements are made - i.e., the linguistic structures of the sentences - change the roles which would be assigned. In order to
resolve this problem we must bring world knowledge to bear on the interpretation of these statements. According to Fillmore's conclusions, we cannot say that Trypanosomes act as Agents because they do not really instigate an Action. Rather, the variation in the antigens they produce is affected by the host's immune response - i.e., it is an automatic reaction on the part of the microbes much as when certain plants release chemicals which repel pests when they are damaged. However, Sentence (1) clearly indicates that Trypanosomes act as Agents. Sentence (3) goes so far as to indicate that Trypanosomes 'use' an Instrument, indicated by the phrase 'by varying their antigenic properties'. But our world knowledge indicates that the correct interpretation of the relations being expressed in these sentences corresponds more closely to Sentence (2) because we know that a microbe cannot intentionally use a property as an Instrument. They are, instead, reacting to a stimulus in an affective manner. In a sense, then, Fillmore's rule for natural forces provides a means of clarifying the true nature of relations between nominals of this kind.

The situation becomes more complicated, however, when the following statement is considered:

(4) Nematodes migrate from the point of entry to the ankle of a rabbit.

Here we have an animate entity which has no more awareness
of its actions than a **microbe**, but initiates an **Action** and is, therefore, an **Agent**. Of course, a **nematode** is a much more complex organism than a **Trypanosoma**, but the ability of a **nematode** to form an intention is still highly questionable.

So far this discussion has been conducted from the point of view that intentionality is a determinant in the assignment of **Agent** or **Instrument** to certain entities. The point of the case grammar model is to describe the relations holding between elements in a linguistic structure — i.e., a sentence. The point of this study was to determine how often the subject specialists discriminated consistently between statements in abstracts which matched their query statements and statements in abstracts which did not match their query statements in terms of the functional relations between the elements in the various statements. Reinterpretation of statements with a view to satisfying certain questions about the nature of entities would seem to subvert the purpose of the analysis. Because this question is not really within the scope of this study, the conclusions reached for dealing with this particular problem was to consider statements like Sentences (1) and (3) above to be metaphors and to treat the structure of the statements as they were given. Any indexing system which uses functional relations indicators, however, would have to deal with this
problem through clear definitions. An automatic language analysis system or automatic indexing system using functional relations would also have to deal with the problem of resolving several structures which contain the elements in functional relations which differ and which may not reflect the 'true' nature of the relations which world knowledge indicates should hold between the elements.
6.1.2. The Source and Goal roles.

The locative and temporal forms of the Source and Goal roles with verbs of motion or in the description of a process did not pose any problems in the analysis and do not require any further clarification. The other definitions of Goal do pose problems, first for the overall definition of the role and second for the assignment of the case role in sentences involving causation (discussed under Goal and Object which follows). The definitions which might be called into question include:

1. Goal as thing impinged on by an Action;
2. Goal as the end result or product of some Action or Process; and
3. Goal as the person or entity which receives some Object, or Goal as destination.

It seems reasonable to conclude that all of these definitions have a common feature - i.e., the idea of Goal as an end, either of a process or a movement. The creation of other roles to accommodate the various definitions of Goal is not necessary because only one of the types of Goal can appear in any one sentence, which satisfies one of the requirements of the grammar.
6.1.3. The Object and Goal roles.

The definitions of these two cases seemed to be both adequate and discrete when analyzing single statements, but comparison of one statement with another brought up a serious question. In the statement,

(4) (Someone) taught computer literacy to children.

Agent Action/Process Goal

Object

Children is the nominal element directly affected by the verb and, therefore, assumes the Object function. The definition of the Object as the thing changed or affected by the Action or Process described by the verb makes this assignment seem clear. Computer literacy as Goal is a bit more problematic, however. Literacy in computers can be seen as something that comes into existence as a result of teaching or as a later state of knowledge in the Object, children, but computer literacy can also be seen as something which is transferred from the teacher to the children, in which case computer literacy would be the Object and the children would assume the Goal function - the thing impinged upon by teaching, or as the receiver of the Object.

This situation is further complicated by the fact that it is not certain that teaching implies learning. Therefore, teach can be seen as either an Action-Process type verb or as an Action verb. The case frame for the two interpretations would be:
(4a) (Someone) - teaches - computer literacy -
   Agent     Action     Object
   to children.
   Goal

(4b) (Someone) - teaches - computer literacy -
   Agent     Action/Process  Goal
   to children.
   Object

The question of the verb type can be resolved simply by defining the lexical features of the verb teach. However, the fact remains that there is a kind of ambiguity in the interpretation of the roles which the two nominals, children and computer literacy, should assume in the statement.

The next sentence also illustrates the problems of assigning the Object and Goal roles in sentences containing Action-Process verbs, perhaps more clearly.

(5) Corona virus - causes - demyelinating disease -
   Agent     Action/Process   Goal
   in mice.
   Object

In this sentence demyelinating disease can only assume the Goal role - i.e., as the thing which comes into existence as a result of the Action described by the verb. However, mice can be seen as either the Object, the thing affected by the action described by the verb, or as the Goal, the thing impinged on by the action described by the verb. To further complicate this situation, the Location role could also be assigned to mice since demyelinating disease
describes an affect upon the myelin sheath and mice are simply the location where the affect occurs. For example, in the next sentence:

(6) Estrogen - affects - gene expression -
Instrument Action/Process Object
in mammal cells.
Location

where gene expression can be seen as the thing affected and mammal cells are simply the location of the genes. Mammal cells in this sentence could not be assigned the Goal case because they do not represent any of the three categories listed earlier in the discussion of Agency.

This confusion over the application of Goal, Object and Location roles is not due to a problem with the definition of the cases, but to a problem in the application of the model to sentences containing Action-Process verbs. Action-Process type verbs were very common in the query and abstract statements and were used to describe causative and affective relationship between nominal elements. The problem with Fillmore's definitions of the case roles arises from the fact that he includes affective or causative verbs and Action verbs under a more general category he calls 'impingement' verbs. The next two sentences were used by Fillmore to illustrate his argument:

(7) Mary - hit - the wall - with a stick.
Agent Action Goal Object
In this sentence, according to Fillmore, the wall is the thing impinged upon, the Goal, and the stick is the thing which moves, the Object.

(8) John - hit - the ball - over the fence.
    Agent     Action     Object     Goal

In this sentence Fillmore points out, the ball is both the Object, the thing which moves, and the thing impinged upon, the Goal. He then suggests that sentences like sentence (8) really contain both a stated action and an implied process. He suggests, but does not recommend necessarily, that such sentences should be divided into the two parts containing the two verbs, the Action verb and the resulting Process verb as in:

(8a) John - hit - the ball; the ball - went over the fence.
    Agent   Action     Goal       Object     Process

In sentence (8), then, as Fillmore notes, "...we have a situation in which one event serves as the immediate cause of some other event," and, further, "...the need to deal with causation as a consequence relation between two events comes up in the problem of determining the case structure of certain kinds of 'impingement' verbs (Fillmore, 1971, p. 46)."

Returning to sentences from the data, we have Sentence (5) which has the same structure as Sentence (8) and could, therefore, be restated as:
(9) (Someone) - teaches - children; children - Agent  Action  Goal  Object
learn - to be literate in computers.
Process  Goal

Sentences (6) and (7), on the other hand, do not contain 'impingement' verbs and so would not pose the same problem in the assignment of roles.

While there will be no conclusion drawn here about how to deal with verbs of 'impingement' in the application of the functional relations model, it is important to note that any attempt to use such a model in a bibliographic retrieval situation would require some decision for these verbs in order for any model to be applied consistently.

This suggestion is further supported by the fact that disagreements between the researcher and the second language analyst for these kinds of verbs and case roles showed that the definitions provided by Fillmore really tell us very little about which alternative should be used. There may also be other instances where Action-Process verbs would be better stated as two distinct sentences - i.e., where one even is the immediate cause of another event. The functional relations would then be between the two events, with the event describing the Action being assigned the Source role - i.e., the earlier state - and the event describing the Process being assigned the Goal role - i.e., the later state. (This point is discussed in more detail in Section 2 of this chapter.)
6.1.4. The Path Role.

The Path case role proved to have an interesting application in some of the query and abstract statements where a chain of events was involved in a causative or an affective relation describing the effect of a microbe on a host or of a chemical on a process. These chains of events were assigned the Path role by both the researcher and the second language analyst. While the word mechanism was used to describe this chain of events by the subject specialists, the mechanism was a chain of events and not a means or an immediate cause as is required of a nominal element assuming the Instrument case.

This application of the Path role satisfied the condition that the Path can include an infinite number of phrases describing various stages in time or space in a movement from a Source to a Goal. While Fillmore intended the Path role to be assigned to locations, it was decided that in the analysis performed here, the Path role was the most appropriate assignment to stages in a Process.
In summary, the discussion in this section described the most serious issues faced in the assignment of functional relations to the keyterms in query and abstract statements. The fact that the analysis encountered all of the problems discussed by Fillmore (1971) and only the problems discussed by Fillmore also indicates that language expressions in the sciences, social sciences, and humanities have the same structure as the simple sentences used for linguistic evidence. The next section, however, addresses problems which would appear to be inherent in language expression in the bibliographic retrieval environment which are not addressed by the case grammar model but which are very important in query negotiation, search strategy formation, and aboutness decision-making.
6.2. Language Behavior in Aboutness Decision-making.

In the bibliographic retrieval system environment a common relationship between keyterms in queries is a causal or affective relationship. The problems encountered in assigning functional roles to the keyterms in statements describing causal and affective relations were discussed in the preceding section. The purpose of this section is to discuss briefly language behavior patterns displayed by the subject specialists in their decisions involving functional relations of this type.

A causal relation will be considered here to be a relationship between two nominal elements where one nominal element is perceived to have an affect on the state or condition of another nominal element. In statements describing causal or affective relations, the verbal element describes the nature of the relationship. For example, the verb \textit{causes} indicates a direct causal relationship, while the verb \textit{associated with} indicates that A existed before B occurred and that A \textit{may} affect or cause B. The verbs \textit{is associated with} or \textit{is related to}, however, merely indicate that A and B are both present, but not that A necessarily affects or causes B. For example, the following query statements appeared in query and abstract statements for one of the queries:

\begin{verbatim}
(10) Leaf characteristics - resist - rust.
      Instrument Action/Process Object
\end{verbatim}
(11) Leaf characteristics - influence
Instrument  Action/Process
resistance to disease.
Object

(12) Leaf characteristics - are associated with
Object  State
resistance to rust.
Object

(13) Leaf characteristics - and - resistance to rust.
Object  State  Object

Sentences (10) and (11) describe an affective relationship between the two nominal elements leaf characteristics and resistance to rust. Sentences (12) and (13), on the other hand, merely indicate that plants with certain leaf characteristics also resist rust, but no proof of a relationship is being claimed. The subject specialist who submitted this query accepted abstracts as about which contained any of the above sentences.

A more complex example of this language pattern involving causal relationship occurred in the following query. The following statements illustrate a hierarchical nature in this pattern as in:

(14) Corona virus infection - causes
Instrument  Action/Process
demyelinating disease - in mice.
Goal  Object

(15) (Someone) - injected - corona virus
Agent  Action  Object
into mice; mice injected with corona virus
Goal  Object
developed - demyelinating disease.
Process  Goal
(16) Corona virus infection - is associated with
     Object
     State
     (the development of) - demyelinating disease -
     Object
     in mice.
     Goal

Sentence (14) contains an Action-Process type verb denoting a cause and effect relationship between the nominal elements *corona virus infection* and *demyelinating disease*. In Sentence (15) the Action-Process verb has been broken into its two parts - i.e., the first sentence describes the Action and the second the Process. Sentence (15) is clearly not equivalent to Sentence (14) because in Sentence (15) there is no stated direct relationship between *corona virus infection* and *the development of demyelinating disease*. However, Sentence (15) does indicate that *corona virus infection* preceded *the development of demyelinating disease*, so one condition of a causal relation - i.e., that the cause must precede the effect in time - has been satisfied. Therefore, Sentence (15) describes a stronger relationship between the two nominal elements than does Sentence (16) which really only states that the two events are present.

In the examples just given, the change in the verb type also results in a change in the functional roles assumed by the nominal elements. That is, in the sentences stating a causal relation - i.e., Sentences (10), (11), and (14) - the cause is an Instrument and the effect is either an Object.
or a Goal. In Sentence (15), on the other hand, we have a Source event and a Goal event - i.e., an earlier or a later state. In Sentences (12), (13), and (16) the nominal elements corona virus infection and development of demyelinating disease are merely conjoined Objects.

It could be concluded from the above patterns displayed by several subject specialists that people do not distinguish between these various degrees of specificity in stating causal relationships. However, the following statement from one of the subject specialists shows that there is every reason to believe that such distinctions are made:

"[I want] those articles that deal with hormonal regulation of gene expression in mechanistic or in molecular biological sense as opposed to - there's a great body of literature that deals with treating an animal or cells with X number of hormones and getting X level of gene expression, and there's absolutely nothing in between."

In other words, the statement of the relationship between hormones and gene expression had to include the relations:

(17) Hormones - affect - gene expression - by...
    Instrument Action/Process   Object   Path

for an abstract to be accepted as about by this subject specialist. The two alternative functional frames
describing consecutive events or simple association which were accepted by the previously discussed subject specialist were not accepted by the subject specialist who made the statement quoted.

Acceptance or refusal of alternative functional relations frames appears to be determined by the stage of the development of knowledge in a research area. Research in the relationship between hormones and gene expression had progressed to the point where the affective nature of the relationship was established. The relationship between corona virus and demyelinating disease, on the other hand, is not so well understood and investigation of the relationship between leaf characteristics and resistance to disease was very new at the time of the search. However, it also should be noted that the use of statements indicating consecutive (and possibly related) events or of statements indicating only an associative relationship also appeared in abstracts which then went on to describe a direct causal relationship. The use of these more cautious statements - i.e., statements describing something less than a direct causal relationship - would appear to be a matter of style and not a true indication of the known nature of the relationship between the nominal keyterms of interest.
In summary, the within type categories for verb types - i.e., causative or associative relations - were distinguished by the surrounding case frame and by the nature of the relations between the nominal elements indicated by the verbal frame. These categories do not indicate, however, that any change in the definitions of the verb types (such as the matrix proposed by Cook (1979)) is necessary because the case roles serve to distinguish between these different categories or relationships.

While the language expression patterns in queries and abstracts and in the language patterns of the aboutness decisions were well described by the functional relations based on Fillmore's case roles and verb types, there was also clear demonstration that the language used to express relationships between entities or events is often metaphorical. The use of metaphorical expressions appears to be based on economy of expression. That is, saying corona virus causes demyelinating disease is a summary statement of a complex chain of events between the development of corona virus infection and the development of demyelinating disease. World knowledge determines how much of that underlying process we understand when reading such a statement.

The conclusion reached on the basis of the data in this study is that simple statements of topics in queries and abstracts are more akin to general descriptions of what the topic is about than clear statements of fact. Often, the application of the rules for assignment of case roles and verb types served to clarify the real nature of the relationships among key terms or key phrases because the assignments were based on language structures actually appearing in the text. Further exploration of this language
is indicated by these results because they provide important information about the use of paraphrase and the structure of topic statements in the bibliographic retrieval environment.
CHAPTER 7

CONCLUSIONS

The results obtained in this study clearly demonstrate that simple co-occurrence of keyterms in an abstract, even where all keyterms have the desired meaning, is not sufficient for an abstract to be judged to be about a query. The conditions for a judgement of relevance under the definition used in this study were demonstrated by the subject specialists to be:

1. the keyterms in relevant abstracts had to have the meaning described in the query;

2. the keyterms in relevant abstracts had to be functionally related to each other in the ways described in the query; and

3. the required functional relations between keyterms were indicated in the query.

The analysis of queries and abstracts using the functional relations model demonstrated that the case roles and verb types were adequate, but not entirely satisfactory, for describing the relations between keyterms in queries and abstracts, and for distinguishing between the texts of abstracts judged to be about the query and abstracts judged to be not about the query.

In summary, abstracts judged to be about the topic(s) named in a query did contain the desired keyterms in
functional relations which matched the functional relations between the keyterms in the query. Abstracts judged to be not about the topic(s) named in the query did contain the same or equivalent keyterms, but not in functional relationships which matched the functional relationships described in the query.

The very high level of agreement between the subject specialists' relevance decisions and the decisions based on functional relations indicates that:

1. the subject specialists' descriptions of their topic(s) of interest were adequate for determining their intended topic, but not the specific query statements which would be accepted when they examined the abstracts;
2. the subject specialists were very consistent in their language behavior when discriminating between about and not about abstracts.

The high level of internal consistency in the relevance decisions of the secondary judges also demonstrated the consistency of language behavior in relevance assessments where the definition on which the decisions are made is aboutness or match in topic. The difficulty encountered by several of the secondary judges in determining which query statements were used by the subject specialists raises serious questions about the performance of someone other than the user in searching and in evaluating search
results. Problems in interpretation occurred at both the word level and the functional relations, or query statement level. That is, for some queries the secondary judge did not accept whole groups of keyterms as equivalent to a desired keyterm, even where the rejected quasi-synonyms had been mentioned by the subject specialist. For other queries, the secondary judges rejected query statements included in the query or accepted query statements which were not considered to be sufficient for relevance by the subject specialist. Of course, it is recognized that the secondary judges did not have the opportunity to discuss the queries with the subject specialists, so could not clarify terms or set limits on statements which they did not fully understand. A scan of the contingency tables indicates, however, that the majority of the secondary judges performed surprisingly well given the limits on the information they had available to them when making their relevance assessments.
7.1. Discussion of the Findings.

The conclusions described above were based strictly on the statistical results of this study and on observations of possible language behavior patterns found in the data analysis. The more general discussion which follows includes generalizations from the findings of this study to possible future research projects and is intended to suggest the implications of these findings for research in library and information science. Because this study was conducted within the limits of the on-line bibliographic retrieval system environment, this discussion and the suggestions arising are intended for the same environment.

The findings of this study and observations from the data have implications for several different areas of interest in bibliographic retrieval system research including:

1. relevance as a theory and as an operational definition;
2. testing of bibliographic retrieval systems;
3. query negotiation and search strategy formation;
4. indexing systems, and
5. improvement of search support mechanisms.

The most important implications arising from this research are for relevance as a theory and as an operational definition in bibliographic retrieval system research.
As was discussed at some length in the Introduction, each study of a bibliographic retrieval system which involves some form of 'relevance' assessments must define 'relevance' and how it is being operationalized in that particular study. The result of this situation is that the results of each study can be called into question by disagreement about the most valid definition of relevance alone.

The one principle that has remained essentially unquestioned is that 'relevance' assessments involve human judgements for each retrieved item, and that the best human judge is an actual user of the system. The drawback of this conclusion has been that it was also believed that the most commonly used definitions of relevance - e.g., aboutness, pertinence, and utility - are inherently subjective because they rely on human judgements. Another drawback at a more practical level is the difficulty and expense of obtaining subjects and relevance decisions in either an operating bibliographic retrieval system environment or in a laboratory setting using a test collection. Real users, it has been assumed, are expressing subjective relevance decisions which can vary from one time to another. The decisions of panels of experts, on the other hand, are suspect because they probably do not reflect real life decisions of real life users. And, finally, the choice of the definition of relevance on which
the performance measure will be based and the construction of the instructions to the judges who will be making the relevance decisions are fraught with concerns about the effect on validity of any results that arise from interfering in any way with the normal relevance assessments of users.

The surprising strength of the relationship between functional relations between keyterms and the aboutness decisions of the judges suggests that relevance assessments could be made using textual information alone or in combination with some form of user assessment. Development of the method for determining which query statements in a query are to be matched requires further testing. But such a method would require some systematic means of establishing both the meaning of key concepts and the functional relations between keyterms required for each query. How such information would be obtained is the central question because the validity of any study using this method would rest on the procedures used to construct the list of query statements used in relevance assessments. One possible solution would be to use a method similar to the one used for gathering data in this study. That is, aboutness assessments could be made on a sample of abstracts which contain a variety of quasi-synonyms for keyterms and a representative variety of functional relations between
the keyterms or quasi-synonyms. The researcher could then use these decisions as a basis for the evaluation of the remaining retrieved abstracts using textual information alone.

Determining relevance and nonrelevance of abstracts using the method described above has the advantage of providing a more objective means for determining relevance because the decisions would be independent of such considerations as author, source, currency, authority, etc. Use of such a basis for relevance assessments could be very helpful in studies of such aspects of bibliographic retrieval performance as:

1. assessments of indexing systems and practices,
2. exploration of the effects of different approaches to search strategy formation or query negotiation on retrieval performance;
3. creation of test sets of abstracts for laboratory experiments or for teaching purposes,
4. comparison of the retrieval performance of different data bases, and
5. searching queries which fall between two or more disciplines.

Determination of the query statements to be used and the language patterns required for relevance from a sample of retrieved records or even at the query negotiation stage would enable the researcher to later determine the relevance
of retrieved records independent of actual decisions of actual users for each abstract. Such a means of defining relevance would, of course, be limited to tests in areas of retrieval performance which are not directly concerned with user satisfaction.

Another interesting use of the findings of this study would be the development of query negotiation and search strategy formation techniques which exploit functional relations in the determination of the user's intended query or queries. It would appear that the more information the searcher obtains about the background of the subject to be searched and the uses to which the inquirer hopes to put the information retrieved, the better. The comparison between the final query statement, or summation of the topic of interest to the subject specialist, and the actual decisions of the subject specialists showed that background statements were also considered to be sufficient for an abstract to be accepted as about the topic. The common practice of relying on a summary statement as a guide in searching may be ill advised given the language behaviors found in this study. A clear understanding of the relative importance of each concept and the nature of the relationship between concepts e.g., causal relations, associative relations, etc. - could serve to aid the searcher in search strategy formation by
suggesting priorities to be considered when altering a search strategy in response to initial retrieval results.

The results of the test of the secondary judges' ability to anticipate the subject specialists' relevance decisions suggests that, at this point in time, the best alternative is to involve the user in the search process from the beginning of the search to the end. In other words, the user should be present during the search to evaluate the results as the search progresses. But this is a poor solution at best. The aim should be to develop systems which provide occasional users with the support necessary to help them develop their own search strategy. Of course, there is continuing research toward this end and the results of this study should provide some encouragement for those involved in such endeavors.

There are strong indications from the results of this study that use of pre-assigned role or relational indicators for indexing and searching is not likely to produce satisfactory results for several reasons. First, the variety of alternative functional relations accepted by the subject specialists in this study and the wide variety of ways in which these relations were expressed in the abstracts indicate that it would be unrealistic to expect at this point that indexers could possibly perceive all
possible relations between keyterms in abstracts under any conditions, let alone under the conditions in which indexing is accomplished at present. The most telling argument against the use of role indicators in pre-coordinate indexing is that the results of this study and of Courrier's study clearly demonstrate that we do not yet know which roles are the most useful or the most obvious for searchers and indexers and users, nor do we know how to define functional relations so that all users of the indexing system know which relations are the ones they should use. To expect all three participants - i.e., users, indexers and searchers - to agree on specific relations between keyterms seems to be more than a bit unrealistic.

Given the reservations discussed above, the most promising research in indexing systems at present, at least for bibliographic retrieval systems, is in string indexing systems which preserve meaning relations between keyterms through such traditional devices as word order, punctuation and judicious use of prepositions. It is also likely that any future use of relational indicators will be applied to the kinds of normalized expressions of topics these systems are attempting to construct.

Finally, the suggestions for the development of more ambitious systems for automatic analysis of query statements and abstracts is that it is highly unlikely that any system
can be expected to recognize the similarities between
the widely divergent expressions of relations that the
human subjects in this study were able to recognize and
distinguish between. Rarely did two abstracts in the same
set state the relations between keyterms in the same way
and world knowledge was often required to discern the
more oblique, but still readily recognizable, statements
of relations. However, the regularity of the language
behavior found in this study indicates that automatic
systems could be developed which could exploit these
regularities and aid human users in clarifying queries,
determining alternative statements of relations,
conducting searches, and even assigning key phrases and
keyterms in indexing. The first task in this endeavor
would seem to be the development of a clear set of relations
which are discrete, atomic, and 'intuitively' obvious.
7.2. Summary Statement.

In summary, the most useful result of this study is the suggestion that there is a more effective and, possibly, more objective means of determining relevance under test conditions and the indication that an operational definition of relevance using functional relations as a means of describing relations between keyterms can be demonstrated to be based on actual language behavior patterns of real users. There is no defense arising from this study for use of role indicators or some other method of fixing relations between keyterms in indexing systems in the bibliographic retrieval environment as it now exists. However, there are strong indications that functional relations could prove to be helpful in query negotiation and in the searching process because they would help the user or searcher to think of language expressions which might be used and should be taken into consideration when entering and combining search terms. The first direction for future research, then, appears to be exploration of the use of functional relations within the limits of the current state of bibliographic retrieval systems to determine how these relations can best be used in future development.

Suggestion of possible directions for research in the use of functional relations in manual indexing systems or in automatic indexing or abstracting systems would be
premature given the limited conclusions which can be drawn from this test. A more likely area for exploration would seem to be use of functional relations in some kind of search support or even indexing support mechanism. But the results of this study do indicate clearly that the development of some kind of functional relations model in the bibliographic retrieval environment is a useful pursuit as it is based on real human language behavior.
APPENDIX I
PERMISSION OF SUBJECT SPECIALISTS

Please sign below if you understand and agree to the conditions of your participation and the guarantees for your privacy.

The search performed for you will be provided free of charge under the following conditions:

1. The search strategy performed may be affected in part by the requirements of the researcher.
2. The evaluation of the search results as described in the accompanying instructions will be completed before you receive full records of your search results.
3. You agree to allow the query negotiation to be tape-recorded.
4. You will permit two other persons (who will be acting as secondary judges for purposes of this project) to hear the tape recording of your query negotiation or a transcript of the recording.
5. You will allow examples from your query negotiation and from your search results to be used as examples in the dissertation of which this search is a part and in any published articles which result.
Appendix I continued

In return for your cooperation in this project I will provide the following guarantees for your privacy:

1. At no time will your name be given to any other person, nor will it be connected to your query.

2. The list containing your name and the code used to protect your privacy will be destroyed upon completion of your part in the project.

3. The recording of your query negotiation will be destroyed upon completion of the study unless you specify otherwise below (Page 3).
Appendix I continued

Query negotiation is currently one of the most important areas of research in Library and Information Science. Your consent to any of the following could provide an important source of data to further our efforts. Please indicate (X) any further permission you will extend for further analysis or use of your taped interview by persons other than this researcher.

Thank you.

I will permit my tape to be used by faculty and/or Ph.D. students for individual research at the University of Western Ontario.

I will permit my tape to be used in Masters level classes for instructional purposes at the University of Western Ontario.

I will permit my tape to be used for research and/or instructional purposes outside the University of Western Ontario.

NAME ______________________ DATE __________________


APPENDIX II

SEARCH EVALUATION FORM

SEARCH EVALUATION FORM
DISSERTATION RESEARCH DATA COLLECTION
DEE LEWIS
SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF WESTERN ONTARIO

<table>
<thead>
<tr>
<th>Accession No.</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX III

INSTRUCTIONS TO SUBJECT SPECIALISTS

Instruction
Re: Evaluation of Search Results

Dear:

Please read each title and abstract carefully. Indicate beside the Accession Number on the accompanying form which corresponds to the Accession Number for the title/abstract whether or not the title/abstract is about the query you submitted. Check YES if there is any mention of the subject you described in your query in either the title or abstract. Check NO if there is no mention of your subject in either the title or abstract.

Please do not allow personal criteria such as usefulness, style, point of view, particular methodology, redundancy, etc. to influence your decisions as far as possible. We really want to know simply whether or not each title/abstract contains a reference to the topic you described in your query.

After you have completed your evaluations, please return the form indicating your decisions to Dee Lewis, care of the School of Library and Information Science, The University of Western Ontario. The full record of your search results will be sent to you immediately upon receipt of your evaluation.

Thank you for your cooperation.

Sincerely,

Dee Lewis
APPENDIX IV

SECONDARY JUDGES' SUBJECT KNOWLEDGE FORM

Please indicate on this form the level of subject knowledge you have in the areas listed below. If you have a B.A., B.Sc., M.A., etc. in the area please indicate the highest degree. Some knowledge means that you have read in the area or taken a couple of courses and are a little familiar with the terminology, but that you do not have a wide knowledge of the subject.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Have Studied</th>
<th>Some Knowledge</th>
<th>No Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socioeconomics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microbiology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Science</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Political Science</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audiology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women's Studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archives and Conservation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX V
SAMPLE QUERY

QUERY 13:

Subject Specialist (S): I'm interested in Type A behavior. Type A behavior is a kind of behavior related to coronary heart disease. And the components of Type A behavior are things like hostility, speed and impatience, stress related type things. I'm particularly interested in a test of Type A behavior. The name of that test is the Jenkins Activity Survey. And, indeed, the whole focus of my study is the construction and validation of the Jenkins Activity Survey.

There are a number of measures of Type A behavior. The Jenkins Activity Survey was the first touted as one of the best measures. It has not been turning out so well. The validation tests on the Jenkins Activity Survey have been suggesting that another measure, the structured interview, is much better than the Jenkins Activity Survey. This is a little surprising. So I think that if we go back and look at the construction of this measure, that is where the problem lies.

Interviewer (I): What is the survey like?

S: It's made up of about 65 objective questions. So it is a typical personality inventory where they ask things like "When I have to wait in line for something I feel very impatient — Yes, always; Yes, sometimes; Not usually; No, never." They have one for students, one for women, one for men. I would be interested in any of those.

I: What do you mean by construction?

S: Psychologists don't just sit down in a vacuum and write up a whole bunch of items. Normally what they do is they have a theory and they write items which are related to that theory. Then they have particular statistical procedures which they use to evaluate these items. The test construction refers to the method that you use to come up with your items and that you use to examine the quality of those items.
Appendix V continued

I: So you're interested in how the questions were created and worded, etc.

S: Yes. The second, a closely related, aspect of test construction is test validation. To validate a test you simply engage in a study to determine whether the test measures what it's supposed to measure. When you look at the validity studies on Type A behavior, in general, when they use the structured interview to predict coronary heart disease, they have a certain hit rate. When they use the Jenkins Activity Survey, which is supposed to measure Type A behavior, they have a much lower hit rate when they attempt to predict coronary heart disease. And I think the problem is that when they constructed the Jenkins Activity Survey, they didn't construct it to predict heart disease. They constructed it to predict the structured interview, and the structured interview was originally constructed to predict heart disease.

At the beginning there were a couple of cardiologists who said, there is this thing called Type A behavior. The way we measure Type A behavior is through the structured interview. That's how they did it originally. What they then did was classify people on the basis of this interview as Type A or Type B and they used this classification to predict coronary heart disease. And they got a certain hit rate which wasn't marvelous, but which suggested there was something there. These interviews take about an hour and you have to be trained to conduct the interview. Jenkins said, "We can do this better. We can make an objective measure using paper and pencil to obtain the interview content. You could get 200 people in a room, give everybody a paper and pencil and the test. However, Jenkins said the Jenkins Activity Survey would predict the structured interview, not coronary heart disease, but Type A or Type B behavior. But what they were really trying to predict was not Type A or Type B behavior, but coronary heart disease.

If you take the structured interview as the standard, the Jenkins Activity Survey was constructed to predict the structured interview which was not a direct prediction at all of coronary heart disease. Then what they do is they take the Jenkins Activity Survey and the structured interview and say which is better at
predicting coronary heart disease. It turns out that the Jenkins Activity Survey is much poorer. We're not sure exactly what is wrong with the Jenkins Activity Survey which is why I want to go back to its original construction and the early attempts to validate the Jenkins Activity Survey so I can see where they have gone wrong. Because generally what you find when you have a structured inventory like the Jenkins Activity Survey is that it is often a more reliable and, therefore, a more valid test than an interview.

I: So you want material about the construction and the validation of the Jenkins Activity Survey?

S: Yes.
APPENDIX VI

INSTRUCTIONS FOR SECOND LANGUAGE ANALYST

Please read the description of the case roles and verb types and their definitions. You may, of course refer to these descriptions at any time during your analysis of the abstracts. There are two tasks to perform in this analysis:

1. the assignment of case roles and verb types to the specified keyterms in the query statements and the abstracts;
2. the decisions as to whether or not each abstract contains a statement which matches at least one of the query statements.

The instructions for performing these tasks are included in the following pages together with a description of Fillmore's Case Grammar and the format you must use when indicating the functional relations between keyterms in the query statements and in the abstracts.
FILLMORE'S CASE GRAMMAR. (A Summary)

1. The Background of the Grammar.

Essentially, Fillmore proposed that the deepest level of categories underlying a grammatical structure—i.e., a sentence—are generalized cases. He suggests that "...the case notions comprise a set of universal, presumably innate, concepts which identify certain types of judgements human beings are capable of making about the events that are going on around them, judgements about such matters as who did it, who it happened to, and what got changed." These judgements are made at the linguistic competence level, or at the psychological level of the basic knowledge all people have about the descriptive patterns of their language.

For example, a competent speaker of English can understand that the following sentences are all about the same subject or contain essentially the same general information:

1. **Elephants** eat **peanuts**.
2. **Peanuts** are eaten by **elephants**.
3. That **elephant** ate a **peanut**.
4. Those **elephants** are eating **peanuts**.
5. **Peanuts** are among the foods fed to **elephants**.
6. **Pachyderms** sometimes eat **Goober's peas**.

In each of these sentences the relations between the words **elephant**, **eat**, and **peanuts** are the same as follows:
Elephants - eat - peanuts.
Agent  Action/Process  Object

By 'case', then, Fillmore means something other than grammatical subject or grammatical object. The case roles in this theory correspond more closely to the elements of a journalist's report: who, what, when, where, to whom, with what, and why.

According to Fillmore's model, a sentence is made up of two components:

1. a predicate, or the verbal element; and
2. its arguments, or noun phrases.

The meaning relations between the predicate and its arguments can be explicated unambiguously regardless of the surface syntactic relations used in the expression of the sentence. This is because the 'deep structure' relations between the noun phrase(s) and the verbal element hold constant despite variations in surface representation - e.g., active and passive versions of the 'same' sentence. In the above example, elephants eat and peanuts are eaten in each of the sentences. Thus, grammatical subject and grammatical object and the case notions of specific languages are not in any way implied by Fillmore's notion of 'case'.
THE VERB TYPES. There are four verb types, or types of predications, in Fillmore's theory: State, Process, Action, and Action-Process. They are defined as follows:

1. **State**: a fixed condition in time and/or space.
   a) The car is parked in the garage.
   b) The house overlooks the river.
   c) John is tall for his age.

2. **Process**: an ongoing event which involves some change in the state of the nominal element (Object) but not by direct action.
   a) The leaves are changing color.
   b) The children are learning to read.
   c) John is growing old.

3. **Action**: an activity which does not produce a change in a nominal element.
   a) Mary is running to catch a bus.
   b) The car sped down the highway.
   c) The child is reading.

4. **Action-Process**: an activity which produces a change in a nominal element.
   a) The children are being taught to read.
   b) Mary is building a house.
   c) The cat is washing its fur.

The verb provides clues, then, as to what case roles the nominal elements in a sentence may assume.
THE CASE ROLES.

AGENT. The case of the typically animate perceived instigator of the action identified by the verb as in:

(1) John opened the door with the key.

INSTRUMENT. The case of the inanimate force of object causally involved in the action or state identified by the verb as in:

(2) John opened the door with the key.

Note that the Instrument must be the immediate cause of the event.

In the case of a psychological event, the 'stimulus', the thing reacted to, is also an Instrument as in:

(3) The noise frightened Tom.

In addition, the Instrument role can be occupied by a sentence when that sentence identifies an event which is understood as having some other event or state as its consequence as in:

(4) The creaking of the floor made Tom jump.

EXPERIENCER. Typically, in a sentence describing a psychological event, we have the Experiencer and some other noun phrase which indicates the cause or the content of the mental state or psychological event as in:
(5) John was afraid of the dog.

This sentence is a description of a mental state. Note that the dog is the Agent. The dog cannot be the Instrument because it is an animate being.

(6) The noise reminded me of the accident.

This sentence is a description of a psychological event. Note that the use of Agent and Instrument is consistent with the rules described above for these two cases.

**OBJECT.** The Object is the semantically most neutral case, the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself. Probably the concept should be limited to things which are affected by the action or state identified by the verb. The term is not to be confused with the notion of direct object, nor with the name of the surface case synonymous with the accusative. To explain further, we have the Object case when:

1. a person or object is affected by the action or state identified by the verb as in:

(7) John broke the cup.

(8) Susan ate the apple

(9) Harry washed the dishes.
2. there is a nonpsychological verb which indicates a change of state as in:

(10) *Bobby* is growing.
Object

(11) *Mary* is dying.
Object

3. an entity which moves is the Object and the thing to which it moves is the Goal as in:

(12) *Mary* hit *the fence* with *the stick*.
Goal Object

(See the Goal case for explanation.)

GOAL and SOURCE. Depending on the type of predicator, the Source and Goal are interpreted as:

1. earlier and later locations as in:

(13) Mary walked from the house to the garage.
Source Goal

2. earlier and later states as in:

(14) The small boy became a large man.
Source Goal

3. earlier and later time points as in:

(15) The marathoner ran from morning until night.
Source Goal

Where there is a transfer or movement of something to a person, the receiver as destination is the Goal as in:

(16) Norma gave *me* the book.
The Goal case is also assigned to nouns which are the later state, or end result of some action or change. That is, the Goal specifies the end-result role of a thing which comes into existence as a result of the action identified by the verb as in:

(17) I - wrote - a poem.
    Agent          Goal

(18) I - constructed - a bridge.
    Agent          Goal

A sentence embedded as Goal is one which identifies the resulting state or event in a causative construction.

And, finally, an entity which moves is the Object and the thing to which it moves is the Goal as in:

(19) Mary - hit - the fence - with a stick.
    Agent          Goal      Object

PATH. In addition to Source and Goal there is another complement to verbs of motions as in:

(20) He - walked - from the cemetery gate -
    Agent          Source
    to the chapel - along the canal.
    Goal           Path

A sentence with the Path designated can contain an unlimited number of Path expressions, as long as these are understood as indicating successive stretches of the same Path.
LOCATION and TIME. These cases are optional complements of essentially any predicator, and indicate the place or time in which an event occurs. They are not to be confused with Source and Goal constructions which designate relations in time or place with verbs of motion. The difference between Location and Time and Source and Goal is further evidenced by the fact that all can occur in the same sentence as in:

(21) He – walked – from the museum, to the park
    Agent                      Source               Goal
    along Central Avenue – in Anytown –
    Path                        Path               Location

    on Tuesday afternoon.
    Time
RULES OF THE GRAMMAR.

1. Each case role describes only one relation to the verb.

2. All possible case roles a noun phrase can assume in relation to the verb are covered.

3. No role, except where noun phrases are conjoined, can appear more than once in a proposition as in:

   (22) John and Mary - walked - to class.

   Agent  Agent

   A good way of determining whether two nominal elements are conjoined and bear the same relation to the verb is to break the sentence into two equal sentences and see if the structure is the same for both.

4. Tense is not relevant to this grammar. The meaning relations here are not changed by use of the past tense, for example.
INSTRUCTIONS FOR THE ASSIGNMENT OF FUNCTIONAL RELATIONS.

1. QUERIES.
   1. Read each query statement and determine the functional relations between the nominal element(s) and the verbal element in each statement. Indicate these relations using the Semantic Network Diagram (See Attached).

2. ABSTRACTS.
   1. Read each abstract and underline the keyterms in the abstract which correspond to the keyterms in the set of query statements corresponding to the abstract. (KEYTERMS are words or phrases mentioned in the queries such as STOMATA, LEAF CHARACTERISTICS, DISEASE RESISTANCE, RUST, ETC. in Query 1.)
   2. Look carefully at the sentences which contain the keyterms. In some abstracts the keyterms may be several sentences apart. However, if you perceive that the keyterms are related to each other and the topic described is about the topic described in one or more of the query statements, bring the keyterms together into one sentence. If the abstract does not contain any statement which is about any of the query statements, try to determine the relations between the keyterms or the relations they have to
other words in the abstract. Note also that one abstract may have matching statements for several query statements. If so, pick one.

3. Now show the case roles assumed by each of the nominal elements (keyterms) and the verb type of the verbal element (may or may not be a specified keyterm) for the abstract statement you have chosen. If no specific verb from the text of the abstract provides the verb meaning you think is implied, supply one you feel expresses the meaning you perceive and the verb type you think is involved.

5. Map the statement you have selected using the Semantic Network Diagram.

4. Indicate whether or not you think the abstract contains a statement which matches a query statement in functional relations between the nouns and the verb.
APPENDIX VII

QUERY STATEMENTS:

The individual query statements are given first, followed by a list of synonyms for each key concept.

1. a) **Leaf characteristics** and **resistance to disease**.
   
b) **Leaf characteristics** influence **resistance to disease**.
   
c) The **leaf characteristics of varieties of plants which are resistant to disease** are compared to those of varieties which are not resistant to disease.
   
i) **leaf characteristics** [stomata, hairs]
   
ii) **resistance to disease** [resistance to rust, or any other particular disease of plants]

2. a) **Henry James use of impressionistic techniques in his writing**.
   
b) **Impressionism (art analogy) in Henry James works**.
   
c) **Henry James and his relationship to the impressionist movement**.

3. a) The **mechanism by which corona virus is associated with primary demyelination in mice**.
   
b) The **mechanism by which corona virus causes primary demyelination in mice**.
   
i) corona virus [mouse hepatitis virus, JHM, MHV]
   
ii) primary demyelination [persistent demyelination, demyelinating disease]
   
iii) mice [in vitro, in vivo, in cells]

4. a) Studies which look at the **factor structure of a Wechsler scale**.
   
b) **Factor analysis of the factor structure of a Wechsler scale**.
c) How does age affect the factor structure of the Wechsler scale(s).

d) Factor analysis of the Wechsler scale(s).
   i) factor analysis [cluster analysis]
   ii) Wechsler scale [WAIS, WISC, WISC-R, etc.]

5. a) Automation of social service information files by community information services.
   
b) Evaluation of software or hardware for automating social information files by a community information center.
   
c) Feasibility studies for automating social information files.
   
d) Automation of social service information files.

6. a) Mechanism by which hormones regulate gene expression in mammals.
   i) hormones [estrogen, progesterone, steroids, testosterone, insulin, thyroid, etc.]
   ii) mammals [in vivo, in vitro, mice, chicks, viruses, MCF-7 cells, etc.]

7. a) Programs for teaching computer literacy.
   i) programs [curriculum guides, computer programs software, etc.]
   ii) computer literacy [basic computer skills, computer related skills]

8. a) Online generation of index displays.
   b) Online display of relations between concepts.
   i) online display [computer generation, computer display, etc.]
   ii) relations between concepts [links and nodes, relations in a thesaurus, etc.]
9. a) Automation of records management in educational institutions.
   b) Management of educational records.

10. a) Curriculum guides for teaching physical skills to K-3 children.

11. a) Development of a classification scheme for normal and abnormal personality traits.
   b) Comparison of normal and abnormal personality characteristics.
   c) Describes dimensions of personality traits.

12. a) Methods of deacidification of paper.
   b) The effect of acid on paper.
   c) Methods of preserving library materials.

   b) Validation of the Jenkins Activity Survey.
   c) The Jenkins Activity Survey predicts coronary heart disease.

   b) Retrieval of the text of journals online.

15. a) Fundamental processes of human thinking as they have investigated in artificial intelligence.
   b) How humans approach problem solving.
   c) Simulating human thinking or memory on a computer.
   d) Associative relations between concepts in the human mind.

   i) human thinking [memory, inference, informal logical processes, etc.]

16. a) Information gathering habits of researchers.
   b) Information needs of researchers.
c) Evaluation of information services to researchers.

d) The nature of the relationship between information producers and information users.

17. a) Higher education of women in the 19th century.

b) How higher education affected women's lives in the 19th century.

c) When were women's colleges opened and where in the 19th century.

d) The role of women in education in the 19th century.

e) When did men's colleges start admitting women.

f) In what disciplines were women admitted in the 19th century.

18. a) The modernization and development process in Tanzania.

b) How is modernization affecting society in Tanzania.

c) Evaluation of economic policies in Tanzania.

d) How do social factors affect economic development in Tanzania.

   i) Modernization [economic development, political development, social development; etc.]

   ii) Social factors [history, culture, tradition, religion, social structure, political doctrine, international relations, etc.]

19. a) How does social class affect liberal/conservative attitudes and behavior.

b) Is social class related to political party support (liberal/conservative parties).

c) Use of Liberal/Conservative Scales to measure political or social attitudes.

20. a) How is the informal sector related to the mode of production in Africa.

b) Articulation of mode of production.
c) What affects the informal sector.

21. a) The mechanism by which superovulation affects fertility in mammals.

b) The mechanism by which [ ] affects superovulation in mammals.

c) Methods for transferring embryos.

i) fertility [ova, ovulation, embryo development]

22. a) Women writers are associated with what characteristics.

b) How does the literary establishment exclude women writers.

23. a) How does trypanosoma infection affect the immune response of the host.

b) Descriptions of the ultrastructure of trypanosomes.

24. a) The use of parasites as biological tags - e.g. to indicate feeding habits, migration patterns, population densities, ecology, etc.)

b) Measurement of the acoustic qualities of vowels.

25. a) Changes over time in the acoustics of speech of normal and hearing impaired persons.

b) Spectrographic displays of deaf speech.

c) Acoustic qualities of the vowels in deaf speech.

26. a) Populist movements and economic depressions.

i) populist movements [right and left wing movements, political attitudes and actions, people's responses to movements, etc.]

27. a) A comparison of men's and women's management styles.

b) What successful leadership styles are used by women.
c) Cooperative management styles compared with competitive management styles.
   i) management styles [leadership styles]
   ii) competitive styles [authoritarian, hierarchical, formal, etc.]

   b) Female archetypal patterns in literature or the arts.
   c) Psychological or archetypal criticism of children's or women's literature.
   d) Feminist criticism of literature.
      i) search [quest]
      ii) identity [self-knowledge]

29. a) Migration patterns of nematodes in the tissues of hosts.

30. a) Use of multivariate (or bivariate) statistical distributions for simulations based on real data.
    b) Use of multivariate (or bivariate) distributions for prediction in simulations using real data.
       i) simulation [model]
REFERENCES


END

270385