INTERDISCIPLINARY PERSPECTIVES IN SUBJECT ANALYSIS

Madalena Martins Lopes Naves (1), Célia da Consolação Dias (2)

School of Information, University of Minas Gerais, Brazil, madalena.naves@gmail.com (1) celiadias@gmail.com (2)

Abstract

In Information Science context, Subject Analysis is considered one of the most important stages in Indexing and the indexer is the reponsible for this activity. An interdisciplinarity with Linguistics, Cognitive Science and Logics is observed during the development of this activity by the indexer. The study discusses factors of these areas that interfere in Subject Analysis. Linguistic factors act in concert with logical and cognitive factors in the intellectual activities that comprise Subject Analysis. So close is their interaction that it is not possible to determine where the influence of one begins or finishes. This reference base is required in order to understand what the text in question is about and in order to assimilate new knowledge. When analyzing a text, the indexer searches the reference base for stored information about the theme of the text and if found comprehension occurs; however, if not found, comprehension is thwarted.

Keywords: subject analysis; Linguistic; Cognitive psychology; Logic

1 Introduction

Subject Analysis is performed by an indexer. It is the process whereby an indexer reads a text, extracts concepts from it and determines its aboutness. Considered the intellectual aspect of indexing, Subject Analysis is affected by many factors: the text being indexed, in particular, the author's ability to express his subject; the indexer's background, including his knowledge specialties, indexing expertise, judgment and experience; institutional factors, such as the objectives of the institution, its indexing policy, the types of vocabulary used (natural or artificial language); and user profiles and information needs. Of these the chief factor, the factor most directly affecting the process is the human factor, the indexer. The indexer when thinking, making abstractions, interrelating and defining the subject of a document is himself influenced by factors of a linguistic, cognitive and logical nature. The purpose of this paper is to examine these interdisciplinary factors.

2 Literature Review

Two specialists in the area of Information Science who have given extensive consideration to the interdisciplinary factors in indexing are Clare Beghtol, a Canadian researcher, and Pinto Molina, a Spanish educator. Beghtol in *Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality, and the cognitive act of classifying documents* (1986) looks at linguistic and cognitive factors and draws out some of the implications of Van Dijk's work for the theory of bibliographic classification. This is relevant to a discussion of Subject Analysis in that it treats of indexers' intellectual activity in interpreting the thematic content of documents.

Also interdisciplinary in its approach is the article by Pinto Molina entitled *Interdisciplinary approaches to the concept and practice of written text documentary content analysis (WTDCA)* (1994). The purpose of Molina's WTDCA model is to deconstruct the object of a written text in order to determine its content. The deconstructive process, which at times is uncertain and ambiguous, begins with the text, which is considered the necessary point of departure for analytic and documentary operations. Pinto Molina believes that any operation executed with respect to a text must be studied from three mutually interactive points of view: the cognitive, the linguistic and the logical. She therefore develops her model with supporting contributions from the three disciplines: Cognitive psychology, because textual analysis is a cognitive process; Logic, because all operations involved in the process must be rationalized; and Linguistics for obvious reasons. In a pedagogic fashion, Pinto Molina (1995) distinguishes three basic stages in textual deconstruction:

(1) reading and comprehension. At this stage of the process the contributions of Cognitive Psychology are fundamentally important.

(2) inference and interpretation. All three disciplines contribute to this, the most creative stage of the process.

(3) synthesis. At this stage the analyst adapts content and structure to meet certain documentary demands.

Other researchers corroborate with Begthol and Pinto Molina relating to the linguistic aspects. In this way, it is emphasized the studies of Moreiro-Gonzáles (1994); Kobashi (1996), and Fujita (2003). Moreiro-Gonzáles (1994) who punctuates the linguistic factors when doing a documentary, points out the phases of reading-analysis, synthesis and representation. Kobashi (1996) also mentions the interface between linguistics and a documentary analysis, as well as Fujita (2003), who in

her study of documentary reading proves the importance of knowledge of the textual structure to facilitate the indexer task of locating relevant information in a specific document (Silva; Fujita, 2004) and more recently, the study of Dias & Naves (2013).

It can be verified in the specialized literature about this subject that there is a lack of current studies, reason why the bibliography does not present more recent items. However, it is believed that this paper may be useful to the theoretical fundamental of new research considering the relevance of the interdisciplinarity that serves as a support for the study of subject analysis.

In the study of indexing languages, as in several other topics in the area, there is a certain conceptual confusion, perhaps arising from the use of the terms "documentation" and "indexing". There is a perception that in the studies developed by French experts or by those who rely on them, there is a strong influence of the term "documentation" because they come from the theory of Otlet and La Fontaine. About the studies of English language, especially those of North American origin, a strong influence in the adoption of the term indexing is already noticeable. Therefore, since there is this duplicity of the terms representing the same concept, it was decided to maintain in this article the terms used by the cited authors, who use either documentary languages or indexing languages.

Following Beghtol and Pinto Molina, the thesis of the present paper is that the theoretical base of Subject Analysis, an operation performed upon texts, is interdisciplinary, made up of contributions from Linguistics, Cognitive psychology and Logic. This thesis is developed in what follows.

2.1 Linguistic factors

In textual analysis, Linguistics and Documentation meet at the point of content extraction. The similarity between the two disciplines, which is evident at a certain level of abstraction, is less studied than it might be. Smit (1987) observes that with few exceptions there has been no interdisciplinary synthesis and she offers some reasons for this:

- Until recently Linguistics research has been focused in the area of syntax, which, though important, is not immediately relevant to documentation, and
- the practice of documentary analysis for long has been based on operations implicit in a manual indexing culture.

In this context, Kobashi (1996) highlights that the interface between linguistics and documentation in the late 1960s arose through the implementation of computers in the documentary work, "by automatic indexing experiments, abstraction elaboration, and information retrieval". (p.6)

With the explosion of scientific information and automated documentation procedures, it became imperative to make implicit indexing operations explicit. Linguistics is fundamental to automated documentation. The same author in another paper (1978) also observes that interest in explicating the process of documentary analysis (and also the linguistic questions related to it) dates from the 1970's. The discussions that emerged at that time were not limited to questions of syntax and semantics because when the text as a whole, rather than the individual sentence, becomes important logical considerations come to the fore.

Another instance of relating Linguistics to indexing is found in the work of Cintra (1983), which identifies many linguistic phenomena occurring in the conversion of natural language into documentary language. Different linguistic subdisciplines are embraced in the study of a language, and of these that which is most relevant to documentary language is semantics.

Some difficulties arise regarding the interrelationship between Information Science and Linguistics, in part resulting from the fact that scholars in the one discipline are not sufficiently knowledgeable in the other; for instance, information scientists may have expectations that Linguistics is not able and probably will never be able to fulfill. This is the opinion of Baranow (1983), who cites an FID (International Federation of Documentation) publication written in the 1970, when the relationship between the two disciplines attracted particular interest. One of Baranow's conclusions is that it is difficult to link linguistic techniques with techniques of information retrieval. The most apt application is in the area of automatic indexing, where research using linguistic models has shown some promise.

Focusing didactically on the interface between Linguistics and Information Science, Baranow (1983), in addition to introducing general linguistic concepts, proposes a Linguistics course for information scientists which would include the study of syntax and semantics as well as Psycholinguistics and Sociolinguistics. He cites a work by Basilio (1979) that deals with morphology, in particular with the analysis of lexical structures within the context of Information Science. He shows, for instance how morphological rules governing word formation have applications in both automatic and manual indexing.

Linguistics is defined as the science that studies natural languages. Its importance to documentary analysis is not restricted to its methods of analysis. In that documentary analysis has language as its object, in particular documentary language, a theoretical excursion into Linguistics would be of use to the information scientist, not only to learn about the particulars of sentence pro-

cessing but also to understand the elaboration of methodology and the explication of results (Cunha, 1987).

At the same time Izquierdo Arroyo (1990) argues that the term Linguistics is that composes it must be taken in a much broader sense than the usual. Linguistics comes to be understood as the science of human natural language, and thus, conceived would be impertinent to the purpose of the documentary language. According to the author the aim of the Documentary Language is designed artificially to a communication purpose.

Cunha (1987) argues that the process of documentary analysis utilizes linguistic concepts in text recognition and in the structuring of documentary languages. While interdisciplinary studies crossing the borders of Linguistics and Documentation did not exist in 1987 when she wrote, she foresaw that the latter could borrow pragmatic approaches from the former, particularly in the area of indexing. Today, the Documentation literature attests to an increasing preoccupation with the intersection of the two disciplines.

That Linguistics must have some relevance to indexing should be apparent to anyone attempting to index a document, that is, to translate its content into words that can be used to retrieve it. But it was not until the 1980's that researchers broaching the two areas began to adopt scientific procedures in indexing methodology (Navarro, 1988). This advance was occasioned in part by the necessity of systematizing the relationship between natural languages and documentary languages for the purpose of developing automatic indexing procedures. This in turn led to theoretical considerations of the relationships between natural and documentary languages.

One of the most serious linguistic problems encountered in indexing is terminological in character. This problem is encountered especially when indexing documents in the humanities area, more than in scientific and technical areas. Bell (1991) for instance attributes terminological limitations in describing human life and its relationships in part to the fact that language evolves more slowly than changes in society.

Dahlberg in her writings addresses many of the linguistic and philosophical problems encountered in the organizing of knowledge (see Dahlberg, 1992). She observes that the content of disciplines like Logic, Epistemology, Ontology, Phenomenology, Etiology and Metaphysics cannot be communicated except through the medium of language. When such disciplines generate new concepts, these need to be named. It is important that the concepts are well contextualized and explicit, since the same concept can receive different terms depending on the area in which it is inserted. In creating verbal forms for concepts, the terms that establish the discipline's terminology emerge.

.

2.2 Logic factors

Linguistic factors act in concert with logical and cognitive factors in the intellectual activities that comprise Subject Analysis. So close is their interaction that it is not possible to determine where the influence of one begins or finishes. Cunha (1987) observes that Logic, as a Science, is probably less relevant than Linguistics to the concerns of librarians, even in matters relating to automation. However, in the same way document analysis makes use of the theoretical constructs of Linguistics in its pragmatic activity, those of Logic are relevant to the parameters of the activity as a whole. General Logic is relevant in that it deals with how scientific knowledge is acquired and how principles, hypotheses, general laws and theories are constructed. Formal logic has application in that it focuses on how reasoning is validated.

Within Philosophy, Logic is considered a normative Science Santaella (1992) elaborates upon Logic's normative character by referencing the distinction made by Peirce between utens logic and docens logic. People, in the normal course of life, reason by instinct and by habit; and this leads them to have opinions about many important subjects. Implicit in this reasoning is utens logic. On the other hand, when one encounters extraordinary facts that require understanding within the context of theory, utens logic is not sufficient. What is needed is reasoning based on a valid foundation, reasoning that is trustworthy in advancing knowledge. Governing this reason is docens logic.

Peirce holds that our conduct may be supported by utens logic only up to a point; when facing a complicated situation we must reason differently, we must employ a more rigorous scientific logic (Santaella, 1992, p. 124). Subject Analysis is a complicated process. The effort involved in the process requires inventive capacity, generalization and theoretical construction on the part of the indexer, activities that presuppose not just utens logic, but docens logic as well.

Another differentiation made by Peirce is between (a) logic in the strict sense, being the science of necessary conditions to reach truth and (b) logic in the wide sense, being the science that deals with the laws of thinking. The latter he called Semiotics. Semiotics treats of the general theory of signs (because thinking always employs signs) and the evolution of the laws of thinking and the conditions of meaning required for communication to take place between one mind and another.

Semiotics is divided into three parts: (1) speculative grammar, which studies how signs function; (2) critical logic (or logic proper), which was first developed as a unified theory of abstraction, deduction and inference; (3) methodological or speculative rhetoric, which stud-

ies the general conditions of a symbol's relationship to other symbols and its interpreters.

In the process of analyzing a text, two types of reasoning or logical argumentation are used: deductive logic (as represented by the Aristotelian syllogism) and inductive logic (of the kind used in probabilistic inference). In deductive reasoning, the factual content of the starting propositions (premises) is not expanded. That is, deduction does not create new knowledge; its conclusions are inherent in, and, thus, determined by its premises. If the premises are true, the conclusion will be valid. (However, a false premise can imply a true conclusion.). In inductive reasoning the factual content of the premises is increased, new knowledge is created.

Formal logic, which treats exclusively the form of propositions, is a subset of General Logic, and as its name indicates, is restricted to the formal manipulation of symbols, as in an algebraic calculus. The relevance of formal logic to Subject Analysis is linked to its use in relating concepts whose meanings are relatively fixed, for instance in the broadening and narrowing of concepts.

One of a scientist's activities is discursive reasoning and this is reflected in the use of language where expressions like *therefore* and *then* precede conclusions derived through argumentation from facts. Two of the principles governing logical argumentation are the identity principle (if $a = c$ and $b = c$, then $a = b$ and the *dici de omni, dici de nullo principle* (if all men are mortal, then no man is immortal). Logical and psychological forces are separate but related determinants of human thinking as it is embodied in scientific texts and in structures of universal concepts. Thus, a third set of factors impacting the process of Subject Analysis is cognitive in nature.

2.3 Cognitive factors

To understand the impact of cognitive factors on Subject Analysis it is important to understand what is meant by *Cognitive Science* and *Cognitive Psychology.* According to Daniels (1986), *Cognitive Science* is multidisciplinary, comprising aspects from the disciplines of *Cognitive Psychology, Linguistics, Artificial Intelligence, Philosophy, Education* and, ideally, *Information Science. Cognitive psychology* is an empirical science based on the perception of humans as communicative beings. A relatively young science, it attempts to identify the mental processes and structures involved in the acquisition, processing and use of knowledge or information. It attempts to distinguish mental processes, such as memory and attention, and mental representations, such as the objects of imagination, the formulation of propositions and the establishment of categories, and complex mental processes, such as comprehension reasoning and problem solving (Pinto Molina, 1994).

According to Belkin (1990) the essence of a cognitive approach to information processing can be expressed in terms of human perception, cognition and knowledge structures. He cites De Mey (1977), who argues that each instance of information processing, perceptive or symbolic, is mediated by a system of categories or concepts, which constitutes a mental model of the mechanics of information processing. The mental model is determined by different types of experiences: individual and collective. De Mey views the information retrieval task as the explication by system analysts and indexers of an author's cognitive structures in a way that relates to user needs. Collective cognition, of the kind frequently reflected in paradigmatic theories, influences the design of classification and indexing systems and, consequently, the relationships among concepts established in the body of analyzed literature.

Language gives rise to thought and language is used to express thinking. Farradane (1976) holds that thinking in its initial form is nonverbal and that it is transformed to a communicative medium through spoken or written language. Spoken language can be complemented by gestures, intonation, articulation or emphasis. In written language emphases can be introduce by the use of italics, unconventional spacing, underlining and other forms of punctuation.

In a philosophic approach, Hannabus (1988) draws a parallel between knowledge and thinking. *We think, therefore we know. Since we know and know we know, we know we think.* Thinking is a form of extended knowing. Knowing is seen in the context of a subjective-objective continuum, somewhere between a total idiosyncratic subjective interpretation of events and a rigid vision of them. Hannabus (1988) believes that insofar as indexers engage in interpretation of texts, the discipline of Information Science must include studies of human behavior, in particular those aspects of behavior related to communication. These studies must begin with the study of human thinking, defining it as a mental process or as a physiological process that takes place in the brain. It is necessary to include considerations of how thinking is produced, that is, how internal or external stimuli trigger mental activity and how thinking is retained in (or lost from) memory, and how it can be differentiated from other activities such as perception. The study of Mind, defined as a complex of thinking and other processes of the brain, and knowledge, defined as structured thinking stored in the Mind, has a place in Experimental Psychology. Results of research in this discipline are relevant to the constant search for solutions to the problems of indexing.

A type of thinking of interest to the present study is syllogistic thinking, which is explicated in the discipline of Philosophy. Syllogistic thinking involves a set

of individual propositions expressed at different levels of generality and between which exist relations of necessity. In the classic Aristotelian syllogism, consisting of two premises and a conclusion, the first premise expresses a general proposition and the second a particular proposition. As a result of the two premises taken in combination, a conclusion is derived based on the general rule of implication. For example, the two propositions: Precious metals do not rust and Gold is a precious metal; logically imply the conclusion: Gold does not rust. The conclusion of a syllogism is a tautological inference in that it is not based on experience but on the relations between propositions.

Some part of human cognitive processing involves deductive or logical inference. This is true in particular of the intellectual operation of indexing, wherein concepts are extracted from a text in order to determine its aboutness. The concept of aboutness is seen as a property of some types of discourse and has been studied for a long time by logicians and linguists, because it represents an aspect of meaning. It can be considered the moment when the indexer can define "what the text is about".

In this context emerges the study of Memory, which plays a significant rule in thinking. *Memory* can be defined as the facility to retain ideas, impressions, and knowledge acquired formerly. The mechanisms by which memory works are described by Farrow (1995) in terms of the acquisition, storing and transference of information in the human mind. Information from the outside world enters the mind through the senses and is stored in a sensory register. This occurs in a very small period of time, maybe a quarter of a second. During this time some information is selected for further processing. Information that is not selected at this stage is lost.

The information that is selected resides in short-term memory. Short- term memory (STM) is a work space managed by control strategies to retain information units or transfer them to Long-term memory (LTM). There is a constant transference of information between STM and LTM. In contrast to STM, LTM stores information permanently in what can be envisaged as semantic nets. Control strategies govern the acquisition and retrieval of information from LTM. When a unit of information in the form of a concept is activated, related concepts in the same semantic net are also activated in proportion to the semantic distance between them.

Concerning this point Leiva (2012) argues that Short-term memory (STM) and Long-term memory (LTM) are closely related devices, despite their functional differences. There is a double interaction between the two structures. On one hand, the information retained in the STM from the sensory input is sometimes transferred to LTM (for example, if the telephone number we first hear interest us, we can repeat it several times and, consequently, transfer it to the LTM to remind it forever). On the other hand, when we want to retrieve some information from the LTM for an immediate use, such information is activated in the LTM. The STM is therefore a device that operates from inputs from both the outside and from the cognitive system itself.

Computer representations of stored knowledge are sometimes represented by a series of associated structures called schemata, frames and scripts (events). When new information is processed a schemata is activated; a frame is a representation of general knowledge; a script is a structure formed by a series of event, e.g. going to a restaurant involves events like being hungry, getting a table, eating dinner and paying for food.

Ingwersen (1982) distinguishes two different kinds of memory: episodical memory and semantic memory. The former consists of individual knowledge, the latter of knowledge that is common to many persons by virtue of similar education and skills. Both types of memory influence how librarians search for information.

Alkinson and Shiffrin (cited by Najarian 1981) look at memory as having three components: (a) a sensory register, which receives sense information; (b) a staging area where information is discarded or transferred to LTM; and (c) LTM or permanent storage. There is considerable controversy over what memory is and how it functions, for instance about the evidence that exists about storage durations in STM and LTM (Garcia Marco and Esteban Navarro, 1993). Different models of memory have been formalized, eg.: a structural model, that distinguishes sensorial, long-term and short-term memory); a functional or operational model that distinguishes superficial and deep processing; and the model introduced by Ingwersen (1982) that distinguishes episodical and semantic memory.

It is hypothesized (Garcia Marco and Esteban Navarro, 1993) that there are two ways in which semantic information is integrated into LTM. The first is by sequential organization whereby things are remembered in the same order in which they were learned. The second is analytic-synthetic in nature whereby what is learned is integrated into a person's idiosyncratic knowledge structures stored in his memory as semantic nets.

Van Dijk and Kintsch (1978) offer a theoretical explanation of indexing as a process of text reduction. In explaining this process it is important to distinguish two aspects of cognition: thinking and memory. T*hinking* is the act or effect of thought, reflection, or meditation. It is a mental process that concentrates the formulating of ideas; a psychic activity; a cognitive phenomenon that is separate from sensory perception.

In Information Science, as in other disciplines, the cognitive point of view considers a phenomenon and the situations that surround it in representational terms. For the most part, these are mental representations of knowledge, intentions and beliefs and of the interactions among them. Cognitive based Information Science typically considers its object of study to be human communication systems, wherein there is an interaction between texts that have a function to perform and individuals who have information needs.

According to Allen (1991), research projects in Information Science focusing on the users' conceptual models began in 1977 and have steadily proliferated. For the most part these projects have applied methodologies and explanatory structures developed within Cognitive Science. Todd (cited by Farrow, 1995) sees it to be a matter of urgent necessity for the indexing professions to develop a basic theoretical foundation for their work. The emergence in the mid 1970's of Cognitive Psychology as a distinct discipline has impelled progress toward establishing such a foundation and contributing to understanding the indexing process.The aspects of Psychology that are of particular interest to Documentation are limited and there are lacunae between Cognitive Science and Information Science. This is the opinion of Garcia Marco and Esteban Navarro (1993), who hold that human cognition is a social process that requires specialized functions and communication. Only as a part of a social process can documentary information as indexing terms extracted from an indexing language, be recorded as knowledge, as a part of humanity's cultural inheritance. However, it seems clear that insofar as social information processes are mediated by psychological processes, Psychology does have something to offer. Garcia Marco and Esteban Navarro (1993) believe that Cognitive Psychology is relevant to information scientists for two principal reasons: (1) Broadly viewed, Information Science and Cognitive Psychology are both cognitive sciences; both look at knowledge representations and both are concerned with how information is processed; and (2) Psychological processes permeate activities involving the processing of information, particularly interface activities between humans and between humans and machines. At the end of the 1970's, the rapprochement between Cognitive Science and Information Science was becoming evident. Classification and cognition was the theme of the 3rd Annual Conference of the German Society of Classification (1979). Dahlberg (1987) interpreted this Conference's results as suggesting that an understanding of basic classificatory concepts was key to general knowledge comprehension and necessary to the performance of documentary activity (p. 79). Documentary activity are the activities related to documentation, such as reading, extraction of concepts and translation into an indexing language. The 2nd ISKO (International Society for Knowledge Organization) that took place in Madras, India, in 1992, had a similar theme: Cognitive paradigms in Knowledge Organization. Beginning with the premise that knowledge is a result of human beings' cognitive activity, Dahlberg sees its organization as requiring the abilities of thought recognition, problem solving and decision making. Among the recommendations proposed by the Conference participants were:

- The intellectual work involved in creating and organizing knowledge could be carried out more efficiently if knowledge representation were reflective of cognitive processes and sensory perception;
- The cognitive processes of learning, comprehension, problem solving and thought expression could profit from using an analytic-synthetic approach. Such an approach could make it possible for users to model interactively the organization of knowledge in computer memory in the form of cognitive paradigms;
- Ranganathan's theory of classification and knowledge organization could provide such paradigms and developments in computer science and artificial intelligence could apply them in information retrieval (Gopinath, 1992).

The central task in comprehending a text is to assimilate not only its surface meanings and grammar, but also its deep structure. According to Luria (1994), this task is as complex as it is important and constitutes the fundamental content of a new science, which taking elements from both Psychology and Linguistics is called Psycholinguistics.

Leiva (2012, p.39) also argues that a structure and some specific content configurate textual typologies, that is why these textual typologies itself open the way to understanding. And an understanding of a text opens the way to indexing, some researches confirm that the more structured a summary is, the more it contributes to indexing.

Cognitive processes are mental activities, such as thinking, imagining, and problem solving (Allen, 1991). Like other human activities they are performed differently by persons of different levels of ability in logical reasoning, the recall of visual images and vocabulary articulation, all of which can affect the effectiveness of information retrieval.

Those seeking to understand cognitive processes pursue research in one of two directions. The first, which is based on textual analysis, is pursued within the domains of Linguistics and Cognitive Psychology. Illustrative of this type of research is the frequently cited text comprehension model developed by Kintsch and Van Dijk (1978). The partnership between these two researchers, one a linguist and the other a psychologist, is significant as an example of interdisciplinary coop-

eration. The authors' model looks at text processing from the point of view of propositional reduction; that is, as reducing a text to elementary propositions, arriving thus at its microstructure (global semantic structure). This model explains comprehension in terms of previous knowledge and logical inference.

The second type of research pursued takes various forms, the principal one being the development of a theory of how knowledge is represented in memory. An example is Beghtol's work, in her analysis of the cognitive processes involved in the generation and comprehension of texts, which borrows from Cognitive Psychology the concept of two types of mental information processing: top-down and bottom-up. Empirical research shows that these two types of mental processing seem to occur continuously and simultaneously as a text is comprehended by a reader.

The two types of mental processing, regarded by some as models of the reading process, are contrary and complementary: the bottom-up type is inductive, guided by data, and models linear reading, which goes from parts of a text to the whole; the top-down type is deductive, conceptually oriented and moves in the opposite direction, employing the vantage of the reader's knowledge base. The double action integrates perceptions and comprehension (Pinto Molina, 1995; Cole, 1994).

Bottom-up processing constructs the meaning of a text by decodifying base units; that is, by recognizing letters, syllables, words and sentences. According to Dell'Isola (1999), the decodification is text based and is data driven from outside-to-inside. It is assumed that reading takes as its object a text and that this text is the point of departure in the search for its author's intention or message.

The top-down model assumes that knowledge structures contained in a reader's mind attribute meaning to texts. Reading is seen as a game of riddles and comprehension in a continuous process of hypothesis elaboration and verification. This model, known also as the psycholinguistic model is reader-based, schema-driven and employs hypothesis testing. It assumes that the reader, rather than the text, is the point of departure in the search for a text's meaning. Top-down processing finds its place in Cognitive Psychology theory that posits reader's mental models and systems that organize knowledge into concept categories (Dell'Isola, 1999).

Smith (cited by Cole, 1994) is the author of one of the most well known top-down models of reading. This model divides the mechanism of perception and reading into two levels: global and focal. In explaining his model Smith makes an interesting analogy with car travel. Global signs are those that are foreseen, like the traffic signs that occur on a motor trip. These function as predications, influencing decisions and events that can occur. Focal signs, on the other hand, are unexpected, ones that are constructed temporarily or for a particular purpose, for example a red signal that indicates that there is road work ahead. As with a car's travel, text reading begins with global predictions about content, theme and treatment, and these usually persist throughout the reading. Focal predications occur at specified places in the reading, in proximity to special words, paragraphs and chapters.

Another type of model used to explain the process of reading is the interactive model. It is predicated on the assumption that reading requires a continuous interaction between reader and text. Widdowson, cited by Dell'Isola (1999), sees reading as a process that combines text information and the information a reader brings to the text. As such it is considered "communication", an interactive process, consisting of an exchange between readers and authors that results in a convergence of meanings.

Models of the reading process are relevant to Subject Analysis insofar as they explain comprehension. For instance, in indexing, the interactive model conceives of an indexer interacting with a written text for the purpose of reconstructing the meaning of the text, based on his previous knowledge. In creating condensed representations of a text, the abstractor is involved in an intellectual activity that consists of identifying words and comprehending the vocabulary utilized. This activity requires as well taking note of cognitive representations, their placement in memory and their integration with previous knowledge.

Monday (1996) looks at cognitive structures and processes activated in the comprehension of textual sentences. He describes those that are involved in creating an abstract. A similar set of cognitive structures and processes is utilized when indexing; what varies is the output, e.g., an index string rather than compressed text. Among these structures and processes are the following:

- Intellectual schemes, which are representations of general concepts designating objects, events or situations. These representations may be likened to Plato's ideas.
- Cognitive units, which consist of concepts, propositions, episodes, rules of production and heuristic procedures. The individual continuously receives stimulation from his environment, incorporating cognitive units into his personal knowledge structures.
- Knowledge organization, which is the structure of intellectual schemes in the brain and its vehicles, e.g., the cognitive units and memory, that constitute the physical structure in which knowledge is stored.

- Mental representations of encyclopedic knowledge. These include short duration representations in memory as well representations that are permanent.

Semantic and Syntactic structures of texts, which are characterized on two levels, a primary microstructure level and a secondary macrostructure level. The former consists of individual propositions and the relations among them; the latter consists of relations between groups of propositions in the general organization of texts. To comprehend a text requires identifying these structures within the text. Full comprehension requires integration of the text's message with the reader's general knowledge.

The act of comprehending a text involves extracting from it meanings and relating these to other experiences or ideas. Comprehension is prerequisite to activities such as paraphrasing the text, abstracting it, responding to questions about it or criticizing it (Weiner and Cromer, cited by Dell'Isola, 1999). Comprehension is influenced by what a reader already knows; the level of comprehension depends on the degree to which knowledge is shared between the reader and the author of the text, the degree to which their mental models overlap.

In any case, the comprehension of a text is influenced by cognitive factors, such as the reader's interest in and attitude toward the text, his world knowledge and his expectations of what the text is about. Comprehension is determined by interpretation, which encompasses reading the text for the purpose of understanding it and developing attitudes toward it such as whether one agrees with what is said and whether what is said is interesting. Beghtol (1986) believes that the interpretation of a text is to be found in the intersection of the many different knowledge structures that exist in the totality of readers' minds.

A given document admits of multiple interpretations, which has the potentiality to introduce a polemical dimension in the form of arguments over which interpretation is to be preferred (Birman, 1994). The conflict or discord created by multiple interpretations is a reality encountered daily by indexers when attempting to determine what a text is about.

Out of comprehension and interpretation arise inferences, which as the nuclei of these processes form the nexus of human communication. According to Giasson (1993), in the making of inferences, the reader necessarily goes beyond literal comprehension. He cites Cunningham, who distinguishes two types of inference: inferences based on a given text, which are logical inferences; and inferences based on a reader's knowledge or conceptual scheme, which are pragmatic inferences. Pinto Molina (1995), in the context of discussing inferences made for the purpose of documentation, categorizes four types of inference;

- Logical inference, which is used by the indexer to establish causes, motivation and conditions defining specific situations.
- Evaluative inference, where inference is colored by the indexer's beliefs about the described situations.
- Integrative inference, which is based on hierarchal concepts and properties and is executed by the indexer at the moment of comprehension.
- Constructive inference, which is based on the indexer's knowledge.

The comprehension of texts is integral to the activities of indexing, classification, abstracting and in the making of text condensations and synopses. What these activities have in common is the analysis of subjects for the purpose of document representation. However, there are differences in the objectives of these activities (Farrow, 1995). Classification has as its objective the representation of a document's meaning translated to a particular classification scheme. The objective of abstracting is the representation in continuous prose of the main arguments of a document. Indexing has as its objective determining the aboutness of a document for the purpose of assigning descriptors to it. Two types of indexing can be distinguished: book indexing and periodical or database indexing. In the former the task is to read the text, distinguishing between relevant and peripheral information and then, applying a combination of top-down (conceptual) and bottom-up (perceptual) processing, to arrive at appropriate descriptors. Periodical or data base indexing is less exhaustive than book indexing, since there is a need for more specificity in the indexing of each article included in the published numbers, and it uses a predominately top-down approach.

3 Conclusions

Indexers in their professional practice cannot depend solely on immediate perceptions to determine what a text is about; rather they must reference a multidimensional net of conceptual and linguistic entities stored in long term memory. This reference base is required in order to understand what the text in question is about and in order to assimilate new knowledge. When analyzing a text, the indexer searches the reference base for stored information about the theme of the text and if found comprehension occurs; however, if not found, comprehension is thwarted. An indexer working in various domains of knowledge is challenged to be a generalist, a person who understands the specifics, characteristics and terminology used by specialists in the different sciences. As the present study has hoped to demonstrate the understanding of Subject Analysis is

enhanced by adopting a generalist view that encompasses the findings of a number of interdisciplinary sciences.

In this sense, it is essential to identify subjects that are involved in the process of Subject Analysis, which facilitates its understanding and application in practice. It is important to emphasize, finally, that studies of this nature can contribute as a theoretical basis for new research about the topic, which is little explored in the literature, but fundamental for a more accurate and effective indexing, taking into account the information needs of the users.

Notes

Acknowledgments: We are very grateful to Professor Dr. Elaine Svenonius, of the University of Los Angeles (UCLA), for her patience in reading our imperfect English and for her valuable revision on content and style.

References

Allen, B. (1991). Cognitive research in information science: implications for design. // Annual Review of Information Science Technology 26 (1991).

Baranow, U.G. (1983). Perspectivas na contribuição da lin-güística e de áreas afins à ciência da informação. // Ciência da Informação 12:1 (1983) 23-35.

Basílio, M. M.de P. (1979). Interface lingüística e ciência da informação: potencialidades na análise das estruturas lexicais. // Reunião Brasileira de Ciência da Informação, 2, 1979, Rio de Janeiro. Trabalhos apresentados. Rio de Janeiro: IBICT.

Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness, intertextuality, and the cognitive act of classifying documents. // Journal of Documentation 42:2, 84-113.

Belkin, N. J. (1990). The cognitive viewpoint in information science. Journal of Information Science 16 (1990) 11-15.

Bell, H. K. (1991). Bias in indexing and aoaded langrage. // The Indexer 17:3 (1991) 173-177.

Birman, Joel (1994) Leitura crítica: questões sobre recepção. // Simpósio Nacional de Leitura, Rio de Janeiro, Leitura, saber e cidadania. Rio de Janeiro, 1994

Cintra, A. M. M. (1983). Elementos de lingüística para estudos de indexação.Ciência da Informação 12:1 (1983) 5-22.

Cole, C. (1994). Operationalizing the notion of information as a subjective construct. // Journal of the American Society for Information Science 45:7 (1994) 465-476.

Cunha, I.M.R.F. (1987). Análise documentária. // Smit, J.W. Análise documentária. 2.ed., Brasília: IBICT. 38-60

Dahlberg, Ingetraut. (1987) Classification and "TheTree of Cognition" (Editorial). // International Classification 14:3 (1987) 125-126.

Dahlberg, Ingetraut. Knowledge organization and terminology: philosophical and linguistics bases. International Classification 19: 2 (1992) 65-71.

Daniels, P J. (1986) Cognitive models in information retrieval: an evaluative review. // Journal of Documentation 42:4 (Dec.1986) 272-305.

De Mey, M. (1977). The cognitive viewpoint. Ghent: University of Ghent, 1997.

Dell'Isola, R.L.P. (1999). O contexto e a compreensão lexical na leitura em português-língua estrangeira. 1999. (Tese, Doutorado em Letras, Linguística Aplicada) Belo Horizonte: Faculdade de Letras da UFMG, 370p

Dias, E.W. & Naves, M. M. L. (2013) Análise de assunto: teoria e prática. 2.ed.rev. Brasília: Briquet de Lemos, 2013.

Farradane, J. (1976) Towards atrue information Science. // The Information Scientist 10:3 (Sept. 1976) 91- 101.

Farrow, J. (1995). All in the mind: concept analysis in indexing. // The Indexer, 19:4 (1995) 243-247.

Fujita, M. S. L. (2003) A leitura documentária do indexador: aspectos cognitivos e linguísticos influentes na formação do leitor profissional. 2003. 321f. Tese (Livre Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

Garcia Marco, F. J.; Esteban Navarro, M. A. (1993). On some contributions of the cognitive sciences and epistemology to a theory of classification. // Knowledge Organization 20:3 (1993) 126-132.

Giasson, J. L. (1993) A compreensão na leitura. Lisboa: Asa, 1993.

Gopinath, M. A. (1992). Cognitive paradigms in knowledge organization: technical on the 2nd International ISKO Conference. International Classification 19:5 (1992) 217-222.

Hannabus, S. (1988). Knowledge representation and information seeking. // Library and information seeking 37:3 (1988) 7-15.

Ingwersen, P. (1982). Search procedures in the library – analysed from the cognitive point of view. // Journal of Documentation 38:3 (1982) 165-191.

Izquierdo Arroyo, J. M. (1990) Esquemas de lingüística documental. Barcelona: DM, 1990, 1, item 1.2 http://www.mdp.edu.ar/humanidades/documentacion/licad/archivos/modulos/proces/archivos/bibliografia/procesamiento/Eje1/P004.pdf . (27/02/2018).

Kintsch, W.; Van Dijk, T.A. (1978). Toward a model of text comprehension and production. // Psychological Review 85:5 (1978) 363-394.

Kobashi, N.Y. (1996) Análise documentária e representação da informação. // INFORMARE, Cad. Prog. Pós-Grad. Ci. Inf. Rio de Janeiro 2:2 (jul./dez. 1996) 5-27.

Leiva, I.G. (2012) Aspectos conceituais da indexação. // LEIVA, Isidoro Gil; FUJITA, Mariângela Spotti (ed.). Política de indexação. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária. 260p

Luria, A. R. (1994). Desenvolvimento cognitivo. 2.ed. São Paulo: Ícone,1994.

Monday, I. (1996). Les processus cognitifs et la rédaction de résumés. // Documentation et bibliothèques 42: 2 (1996) 55-62.

Moreiro González, J. A. (1994). Documentación y Lingüística: conceptos de relación esenciales. // Ciências de La Información 25:4 (1994) 202-212.

Najarian, S.E. (1981). Organizational factors in juman memory: implications for library organization and access sys-tems. // Library Quarterly 51: 3 (1981) 269-291.

Navarro, S. (1988). Interface entre lingüística e indexação: revisão de literatura. // Revista Brasileira de Biblioteconomia e Documentação 21:1/2 (1988) 46-62.

Pinto Molina, M. (1995). Interdisciplinary approaches to the concept and practice of Written Documentary Content Analysis (WTDCA). // Journal of Documentation 50:2 (1995) 111-133.

Santaella, L. (1992). A assinatura das coisas: Peirce e a literatura. Rio de Janeiro: Imago, 1992.

Silva, M. R. da., Fujita, M. S. L. (2004). A prática de indexação: análise da evolução de tendências teóricas e metodológicas. // Transinformação 16: 2 (maio/ago 2004) 133-161.

Smit, J. (1978).Documentação e lingüística: interrelação e campos de pesquisa. // Revista Brasileira de Biblioteconomia e Documentação 11:1/2 (1978) 29-32.

Smit, J. coord. (1987). Análise documentária: a análise da síntese. Brasília: IBICT, 1987.

Van Dijk, T.A. (1976). Complex semantic information processing, 1976.

Walker, D.C. et al. (1976) Natural languages in information science. (pp.127-163). Stokholm: Skriptor. (FID 551), 1976.

Van Dijk, T. A & Kintsch, W. (1978). Cognitive psychology and discourse: recalling and summarizing stories, 1978

Dressler, W. U., ed. (1978). Current trends in text linguistics. Berlin: Walter de Gruyter, 1978.

Received: 2018-07-19. Accepted: 2018-10-05