Observing documentary reading by verbal protocol

**Mari?ngela Spotti Lopes Fujita, Maria Isabel Asperti Nardi
and Silvana Aparecida Fagundes**

**Departamento de Ci?ncia da Informa??o, Universidade Estadual Paulista (UNESP),
Mar?lia, S?o Paulo, Brazil**

Abstract

Verifies the applicability to research on indexers' reading strategies of the process observing technique known as Verbal Protocol or Thinking Aloud. This interpretative-qualitative data collecting technique allows the observation of different kinds of process during the progress of different kinds of tasks. Presents a theoretical investigation into "reading" and into formal methodological procedures to observe reading processes. Describes details of the methodological procedures adopted in five case studies with analysis of samples of data. The project adopted three kinds of parameters for data analysis: theoretical, normative, empirical (derived from observations made in the first case study). The results are compared, and important conclusions regarding documentary reading are drawn.

Introduction

Throughout the 1970s, researchers observed the activity of reading taking into account only individuals' answers to and product analysis of the comprehension of a text: however, as research developed, it was realized that it was also necessary to observe the reader when he turns the pages and runs his eyes over the text he is reading. That is to say, to observe reader's visible activities and record those thoughts that may influence his search for comprehension. These activities are properly called reading strategies.

In information science, "thinking aloud" techniques have been used in information seeking and interactive information retrieval research since the late 1970s, for example, in the work of Ingwersen, whose Information retrieval (IR) research line has adopted a cognitive approach to IR processes. In a project, which collected data from twenty librarians, Ingwersen et al. (1977) applied thinking aloud techniques to obtain data about the negotiation process in a reference situation of information searching. Ingwersen (1982) reports the main results of an investigation conducted by his research during the period of 1976-1980. The investigation, focusing on cognitive aspects of the information transfer processes in public libraries reference services, was partly experimental, with the use of thinking aloud techniques. It discussed the use of verbal protocols, its advantages and disadvantages, in connection with studies of the interpretation by users of retrieved texts, relevance assessments and search strategy shifts.

It is interesting to note that, in applying thinking aloud techniques, Ingwersen gave special attention to aspects of the technique and has carried out actions to guarantee data reliability and validity which have been observed in the research reported here: collecting data from subjects in a natural situation, in their own professional environment, without inserting any changes in their routines; supplementing thinking aloud data with observation of the subjects' behaviour and actions; and training the subjects to make them familiar with the technique.

Later the techniques have been applied often in library and information science research as well as in human-computer interaction studies worldwide, including processes of professional summarizing, for example, by Endres-Niggemeyer and Neugebauer (1998). However, it is important to make it clear that Ingwersen has made use of thinking aloud techniques in works that focus on information retrieval processes, while Endres-Niggemeyer and Neugebauer's have focussed on summarizing processes in general. Thus, the present research has its main value in its objective, which differs from the objects of the other authors and constitute a novelty in the information area: it focuses on reading processes for indexing purposes.

Observation of reading in documentary analysis is important since reading represents the first step, which establishes and

influences all other performance operations. However, reading for documentary analysis is different from reading for other purposes because it is directed towards indexing. In this way, we believe that studying the process of reading for documentary purpose may bring some orientation to indexers and improve the use of indexing methodologies. Therefore, a project of integrated research called "Reading for Documentary Analysis", in which the indexer's reading is observed during the indexing process, is being developed at UNESP-Mar?lia, coordinated by Dr. Mari?ngela Spotti Lopes Fujita. Two reports have already been published. (Fujita, 1996; 1998).

Our main premise is that reading is a fundamental activity of documentary analysis because it results in the selection of concepts and terms, which will represent the document for the user. Therefore, we give emphasis to the fact that the identification and selection of terms take place during reading, that is why observing reading strategies is important. Thus, investigation into reading in documentary analysis focuses on the indexer as a reader who reads documents in order to identify and select concepts. Reading made by an indexer with such a purpose will be called "documentary reading".

To begin with, a theoretical investigation into reading was carried out in order to describe different conceptualisations of reading, reading strategies and also formal methodological procedures to observe reading processes. Following the bibliographical review, it was possible to elaborate a theoretical foundation ranging from the view of reading as a linear process to reading considered as communication, revealing the notion of reading strategies as subsidiary to the study of reading strategies regarding documents. (Fujita, *et al.*, 1998).

In addition to the theoretical foundation, it was necessary to choose a method to observe the indexer's documentary reading process. Although, in the original project, description of the documentary reading processes was intended to be founded only on interviews to be conducted with indexers, it was decided that observation of the process would also be important because it might add support to the interviewee's statements and should be the ideal method to identify different strategies of documentary reading.

Thus, the investigations also verified the applicability of the Verbal Protocol (Ericsson and Simon, 1987) This interpretative-qualitative research data collecting technique allows the observation of different kinds of processes during the progress of different kinds of tasks. Process observations can provide information about individual processing steps such as sequences of eye movements or spontaneous verbalization that can externalise mental processes, keeping the sequence of processed information.

In Brazil, the use of verbal protocols to observe reading processes for the purpose of indexing is novel. We have not heard about any other research that makes use of this methodology.

In foreign literature, the use of verbal protocols to observe indexing processes has been reported only by the Japanese researcher Gotoh (1983) in an article which discusses problems of information processing behaviour in the human indexing process. In that study, the author carried out an experiment with two indexers using verbal protocols.

According to Ericsson and Simon (1987), verbal protocols consist in recording the verbal externalisation of thought during an activity. This is possible because humans are able to externalise their mental processes while information is in the focus of attention. When they do it consciously, the authors consider that they are "thinking aloud". In this way, each subject's words are audio recorded and transcribed literally, resulting in verbal protocols. Protocols are generally defined as verbal reports of subjects' conscious mental processes.

According to Cavalcanti and Zanotto (1994), verbal protocols were introduced into qualitative research through psychology in 1980, and since then its validity in disclosing mental processes has been questioned. Following the triumph of behaviourism over competing viewpoints, verbal protocols have been regarded as suspect (Ericsson & Simon, 1980). According to Ericsson and Simon (1987), when cognitivism gained evidence as a new paradigm, they came again into scene as the main source of data for cognitive research. That revival occurred within the theoretical framework of information processing related to studies of problem solving.

Since then, they have gone beyond the limits of cognitive psychology, and have been adopted by the area of applied ;inguistics, in which they have become important in investigations of reading in a foreign language (Hosenfeld, 1977; Cohen & Hosenfeld, 1981; Faerch & Kasper, 1987). Undoubtedly, research on reading is the chief field in which verbal protocols are used (Cavalcanti, 1983;Paschoal, 1988).

In Brazil, particularly in the Postgraduate Programme in Applied Linguistics at the Pontifica Universidade Catolica de S?o Paulo, protocols have been used as research tools in research for many dissertations and theses. Among these researches, Nardi's (1993) was taken as methodological model for research in our Integrated Project. According to Nardi (1993), although controversial, verbal protocols are now the only available data-collecting instrument which enables researchers to observe the reader's processes during his activity of text comprehension, which accounts for our choice of the method.

Santos (1996), supervised by Nardi, opened the possibility of using verbal protocols to observe processes of documentary reading within the Documentary Analysis Research Group, and, in the first stage of the integrated project Fujita (1998) carried out a theoretical study on reading and documentary analysis reading strategies, and also the first case study. In the second stage, four more case studies were carried out with the same objectives and the same methodology. The results were compared, and important conclusions regarding documentary reading were drawn.

In this paper, the next section details the procedures for applying verbal protocol methodology: those carried out before audio recording reading sessions (research texts selecting; research subjects selecting; informal talk with subjects; and carrying out of individual sessions of familiarization with the "Think Aloud" technique); those carried out during the recording of protocols ("Think Aloud" recording during reading tasks), and those that followed the recording of the reading tasks (retrospective interview recording and literal transcription of the recordings). In this section, we also present the five case studies carried out by the project and some specific procedures relating to specific cases. The third section discusses the theoretical foundations for data analysis (theoretical parameters and normative parameters). The fourth section presents the analysis of the transcriptions of the protocols. The fifth section presents and discusses empirical parameters: strategies that are specific of documentary reading which were observed in the protocols of the first case study and which were used as parameters for the other case studies. The final section presents analytical comments on the applicability of verbal protocols for documentary reading research.

Applying verbal protocol methodology

The recording of "thinking aloud" for the observation of indexers' and summary writers' reading processes, in the five case studies, followed the same general methodological procedures.

Procedures before recording the verbal protocols

Selecting the research texts

- Text typology: to be chosen, the text should not have been indexed by any of the indexers-research participants.
- Theme of the text: specific to the area in which the subjects works.
- Expert counseling: in order to facilitate analysis of data, it was advisable to consult a specialist in the text area to obtain his interpretation of the text content.

Selecting subjects

Criteria for selecting subjects were years of experience in the information system and on indexing and/or abstracting activities. In addition to these criteria, "indexing or abstracting skills" set by number of documents indexed was taken into account.

Informal talk with the subjects

First, informal talks were held with the subjects in which individual appointments were made for the data collection session. During this talk, the research purposes were clearly stated, emphasizing the importance of the work for the development of the documentary analysis field. It was also made clear that the anonymity of the subjects would be preserved, so that they would feel comfortable during their performance on the reading task.

Becoming familiar with the "thinking aloud" task in individual sessions

In each session in which protocol data were collected, before handing the research text to the subject, the researcher read aloud instructions (Appendix 1) to the subject and modelled a "thinking aloud" task for him, aiming at making him understand the nature of the task. When the subject was not able to "think aloud", his data were rejected.

Procedures during verbal protocol recording

"Thinking aloud" recording during reading

The research text was handled to the subject and he was reminded that he would be expected to "think aloud" all the time during reading, which might be understood as him having to try to externalise his mental processes. Also, it was emphasized that the subject should try to forget the presence of the researcher in the room, for the researcher would be

there only to remind the reader that he had to think aloud all the time, to call his attention back from digressions of attention and also to control the tape recorder.

Procedures following verbal protocol recording

Retrospective interview recording (optional)

Immediately after thinking aloud recording, in the same session, a retrospective interview was recorded, in cases where there was a need to clarify some points considered obscure by the researcher. (Such points were usually sequences of the reading and thinking aloud in which the researcher thought it difficult to identify which strategies the subject was using - and took note of the sequence in order to be able to ask the subject about it later).

Literal transcription of the recordings

Transcribing should be done so that the subjects' comprehension, doubts, mistakes, term identification and selection were visible. For better understanding of the processes of the subjects, we made use of specific notations adapted from Cavalcanti (1989)

[.]	passage of the text verbalized by the subject at the first reading
Italic	subject's comments showing his comprehension
?	pauses and continuation of reading
< - -	subject returns to previous passages of the text
bold	terms selected by the subject
(- >)	subject "jumped" (ignored) passage of the text during reading
/	auto-interruption of a thought
((SL))	subject speaks and laughs at the same time
((MT))	subject mutters (meaning irony)
((LG))	subject laughs
(-> -> ->)	subject accelerates the reading rhythm
(~~~)	subject reading at a slower speed, with attention
"."	word or expression commented upon by the subject
{}	inclusion in the transcriptions of descriptions of the subject's meaningful gestures or the researcher's analytical comments
(?.)	omission of an irrelevant passage of transcription
Underlined	relevance of the passage for the reader
Underlined and bold	sequences that best express the phenomenon under analysis

As example, we have chosen the transcription of one subject in order to show the use of some notations (underlined) and their potential for making the analysis of data easier. It is important to note that the reader or indexer underlined some words, which he considered candidates for key words.

> [.*the contamination of the oceans by anthropogenic radionuclides*.] *well, the text presents an abstract which makes things easier? if the abstract is good, I can rely on it.well, let's take a look at this abstract.it should be very short indeed.here he {the author of the text} starts telling the objective of this paper (<-) (~~~) [This paper / summarizes the main sources of contamination in the marine environment and presents an overview covering the oceanic distribution of anthropogenic radionuclides in the FAO regions. A great number of measurements of artificial radionuclides have been carried out on various marine environmental samples in different oceans over the world, cesium-137 being the most widely measured radionuclide] here he does not explain if it has been the most measured because more experiments have been conducted or it has been the most measured because it has been the most found.this has become faulty somehow [radionuclide concentrations vary from region to region according to the specific sources of contamination. In some regions, such as the Irish sea, the Baltic Sea and the Black Sea, the concentrations depend on the inputs due to discharges from reprocessing facilities and from Chernobyl accident]?* **Chernobyl accident.still**? ((FR.My

God!.when was this text written?. 1996?.not long ago.)).but those words are going to be selected, I'm sure.as I'm a chemist, I always like to choose some words of my specific field, when there's something to do with it. I always try to give emphasis to passages dealing with chemistry.here something from the area of chemistry is going to be selected.well, let's take a look at the artwork.

{the subject turns the page and continues}.[radionuclides behaviour in the marine environment].{the subject examines the artwork}.[Radionuclides level in oceans], *with some regions of the world, fish in some regions of the world (?).* **toxic material** *is important, is another key-word.***marine ecosystem**, *another key word.here he says that it is important to monitor, but the subject is not exactly only about that.*

{the subject goes back to the first page} (<-) [? the oceanic distribution of anthropogenic radionuclides in (-> -> ->) measurements of artificial radionuclides have been carried out on (->->->) samples in different oceans over the world].*it would be a quantitative analysis here*? [another radionuclide highly harmful and dangerous even in the smallest amounts is plutonium-239 ?] *yes, I cannot choose this term because it's too specific.it's got to be* **radionuclides** *because it is more general.*(->) [Artificial radionuclides scattered in the oceans come from tests with nuclear artefacts, nuclear accidents and radioactive material liberation].

*{the subject turns over the pages looking for the words he had underlined and reading them aloud as they were encountered}.***radionuclides**, **anthropogenic contamination**, **oceans**, **quantitative analysis**, **radionuclides**, **ecosystem**, **food chain**, **toxicity**. *and maybe other words which could be added to those would be* **chemical**, **physical** *and* **biological interactions**.*but I don't believe there are descriptors such as those.*

Description of some specific procedures according to the specific case study

The investigations on the process of documentary reading, from the documentalist's point of view, have focused on several aspects: a) different documentary processes: during document indexing and abstracting, b) different textual structures: scientific journal and newspaper articles, and c) different knowledge areas: odontology, nuclear energy and agronomy.

Thus, five case studies were conducted which have emphasized different aspects:

- the first, carried out in the context of Brazilian university libraries in the area of oral health sciences, explored an academic, scientific text in the area of odontology (Santos, 1998: Fujita, 1999a);
- the second, carried out in the context of one of Brazil's main newspapers, "O Estado de S?o Paulo", explored a newspaper article (Fagundes, 1997);
- the third and the fourth took place at the Centro de Energia Nuclear (CIN) of Comiss?o de Energia Nuclear, in Rio de Janeiro, and explored scientific journal articles with the objective of observing reading activity for indexing (Rubi, 1999) and for abstracting (Iwashita, 1999);
- the fifth, carried out at the Coordena??o Geral de Documenta??o em Agricultura (CENAGRI), also examined scientific journal article on agronomy (Pires, 1999).

Counting all five case studies, verbal protocols were applied to fourteen subjects: nine were indexers with degrees in librarianship, and five were indexing and abstracting specialists in the area of nuclear energy. Figure 1 shows the studies carried out:

Title of the Case Study	Institution	Number of subjects and their education	Documentary typology	Knowledge area of the document
Case 1 Documentary reading by indexers in the area of Odontology	National Net of Oral Health Sciences (Sub-Rede Nacional em Ci?ncias da Sa?de Oral)	Indexers with degrees in Librarianship (4)	Scientific journal article	Odontology
Case 2 Documentary reading by indexers of newspaper	Filing Department of the newspaper "O Estado de S?o Paulo"	Indexers with degrees in Librarianship (1) and journalism (1)	Newspaper article	Sports, police occurrences

articles				
Case 3 Documentary reading by indexers of the Area of Nuclear Energy	Nuclear Information Centre (Centro de Informa??es Nucleares-CNEN)	Indexers with training in Nuclear Energy (4)	Scientific journal article	Nuclear energy
Case 4 Documentary reading by **abstractors** in the field of Nuclear Energy	Nuclear Information Centre (Centro de Informa??es Nucleares-CNEN)	Abstractors with training in Nuclear Energy	Scientific journal article	Management
Case 5 Documentary reading by indexers in the field of Agronomy	Coordena??o Geral de Documenta??o em Agricultura-CENAGRI	Indexers graduated in Librarianship (4)	Scientific journal article	Agronomy

Figure 1: Case studies using the verbal protocol for documentary reading

The methods followed in all case studies were those already described as the Integrated Project general methodology. However, during the research, the need was felt to adapt some procedures to improve the methods. The procedures for the "selection of research-text" and for "selection of subjects" were adapted to the documentary typology and to the knowledge area of each case study. Other procedures did not have to be changed. The details of the procedures followed in the individual cases are set out in Appendix 2.

Analytical procedures of each case study according to specific objectives.

Case study 1

Transcription analysis of the first case study used the three parameters: theoretical considerations concerning meta-cognitive features of documentary reading, trying to identify the meta-cognitive strategies listed by Brown (1980); the normative aspects, associated with the strategies of ISO Standard and the Indexing Handbook of the Latin American and Caribbean Centre on Health Sciences (BIREME); and the empirical parameter. Thus, emphasis was given to the following aspects:

- Making the reading objectives explicit;
- Identifying the important aspects of a message;
- Allocating attention to important areas;
- Exploring textual structure;
- Behaviour monitoring;
- Engaging in reviewing and self-questioning to check whether the reading objective is being reached;
- Making corrections when mistakes of comprehension are detected;
- Recovering attention from distractions or digressions;
- Associating with documentary language;
- Maintaining thematic coherence;
- Selecting concepts for indexing.

However, for case studies 3, 4 and 5 Kato's reformulation of the aspects listed by Brown (1980) was added to the theoretical considerations, and the normative parameter only took into account the ISO Standard, leaving out the BIREME Handbook.

Case study 2

In the data analysis of this case, we tried to verify whether indexers, during reading, make use of the meta-cognitive strategies listed by Brown (1980) in association with those described by the ISO Standard 5963 (1985), as described in

case study 1.

We also tried to observe whether indexers possessed knowledge about the textual structure of a newspaper text (Van Dijk, 1983), and if they explored it during reading. The following proposal for the conventional superstructure of news discourse was employed:

1. Summary/introduction
1.1 Headlines (with super-, main-, and sub-headlines, and captions)
1.2 Lead
2. Episode(s)
2.1 Events
2.1.1 Previous information
2.1.2 Antecedents
2.1.3 Actual events
2.1.4 Explanation
2.1.4.1 Context
2.1.4.2 Background
2.2 Consequences/reactions
2.2.1 Events
2.2.2 Speech acts
3. Comments
3.1 Expectations
3.2 Evaluation

Case study 3, 4 and 5

In these studies, concerning the theoretical parameter, we took into account Kato's (1987) reformulation of the meta-cognitive strategies listed by Brown (1980), which reduces all of Brown's strategies into two categories: a) defining objectives for reading and b) comprehension monitoring. As to the normative parameter, we considered only the ISO Standard, not the BIREME manual.

Theoretical foundations for data analysis

As may be seen from the transcription extract shown above, it is easy to observe the procedures of a reader who reads passages and makes comments about them immediately. However, identifying strategies depends on different parameters related to reading comprehension as a cognitive activity. Based on these parameters, it is possible to observe what kind of strategies the reader or indexer has made use of, and ,therefore, to determine the extent of his proficiency in documentary reading which will result in better identification of valid concepts for document retrieving. The Project adopted two kinds of parameters for data analysis:

- **Theoretical parameter**: theoretical considerations on the meta-cognitive nature of the reading strategies listed by Brown (1980);
- **Normative parameter**: strategies connected with the ISO Standard 5963 (1985) for theme identification.

Figure 2 shows the strategies typical of each parameter:

Brown's meta-cognitive strategies	Strategies specific to documentary reading proposed by ISO 5963
- Making the reading objectives explicit and/or keeping them in mind; - Identifying important aspects of message; - Allocating attention to important areas: exploring textual structure - Behaviour monitoring by: - reviewing and self questioning in order to check if objectives are being reached; making corrections when mistakes of comprehension are detected; - recovering attention from distractions or	- Analysis of the document with textual structure control; - Identification of concepts (systematic approach through questioning); - Selection of concepts

Figure 2: Reading strategies according to different parameters

Theoretical parameter: different views of readers' behaviour towards the reading of a text and notions of metacognition and strategies

Text and reader interaction develops by means of the use of strategies, defined by Brown (1980) as any deliberate and planned control of activities, which lead to comprehension.

According to Giasson (1993), "the reader approaches reading activity with cognitive and affective structures of his own. In addition to that, he makes use of different processes which allow him to understand the text."

There are different kinds of processes of comprehension that take place at different levels and simultaneously. According to Giasson (1993) there are processes for understanding a sentence (micro processes - sentence level); for achieving coherence among sentences (integration processes - among sentences); for building a mental model of a text (macro processes - textual level); for allowing reader to grasp fundamental elements and raise hypotheses (elaboration processes - text comprehension).

Metacognition in reading allows the reader to have an understanding of his own comprehension, that is, allows him to follow and evaluate his own process of comprehension during the reading of a text and, furthermore, to take necessary steps when comprehension fails (Leffa, 1996).

Kato (1987) distinguishes two kinds of strategies taking into account the degree of consciousness involved in them: the meta-cognitive, understood as the reader's conscious actions, are directed towards an objective or towards the search for a solution to comprehension problems; and the cognitive, used during fluent reading, without obstacles, which are automatic subconscious actions.

The involvement of consciousness is the criterion generally used to distinguish cognitive activities from meta-cognitive ones: cognitive activities would take place below the level of consciousness; the meta-cognitive ones would involve conscious introspection (Brown, 1980). The use of strategies is not easily observable because mental actions, as connections and deductions during reading, cannot be seen, although they can be verbalized. To assign meta-cognitive character to mental actions, Brown (1980) lists the following activities:

- clarifying the purposes of reading;
- identifying the aspects of a message that are important;
- allocating attention so that concentration can be focused on the major content area rather than the trivia;
- monitoring ongoing activities to determine whether comprehension is occurring;
- engaging in review and self-interrogation to determine whether goals are being achieved;
- taking corrective action when failures in comprehension are detected;
- recovering from disruptions and distractions;

and many more deliberate, planful activities that render reading an efficient information-gathering activity.

The use of cognitive and meta-cognitive strategies must tend to a balance, as according to Cintra (1987) who considers that, although any reading involves both kinds of strategies, a text is likely to be more legible if it requires fewer meta-cognitive activities. However, mere automatic reading is likely to lead to incomprehension.

Concerning this, Kato believes that there are moments in reading where a difficult passage requires careful and linear reading, and there other moments in which context-based inferencing allows fluent comprehension. Therefore, she considers proficient a reader who is able to use both kinds of strategies, the ascending ones (which depend on careful analysis of visual input, including language) and the descending ones (based on the reader's prior information and his ability of inferring and predicting), making use of each type of strategy, in a conscious way, at the moment where each of them is required. Moreover, we believe that proficient-strategic reader is the one who not only uses ascending and descending strategies properly but also keeps in mind the reading objective.

The theoretical conception of reading strategies on documentation presented by Cintra (1987) is in accordance with Cavalcanti (1989) when it states that reading for documentary purposes requires author-reader cooperation, inasmuch as the author cannot foretell who will read what he has published. Moreover, it does not recommend linear reading, letter-by-letter, word-by-word, but that the reader should skip parts of the text based on what he is able to predict from his

knowledge of text structure. In addition, Cintra (1987) believes that a reader who easily recognizes textual superstructures can understand the main ideas of a text better than a person who only performs linear reading. Relying on textual structure and his prior information about the subject, the reader can infer meanings and raise hypotheses, which help him to grasp the global thematic coherence.

Theoretical studies of the field show that indexers understand the text in the same way as fluent readers do, but are influenced by specific conditions (very limited time, text comprehending for the purpose of indexing, great number of different types of texts and knowledge areas, repetitive element in his work, etc). In addition to this, when indexers have to choose concepts (which will require exhaustivity and specificity), they will be directly influenced by their awareness of the needs of their usual clients and by their familiarity with the system's indexing policy.

The documentalist reader, even when he is not a specialist in the subject matter of a text, interacts with the text using his knowledge of a specialized documentary language, of different textual structures and of the nature of the information system with which he works. This interactive aspect of documentary reading echoes general reading theories such as those of Cavalcanti (1989) and Giasson (1993), which have described the process of reading as involving the interaction of different knowledge sources.

Normative parameter: strategies associated with the ISO Standard 5963

The ISO Standard 5963 (1985) on methods for examining documents, determining their subjects and selecting indexing terms, suggests the following stages for indexing: examination of the document, identifying concepts and selecting concepts. Particularly with regard to identifying concepts, it suggests a "systematic approach" to the text through questioning. In this way, the Standard suggests a method for document analysis in which the document analysis and synthesis are processes that involve the following stages:

Examination of the document: at the same time as it indicates the necesssity for an overall reading of the text for a full comprehension, the standard points at the unviability of such a procedure; for this reason, it suggests that the indexer can be successful if he pays attention to the following important parts of a text:

- the title;
- the summary, if there is one;
- the list of contents;
- the introduction; the sentences that begin chapters and paragraphs, and the conclusion;
- illustrations, diagrams, tables;
- highlighted words or group of words.

At the end of that item, the Standard points out that it is not possible to analyze a subject only through reading the document title or summary.

Identifying concepts: after examining the document, the indexer must follow a systematic approach to identify concepts, which are essential elements for subject description. Therefore, the Standard suggests the use of some questions prepared to accomplish that aim:

- Does the document possess in its context an object under the effect of an activity?
- Does the subject deal with an active concept (for instance, an action, an operation, a process, etc)?
- Is the object under the influence of the identified activity?
- Does the document have an agent, which has performed the action?
- Does this agent refer to specific ways to perform the action (for instance, special tools, techniques or methods)?
- Are all those factors considered in the context of a specific place or environment?
- Are any dependent or independent variables identified?
- Has the subject been considered from a point of view, usually not related with the field of study (for instance, a sociological or religious study)?

According to Lara (1993), documentary reading has the objective of "identifying and extracting references from original texts to transform it into a documentary text". The author cited the "essential concepts" of the ISO Standard, which we consider as documentary reading strategies of a meta-cognitive nature.

In the item "identifying concepts ", the Standard approaches the issues of term selection by suggesting that "indexers do not necessarily need to represent all concepts identified during document examination making use of indexing terms. Concepts will be selected or rejected according to the purpose for which they will be used."

Then, the Standard mentions exhaustivity and specificity as aspects, which may define the choice of concepts. However, it recognizes that those aspects are related to and dependent on two influencing variables in identifying and selecting concepts: the documentary system and the user of that system.

Exaustivity and specificity are two very important principles that guide both the indexing process and the retrieval process. In indexing, these principles are part of an indexing policy, which influences the quality of retrieval. Exaustivity, in indexing, is related to the number of terms attributed to the content of a document, which has to be proportional to the information required by the users of a system of information retrieval. Specificity relates to the specific content of each term attributed to each topic of the document and involves the decision of selecting only the most specific terms, not the generic ones, to represent the content of the document.

We understand that the "Indexing stages" of the Standard (first the *document examination* and then the *identification of concepts*) may be considered as meta-cognitive strategies of reading, since the first one suggests exploring the textual structure and the second may correspond to identifying important aspects of the message, as pointed out by Brown (1980). Moreover, selecting concepts, meant by the ISO Standard as following the identification of concepts, may be considered as the "outcome" of the indexer's reading.

Analysis of the protocol transcriptions

Considering the theoretical and the normative parameters and also those derived from the observation of data of the cases studies, we will now present examples extracted from the transcription of verbal protocol data of the third case (Rubi, 1999), which can clearly show the readers making use of different strategies:

- *Making the reading objectives explicit and/or keeping them in mind*

   .[this paper summarizes.]. *now this is the most important part in which the author says to the reader what .* **one wants information to be concise and easy for the reader, for the person who is consulting it..**
   .[. present in marine environment and to estimate the dosage to which population is exposed by consuming contaminated marine products] *perfect!.then here in this sentence he explains this paper objective? then **this is an inestimable source of terms for indexing.***
- *Identifying important aspects of the message* .*well,* **the most important thing of this text is already in the title itself**.*the title itself is already showing that the most important thing here, as I've read in the abstract, is exactly* **the contamination of oceans by anthropogenic radionuclides** .
   .[. in biological processes, cesium and strontium, chemically similar to potassium and calcium tend to follow them by laying themselves partially on muscles and bones respectively.] ***This sequence is nice! These heavy materials cause radioactivity effects in the bones.and here, I'd have to choose a concept for biological effects of radioactivity.***
- *Allocating attention to important areas*

   .[This paper summarizes .] ***now this is the most important part*** *in which the author says to the reader what .*
   .*well,* **let me take a look at these illustrations and subtitles** .*conclusion?*
   . **this sentence here is important**.*it is in itself a summary of this paper.*
- *Exploring textual structure*

   .*well,* **let me take a look at these illustrations and subtitles** *?conclusion?*
   .*I'm going to take a look at the conclusion because I want to clarify what kind of text this is.I think **this text does not have a scientific paper structure** with objective, methodology, etc.it does not have a methodology, it's not a research.*
- *Behaviour monitoring to check if comprehension is being achieved by engaging in reviewing and self-questioning.*

   .*let me examine inside (**the body of the text**) because sometimes we can miss something important?*
- *Association with documentary language*

   .[sources of anthropogenic pollution] *I think **there's a descriptor with that name**.*
   . [chemical, physical and biological interaction].*but **I don't believe that there's a descriptor for this.***
- *Maintenance of global thematic coherence*

   .*well,* ***the most important thing of this text is already in the title itself***.*the title itself is already showing that the most important thing here, as I've read in the abstract, is exactly* **the contamination of oceans by anthropogenic radionuclides** .

- ***Selecting concepts (terms or descriptors) (Standard or Handbook)***

  *.well, **I will underline the probable descriptors**. **artificial radionuclides.** I don't know if there is such a descriptor. I will have to check?. if there isn't, I'll have to find something appropriate.*

  .so those are the key-words .{the subject turns over the pages looking for the words she had underlined}. radionuclides. anthropogenic contamination.. oceans.quantitative.analysis.qualitative analysis. radioactivity .ecosystem . food chain.toxicity.and maybe, besides these words, what could also be selected would be chemical, physical and biological interactions.but I don't believe there is a descriptor for that.

Empirical parameters: other strategies and new aspects observed in the process of documentary reading

In the analysis of the protocols of the first case study (Fujita, 1998), we observed aspects related to the indexers' behaviour during reading that are specific to documentary reading and that reveal important results considered as new contributions for the understanding of the process of reading for indexing purposes. These aspects were used as parameters for the other four case studies that followed.

The first case study took place at the Oral Health Sciences Information Subsystem of BIREME formed by Odontology libraries. It was the first application of the verbal protocol technique for the observation of indexers' reading. Some aspects identified in the analysis of these first protocols which had not been predicted by theoretical and normative parameters, such as *connection with the documentary languages used by the system*, *maintenance of thematic coherence* and *selecting concepts*, were evaluated for use as parameters for further research.

We consider it important to discuss these aspects in detail displaying examples extracted from the protocols:

Identifying concepts

The most important finding is that indexers consider the abstract as the main source of concepts identification and, probably, are not aware of the role of text structure in the prediction of important parts that contain important concepts.

The observation of the process of concept identification showed evidence of the use of different strategies, without consistency among subjects: subject 1 used many strategies, including spontaneous questioning; subject 2 and subject 3 associated the concepts found with the documentary language used by the information system they served and underlined the words that represented the concepts; subject 4 extracts most of the concepts from the abstract, and then, during the reading of the whole text, only confirms whether the concepts selected were adequate.

The lack of consistency in concept identification seems to accord with the lack of a pattern in the sequence of operations observed in the beginning of the analysis.

From the results, we may consider that identification may depend on the indexer's skill of exploring textual structure, that is, we believe that if the indexers were able to rely on textual structure while trying to identify concepts, the extracted terms would be more representative of the text, and at the same time, coherent with the retrieval language.

This conclusion is based in the observation that during reading, indexers showed evidence of non-awareness of the role of textual structure in concepts identification. They did not pay attention to parts as "Material and Methods", "Conclusions", etc.. which could have guided their search for concepts.

The identification of concepts to represent a document must be well structured once its products (indexes and abstracts) must satisfy the demands of the users. We believe we need to generate a systematic method for this operation and check its efficacy.

Exploring textual structure

The results reveal that, although, sporadically, we could observe subjects exploring textual structure in detail and very rapidly, this did not always result in concept identification and, furthermore, when concepts were selected, they were not necessarily coherent with the users' language. The use of textual structure exploration should guarantee

identification of concepts.

We believe that the use of the strategy of textual structure exploration related to the identification of concepts may facilitate documentary reading and guarantee consistency of procedures for the thematic treatment of information.

Selection of concepts

The operation of selecting concepts was clearly observed during text reading almost always after identifying concepts, and it has become explicit that there are two distinct operations used by indexers: identifying concepts and selecting concepts . Some indexers preferred to select all terms by the end of the reading activity, and others did it during the activity, simultaneously identifying and selecting concepts. Observe the examples below:

- **Subject 1**: the subject selected concepts for indexing on 5 occasions during reading.
  *.well from this point on, I will be observing the primary terms which I am going to include in this summary that are: **free gingival graft analysis, cicatrisation** too, and then we are going to look at the part about the results.*
  *.**well, the terms I encounter I select right away**.gum obviously, because the text deals with gum..*
  *. I'm writing here **gingival.***
- **Subject 2**: He selected concepts only on one occasion (at the ending of reading activity).
  *.now, I am going to review all the terms I have underlined to find out whether there is anything else I can underline and select the final terms...let's see. **human . surgical . cement. autogenous graft. autogenous graft** and **gingival graft** are the same thing. there are some differences.but gingival graft is more specific. if I don't have the term gingival. I'll select autogenous.clinical cicatrisation I'm going to consider secondary.*

Data shows evidence of concept identification being carried out *during* reading and, therefore, there are two distinct operations: **concept identification** and **concept selection.**These two operations occur *during* (not after) reading, which means that the selected concepts are the result of the interpretation of textual content. This implies the need to distinguish between concept selection during content analysis and concept selection in the translation of concepts to documentary language.

Making the reading objectives explicit and/or keeping them in mind

According to the analysis of the four subjects' strategies, "making the objectives of reading explicit" was the one for which data have offered the least evidence: while subjects 2 and 4 showed evidence of maintaining the objective of reading (concept selection) in mind only in a single moment of the whole process of reading, subject 3 mentioned it only in the retrospective interview and subject 3 did not offer any evidence at all.

The objective of reading for indexing purposes is to represent the text for future retrieval by the system users. The fact that the subjects have not mentioned this objective during the identification of concepts has led us to the following reflection: the purposes of the policy of an indexing system should relate to the importance of representation for the retrieval and this orientation should be present in all indexers' training programs. Besides that, the indexer should be advised to get involved in services to the user to get familiar with database searching strategies.

Association with documentary language

By observing sequences of operations during reading, we have identified what we have called "association with documentary language". That operation, not predicted at first, was easily observed, as we can see in the examples extracted from the transcription of the protocols:

- [Clinical Analysis of Free Gingival Graft Areas] *the main subject that I see here is **free gingival graft**, **I am going to check at the DeCS** if there is the uniterm "free gingival graft". If there is that term, I'll also check if there are the qualifiers related to it .*

- [(<-) it has been assessed clinically and comparatively in humans] . *I've underlined "**humans**", it is not performed on animal, the influence of surgical cement, I underline **surgical cement** that is a term of odontology,.*

.he is using "surgical cement", then, I underline in order to know if those descriptors are going to be primary or secondary, whether it will be meaningful in the text or not."

Association with language occurs at different moments, but almost always simultaneously with the "identification of concepts". Taking into consideration that indexers are not always specialists in the text area, getting familiar with subject is achieved by the language of the system which functions as part of the indexer's previous knowledge. According to theoretical foundation, previous knowledge retrieving is a powerful strategy for reading comprehension.

Maintenance of thematic coherence

In this operation, the indexer gives evidence of his global comprehension of the text not only by the use of his knowledge of which parts of the text are expected to present the main ideas, but also by relating the theme with other parts of the text where the secondary ideas are expected to be encountered. In this way, even if the indexer does not mention all the main topics of an article, he is able to extract from its content terms (descriptors), which he considers important to represent it in a concise way, as the sequences extracted from the protocols can show:

- *.the main focus of my text is in this paragraph.*

- *.I don't have to take a look at this part, because I already know that it isn't important for indexing, then**, I'll take a look at the results**.*

- *.the author hasn't said this, he only cited, it's not even a secondary item.. It's something not worth mentioning. not even as a secondary point.and **in the conclusions I expect to find exactly the major worry of the text**.that means that from the conclusions I can decide which information that I included in the summary can be excluded.I will be able to decide what is really important.. Therefore, I'll examine the conclusions.*

The indexer-reader: metacognition and reading strategies

Considering the results obtained by the analysis of strategies observed by the verbal protocol technique, we conclude that, from the cognitive point of view, the indexer is potentially a proficient reader for his innate and constructed cognitive structure: his previous knowledge should comprise linguistic knowledge, textual structure knowledge and world knowledge which can be used by conscious retrieval of adequate schemata through meta-cognitive processes.

The indexer should be more proficient, more strategic in reading than the ordinary reader once he is a professional reader, once he deals with a great amount of reading. He will need to be aware of the strategies he can use and to be able to consciously select the adequate strategy for each reading event.

We also understand that, as reading has a strong cognitive component, an essential cognitive support for reading comprehension is knowledge of the typology of textual structures and knowledge of the content of the specific area one serves (which can be obtained through experience in the area). Another important factor is that indexers should become aware of their innate and constructed skills, should improve their linguistic knowledge and should be advised as to the importance of conceptual analysis.

Final considerations: analytical comments on the applicability of verbal protocols for documentary reading research

Verbal protocols were successfully used in this Project once it was made possible for researchers to observe the indexers' reading processes in the process of indexing activities, confirming theoretical hypotheses about reading and the use of strategies and, moreover, revealing new aspects of documentary reading. Verbal protocols can reveal a reader's introspection in a natural way, and have advantages over other kinds of introspective techniques, like diaries, questionnaires or interviews, because it is the only one that provides direct access to the mental process of reading *while it is being carried out* by the subject. Considering this, it is the only really introspective technique while the others are of retrospective nature.

However, we must report that, during the first informal talk with the subjects, all of them were apprehensive about

the audio recording of their "thinking aloud". The sessions of familiarization of the subject with the thinking aloud task, in which they got acquainted with the procedures they would be expected to perform, allowed doubt clarifying by the researchers and helped to build the subjects' confidence in themselves and, most of all, in the researchers' promise of not revealing their identities. After having understood the objectives of the study and its relevance in terms of benefits for the area of research on documentary reading, the subjects were willing to collaborate.

By conducting these investigations, the group of researchers has tried to contradict the myth that a good indexer "is born with a gift for indexing", and that reading is performed intuitively without the need of any parameters which would guide such an activity. With the results, we can provide orientation for future indexers.

Some of the results of these researches that can bring benefits to the area of documentary reading are:

It was possible to observe meta-cognitive strategies during reading comprehension and the more proficient the reader is in grasping the content of a document, more strategies he uses.

As to the indexer's previous knowledge, verbal protocols allowed us to observe that even in cases when the indexer does not possess the knowledge of the specific area he serves, he is still able to understand the text and carry out the process of identification of concepts making use of skills and strategies of reading comprehension, relying on his knowledge of language and on textual aspects. Familiarity with the area can be achieved by experience. On the other hand, indexers must be trained in indexing processes which can make him conscious of the need to carry out conceptual analysis and also conscious of his innate cognitive skills as well as the constructed ones.

We understand that reading is a process that comprises different stages: analysis of the text structure and definition of text typology, searching for perceptual "cues", attempting to grasp the meaning of the text followed by concept identification, the final mapping to confirm the meaning and the result of reading: concept selection.

Reading is important for the process of conceptual analysis and the indexer, as a reader, needs to achieve reading comprehension to be able to properly represent the content of a document. The representation of the content of a document is a guarantee of the relevance of the retrieval process that is the objective of indexing.

Acknowledgements

References

- Brown, N. (1980) Meta-cognitive development and reading., In: *Theoretical issues in reading comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence, and education*, edited by Rand J. Spiro, Bertram C. Bruce, & William F. Brewer. (pp. 453-481) Hillsdale, NJ: L. Erlbaum Associates, 1980.

- Cavalcanti, M.C. (1983) Interpreta??o pragm?tica: princ?pios ret?ricos na intera??o leitor-texto em l?ngua estrangeira. In: *Anais do VIII Encontro Nacional de Linguistica*. Rio de Janeiro: Pontif?cia Universidade Cat?lica do Rio de Janeiro.?Departamento de Letras

- Cavalcanti, M.C. (1989) *Intera??o leitor texto*: *aspectos de intera??o pragm?tica*. Campinas: UNICAMP.

- Cavalcanti, M.C. & Zanotto, M.S. (1994) Introspection in applied linguistics: meta-research on verbal protocols. In: *Reflections on language learning*, edited by B. Scott. pp. 148-156. Clevedon: Multilingual Matters.

- Cintra, A.M.M. (1987) Estrat?gias de leitura em documenta??o. In: *An?lise document?ria: a an?lise da s?ntese*, edited by J.W. Smitt. pp.29-38. Bras?lia: IBICT.

- Cohen, A.D. & Hosenfeld, C. (1981) Some users of mentalistic data in second language research. *Language Learning*, **31**(2), 185-313.

- Endres-Niggemeyer, B. & Neugebauer, E. (1998) Professional summarizing: no cognitive simulation without observation. *Journal of American Society for Information Science*, **49**(6) 486-506.

- Ericsson, K.A. & Simon, H.A. (1980) Verbal reports as data. *Psychological Review*, **87**(3), 215-251

- Ericsson, K.A. & Simon, H.A. (1987) Verbal reports on thinking. In: C. Faerch & G. Kasper (eds.) *Introspection in second language research.* pp. 24-53. Clevedon: Multilingual Matters.

- Faerch, C. & Kasper, G. (1987) *From product to process.* In: C. Faerch & G. Kasper, editors. *Introspection in second language research.* pp.1-23. Clevedon: Multilingual Matters.

- Fagundes, S.A. (1997) *Leitura em an?lise document?ria de artigos de jornal.* Mar?lia: UNESP. (Scientific Report)

- Fujita, M.S.L. (1996) *A leitura em an?lise document?ria.* Mar?lia: UNESP, CNPq. (Research Report).

- Fujita, M.S.L. (1998) *A leitura em an?lise document?ria.* Mar?lia: UNESP, CNPq. (Research Report)

- Fujita, M.S.L. (1999a) *A leitura em an?lise document?ria.* Mar?lia: UNESP, CNPq. (Research Report)

- Fujita, M.S.L. (1999b) A leitura do indexador: estudo de observa??o. *Perspectivas em Ci?ncia da Informa??o,* **4**(1), 101-116.

- Fujita, M.S.L., Nardi, M.I.A. & Santos, S. (1998) A leitura em an?lise document?ria. *Transinforma??o,* **10**(3), 13-31.

- Giasson, J. (1993) *A compreens?o na leitura.* Lisboa: Asa.

- Gotoh, T. (1983) Cognitive structure in human indexing process. [In Japanese] *Library and Information Science*, (21) 209-226.

- Hosenfeld, C. (1977) A preliminary investigation of the reading strategies of successful and nonsuccessful second language learners. *System*, **5**, 110-123.

- Ingwersen, P., Johansen, T., & Timmermann, P. (1977) A study of the user/librarian negotiation process. In: *Eurim II: a European conference on the application of research in information services and libraries*, edited by W. E Batten. pp. 203-207. London: Aslib.

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

- International Organization for Standardization. (1985) *Documentation - documents, determining their subjects, and terms.* Geneva: International Organization for Standardization.

- Iwashita, K.H. (1999) *Estrat?gias de leitura para elabora??o de resumos.* Mar?lia: UNESP. (Scientific Report)

- Kato, M. (1987) *O aprendizado da leitura.* S?o Paulo: Martins Fontes.

- Lara, M.L.G. (1993) *A representa??o document?ria*: *em jogo a significa??o.* (Master of Communication Sciences dissertation, Escola de Comunica??es e Artes, Universidade de S?o Paulo, 1993)

- Leffa, V.J. (1996) *Aspectos da leitura.* Porto Alegre: Sagra.

- Nardi, M.I.A. (1993) *As express?es metaf?ricas na compreens?o de texto escrito em l?ngua estrangeira.* (Master Applied Linguistics to Languages Teaching dissertation, Pontif?cia Universidade Cat?lica, S?o Paulo, 1993)

- Paschoal, M.S.Z. (1988) Em busca da elucida??o do processo de compreens?o da met?fora. *Trabalho em Ling??stica Aplicada*, (12), 175-189.

- Pires, R.A. (1999) *Estrat?gias de leitura para indexa??o*: um estudo de observa??o dos processos de leitura com indexadores em Agronomia. Mar?lia: UNESP. (Scientific Report).

- Rubi, M.P. (1999) E*strat?gias de leitura para indexa??o em energia nuclear*. Mar?lia: UNESP. (Research Report)

- Santos, S. (1996) *A met?fora no t?tulo de um texto e a leitura para a an?lise document?ria*. Mar?lia: UNESP. (Scientific Report).

- Santos, S. (1998) *Leitura e estrat?gias de leitura para indexadores: aplica??o de protocolo verbal para observa??o*. Mar?lia: UNESP. (Research Report)

- Van Dijk, T. (1983) Discourse analysis: its development and application to the structure of news. *Journal of Communication*, **33** (2), 20-43.

Find other papers on this subject.

How to cite this paper:

Fujita, M.S.L.J., Nardi, M.I.A., and Fagundes, S.A.F. (2003) "Observing documentary reading by verbal protocol."*Information Research*, **8**(4), paper no. 155 [Available at: http://informationr.net/ir/8-4/paper155.html]

? the authors, 2003.

Check for citations, using Google Scholar

Appendix 1

Instructions on the "Thinking Aloud" procedure (adapted from Nardi, 1993)

What we are going to do now is an activity to familiarize you whith the technique of data collecting which will be used in our research.

All you have to do is to read the text that will be handed to you, in the same way you are used to do reading while you are indexing a text. It is very simple and natural.

During the entire reading task, you have to "think aloud". Try to imagine you are by yourself in a room, reading a text for indexing. In such situations, hasn't it ever occurred to you to think aloud by saying words, externalising your reasonings, your mental mechanisms in order to understand it? In such a process, one "thinks aloud" by verbalizing spontaneously and almost unconsciously his thoughts, questionings, searches for problems of comprehension he may have, his particular way of getting the meaning from a text.

One very clear example of thought externalising during performance of a task (that happens to most people) is to "think aloud" spontaneously while working out a mathematical problem. Could you have an idea of how that technique works? It corresponds to your verbalizing your internal speech, your thoughts.

Now, you may find some passages clear and easy to understand, others may require a "short stop" to think and efforts to understand? <u>All depends on your own way of doing it.</u>

You must remember that whenever you stop to think a little more or to solve any problem, you must try to externalise whatever comes into your mind.

If at any moment, you find difficult to speak and think at the same time, you can give an explanation on how you understood such a passage or how you found a solution for a comprehension problem.

If possible, try to make effort to "think aloud" <u>during</u> the reading process. <u>It is a unique process where speaking is thinking.</u>

Try to take no notice of the researcher. She will be present only to remind you that you ought to "think aloud" all the time and to control the audio recorder. Try to act naturally as much as you can, as if you were by yourself.

Try to concentrate only on the task you have to perform.

Appendix 2

Procedures before data collecting

Selection of the research-texts

Case study 1
<u>Selected text:</u>
BOSCO, A.F. et al. An?lise Cl?nica das ?reas doadoras de enxerto gengival livre. *Revista da APCD*, v. 50, n. 6, p. 515-521, nov/dec. 1996.

Case text selecting was done by SDO/USP following suggestions that the text should be one that had not been indexed by any of the indexers. The text chosen is an article on Periodonty and was published in <u>*Revista da Apcd*</u> with the title "Clinical analysis of free gingival grafting "donor areas".

<u>Consultation with a specialist:</u> in order to have a better understanding of the results, a specialist in Periodonty was consulted with the aim of getting his interpretation of the text content.

Case study 2
<u>Selected texts:</u>
LOB?O volta ao rock da melhor qualidade. **O Estado de S?o Paulo**, Caderno 2, 25 julho de 1997.

SEMIN?RIO prova que roupa tamb?m ? documento. **O Estado de S?o Paulo**, Caderno 2, 25 julho de 1997.

EXECU??O de quadrilha afasta delegado no MA. **O Estado de S?o Paulo**, Caderno Cidades, 25 julho de 1997.

ASSALTANTE do BB de Campinas ? preso. **O Estado de S?o Paulo**, Caderno Cidades, 25 julho de 1997.

The texts chosen were unfamiliar to the two indexers. Each indexer worked within a specific area, therefore, a different text was chosen for each, taking into account the area in which they worked.

Case study 3
<u>Selected text:</u>
FIQUEIRA, R.C.L. and CUNHA, I.I.L. A contamina??o dos oceanos por radionucl?deos antropog?nicos. *Quimica Nova*, v. 21, n. 1, p. 73-77, 1989.

CIN was asked to select a research text which had not yet been indexed, and with a theme that was not of too specific a sub-area, so it could be indexed by any professional whatever sub-area he was used to work with.

The selected text was a Chemistry article extracted from *Qu?mica Nova* with the title "Contamination of oceans by anthropogenic radionuclides".

Case study 4

Selected text:
COVEY, S. MERRILL, R. et al. Primeira coisa primeiro: si, voc? tem que aproveitar melhor seu tempo. Mas essencial ? aproveitar a vida. *Voc? S/A*. Agosto 1998. p. 96-107.

CIN was asked to choose a research text which had not yet been summarized by any of the subjects, with a theme that was not of too specific a sub-area, so the task could be performed by any professional whatever area he was used to work with. The text selected was an article on Management extracted from <u>*Voc? S/A*</u> with the title "First things first".

Case study 5
Selected text:
ALMEIDA, H. A.; KLAR, A. E.; VILLA NOVA, N. A. Compara??o de dados de evapotranspira??o de refer?ncia estimados por diferentes m?todos. *IRRIGA*, Botucatu, v. 4, n. 2, p. 104-119, 1999.

CENAGRI was asked to select a research text that had not been indexed. An article on Agriculture was chosen extracted from *IRRIGA* with the title "Comparison of data of evapotranspiration of reference estimated by different methods". The document follows the textual structure of a scientific article with Title, Summary, Keywords, Abstract, Introduction, Material and Methods, Results and Discussion, Conclusion and Bibliographical References. The article content also includes statistical tables, illustrations and graphs.

Selection of subjects

Case study 1

For the research purposes, four indexers were selected according to their years of experience in the system and in indexing activity and according to their "indexing skill". Skill was assessed by number of indexed records in the *Latin American and Caribbean Health Sciences* (LILACS) database and the *Brazilian Dental Biliography* (BBO), which was checked in "Indexing board LILACS and BBO up to December '96". All of the four selected indexers had degrees in Librarianship.

Case study 2

At the "O Estado de S. Paulo" newspaper, a journalist and a librarian were selected because they usually indexed more than the others, and had several years of practice in documentary reading.

Case study 3

Four indexers who were specialists in the area of Nuclear Energy were selected

Case study 4

The abstractors selected are the same as in Case Study 3. Therefore, the criterion for their selection was the same as in Case 3.

Case study 5

Four indexers with degrees in librarianship were selected according to their years of experience in the system and indexing activity, considering also their indexing ability.

The other procedures before data collection, such as the informal talk with each of the subjects and the session of familiarization of the subject with the "think aloud" task, were carried out in the same way in all case studies.

Procedures during the ongoing of the protocol recording

Audio recording of each subject's "thinking aloud" while the research text was read.

Procedures following the application of verbal protocols

Retrospective interview

Case studies 1, 3 and 4

As soon as each subject was over with the reading aloud task, a retrospective interview was conducted aiming at clarifying some points considered obscure by the research.

Case study 2 retrospective interview was not necessary.

Case study 5

Retrospective interview was only conducted with subjects 3 and 4 to clarify points that remained obscure during the "Think Aloud" procedure. There was no need of interviewing the other subjects once they had performed the task at ease and had furnished all the information necessary.