INTERDISCIPLINARY APPROACHES TO THE CONCEPT AND PRACTICE OF WRITTEN TEXT'S DOCUMENTARY CONTENT ANALYSIS (WTDCA)

Maria Pinto Molina
Department of Documentation
Universidad de Granada. 18071 GRANADA (Spain)

Content analysis, restricted within the limits of written textual documents (WTDCA), is a field which is greatly in need of extensive interdisciplinary research. This would clarify certain concepts, especially those concerned with "text", as a new central nucleus of semiotic research, and "content", or the informative power of text. The objective reality (syntax) of the written document should be, in the cognitive process that all content analysis entails, interpreted (semantically and pragmatically) in an intersubjective manner with regard to the context, the analyst's knowledge base and the documentary objectives. The contributions of semiolinguistics (textual), logic (formal) and psychology (cognitive) are fundamental to the conduct of these activities. The criteria used to validate the results obtained complete the necessary conceptual reference panorama.

INTRODUCTION

THE WELL-KNOWN PHENOMENON of documentary "inundation" has become the main cause for the growing presence of centres dealing with document processing. The basic objective of such establishments is the transformation of the original, or primary, documents into a more manageable format which in turn accurately represent the original. As stated by Lancaster [1, p.1], the main purpose of indexing and abstracting is to construct representations of published items in a form suitable for inclusion in some type of database. The importance of these derived, or secondary, documents, compared with the originals from which they are taken, is increased in said data bases, and this is obviously due to a greater ease of handling. It should be pointed out that the study and analysis of the formal aspects of the original document imply a first and necessary step in the elaboration of these derived documents, but their *content* is the real "hobbyhorse", not only by the greater difficulty entailed in its analysis and description, but also due to the documentary importance of the products derived from this analytical-synthetic process. And the fact is that the more documents there are in a specific centre, the greater the need of analyse their "content" in detriment of simply recording their formal properties.

Close to the terms "knowledge", "information" and "meaning", the word *content* express their superposition and integration into a higher unit of significance. Content is at the same time

knowledge, information an meaning. According to Hirsch [2, p.8], *meaning* is that which is represented by a text; it is what the author meant by his use of a particular sign sequence; it is what the signs represent. Significance, on the other hand, names a relationship between that meaning and a person, or a conception, or a situation. The content we are looking for (*documentary content*, we must not forget this qualifier) is nearest to that "significance". It's about a mental reality that rely on the knower. As Watson state [3, p.277], the conception of meaning as independent of the knower, which assumed that is was a property of objects or events was a result of the transfer to the social sciences of a viewpoint which dominated the physical sciences. Recent writers in the sociology of knowledge see the subjective and objective dimensions of experience as simply different ways of viewing the same world and both are essential for an understanding of any situation. On the other hand, as we will see, from the text's properties we can infer that textual content is many-sided and unlimited. In order to satisfy our analytical aims, it must be specified and limited. Written text's documentary content analysis (WTDCA) force us to determine what part of this many-sided and unlimited content will satisfy some documentary needs. And so the "content" may adopt different forms, from "subject" and "index" to "abstract". The "classification" is a consecuence or effect. Nevertheless, our main concern rest on the abstracting operations and its corresponding produce, the abstracts.

The term "content analysis" denote a family of research methods that attempts to identify and record the meaning of documents and other forms of communication systematically [4, p.251]. Being already restricted to the documentary field, and more concretely to the field of written text or bibliographic documents, *written text's documentary content analysis* (WTDCA) consists, on the whole, of the examination or breakdown that the written textual object must undergo in order to determine its content, and its subsequent description. What we really aim to obtain, beginning with "exploratory", or first hand, information, is the corresponding explanatory, or "representative" information (second hand). We are confronted by a severe problem from the outset with this type of analysis, which is none other than the vagueness and ambiguity that prevail as far as the term *content* is concerned, this being a problem that goes beyond linguistic borders. In Spanish language, it is extremely difficult to establish a clear definition of the terms "contenido" and "materia"; a similar thing occurs in English language when we refer to the words "content", "aboutness" and "subject". The fact is that, as Krippendorff [5, p. 10] maintains in a much wider sense than that which is purely documentary although it is applicable to our current interests, "content analysis trascends conventional notion of content as an object of concern and is intricately linked to more recent conceptions of symbolic phenomena. This may be seen in the context of a changed awareness about human communication, the existence of new media, and the roles they play in the transmission of information in society". In this paper we shall try to throw a little light on the dark epistemological panorama that surrounds the concept of "content" and, consequently, the type of analysis aimed and which, in any case, implies two inverse and complementary movements: *analysis*, or dismantling of textual structure in order to extract its informative content; and *synthesis*, or reconstruction of the content structures obtained in a new and reduced expressive manner that make the mechanisms of documentary searching easier. It

is a *cyclical process* developed between the material reality of forms and the spiritual fantasy of meanings, a unique and exciting "journey" from an apparent world of expressions to another, underlying world of contents, on the border between concrete and abstract, between psychological and logical, between instrumental and ontological. The problem arises with the extreme difficulty derived from the many mysteries that surround, on the one hand, the *complex textual mechanism* and, on the other, the *concept of content* as the starting point on a specifically documentary journey. A series of scientific disciplines decisively contribute in helping to solve this problem and this paper aims to discuss them by pointing out their basic contributions.

1. THE TEXTUAL OBJECT, A NEW INTERDISCIPLINARY PARADIGM

The *text* is the necessary starting point for analytical-documentary operations: not in vain does it possess a maximum capacity of response to the informative-communicative needs of the human species, since documentary communication between men is preferably carried out by means of these unique and complex information units. In fact, a documentary communication model (linguistic or otherwise) has not been discovered to surpass it. In a different line of investigation, though also interesting, Belkin and Robertson [6, p.199] state that the messages received from other humans (structured by them) are linguistic in the most general of senses; that is, semiotic structures constructed by one human (sender) with the intention of changing the image of another human or group of humans (receiver), which we call *texts*. The deliberate (purposeful) structuring of the message by the sender in order to affect the image structure of the recipient implies that the sender has knowledge of the recipient's structure.

  However, the textual object poses the preliminary problem of placing it in the ever more complicated universe of science, since it does not have the qualities required to belong to any of the already existing disciplines, and so we can intuit the advantages that giving it the category of new cross-disciplinary object would imply, it thus being the central symbol of a new scientific paradigm: *textology* or *text science*. In that line Derrida [7, pp. 13-124] speak of "*grammatologie*" as the project of a modern science: the writing science. The writing, instead of secondary notation system attached to language activities, is the content and essence of such activities. There is no linguistic sign before the writing, because without this outward appearance the very idea of sign fall into ruin. In this same line, Hjelmslev regret the least importance payed to the ink substance if compared with the air substance. But the "grammatologie" is destined to not receive its guiding concepts from other human sciences or traditional metaphysics. We believe that the introduction of this new interdisciplinary field would overcome the well-known but unsuccessful *transphraseological* perspective that conceives the text as a mere linguistic extension of the traditional sentence. We would tackle the text with a new phatic mentality in accordance with the growing *communicological* tendency, which puts the emphasis on the pragmatic dimension of language and the consequent consideration of communication in context: Language is a communicative activity parting from the philosophies of action; therefore, words are acts or

"deeds" that may have important consequences [8, p.42]. It should, thus, be sufficiently clear that it is not a question of dealing with the textual object by assigning it a new descriptive limit, or by simply starting a renovation process of the old and well-known linguistic object. On the contrary, we pursue its ascent to the category of central symbol of a new methodology in which its communicative-semiotic value takes on a more important role. But this new central symbol, by virtue of its newness and intrinsic complexity, is extremely difficult to define, although, for starting, we can allow certain definitions that conceive it as the verbal record of a communicative act [9, p. 6]; a discrete significative singularity; an autonomous unit of meaning; a process of significant updating [10, p. 41]; a sequence of linguistic symbols between two marked interruptions in communication; a complete unit of linguistic behaviour and, in short, a "complex signum", a "connected and complete entirety meeting a real or assumed communicative intention" [11, p. 191]. To Barthes [12, p. 108], there is a text in any place where it's carried out a significance activity according with some combination, transformation and displacement rules. The text is plural, radically symbolic and, as the language, it have not neither end nor center. After this variety of opinions the following dilemma arise: is a text really an *objectivization of meaning*, or is it rather simply and plainly an *object of perception?*. Difficult question whose round answer we imagine unattainable. Instead of implant an exclusive and radical situation, we opt for a complementary perspective that contemplate the presence of both elements (content and form) as essential part of texts, and neither of them can be slightly eliminated. An so this difficulty of apprehension of text identity entail the difficulty analysing its content.

From a documentary point of view, we admit that a text is a collection of *symbols which are intentionally structured* by a sender in order to change the structure of the receiver's image [6, p.201]. Though texts may also be spoken, we'll restrict our attention on the written ones, the overwhelming majority in the documentary scientific production, conscious that paralinguistc cues are denied to the writer and also that a printed version of a hand written text is, in an important sense, an interpretation [9, p.4-8]. It seems reasonable to suggest that, whereas in daily life in a literate culture, we use speech largely for the establishment and maintenance of human relationships (primarily *interactional* use), we use written language largely for the working out of and transference of information (primarily *transactional* use) (9, p.13). Anyway, written texts are complex and integrated symbols that play a mainly transference's (*transactional*) role and which require a systematic and in-depth study. The difficulty of a scientific description derive from the fact that text is a sign *open-ended* [13, p. 37]; it has a dynamic and open meaning (and so a dynamic and open structure), and this is something that the documentalist should think about in greater depth. The study of textual macrosymbol demands the interdisciplinary integration of the various semiotic variables of language (syntax, semantics and pragmatics), among which *pragmatics* stands out as the consequence of the phatic mentality which currently prevails in research into textual objects, although we should qualify that we find ourselves, at least for the moment, faced with a field that is not very systematized and is methodologically dispersed. In any case, the aim of every sensible text is to convey some kind of message, and this is done by marshalling the "meaning units" of the text in a rational and purposefull way [14, p.

3]. Therefore, it does not seem strange that the main definers of the textual unit are those of *cohesion* (formal approach) and *coherence* (functional approach) among its constituents, principles that guarantee the necessary semiotic integration. Basicaly, its general structure implies the superposing and inter-relating of two different and complementary structures: *formal* (expressive-superficial), and *content's* (functional-deep) [15]. Besides, the text may have an indetermined number of secondary structures, as the *rhetorical (discourse-level)* one, situated between both of these structures, it being a type of conventional production schema to which the text is adapted independently of its meaning and its expression. At this documentary moment, we'll reduce the *formal* structure to the physical reality of the text. On the contrary, the *content's* structure, situated on the other side of the same textual "money", refer us to its specific communicative (transactional-interactional) function.

Although our purpose is not to go too deeply into the essence of the rescued written textual object, we would like to reflect on some of its basic aspects: firstly, its *transactional nature* (knowledge's transference), which causes a situation of change, at least intentionally, of people's knowledge structure; its *cognitive nature* should also be pointed out, since the text is, above all, the cognitive relationship that is established with the text itself; on the other hand, and from a documentary perspective, we cannot study the textual phenomenon and its associated information if we do not keep in mind the phenomena that relate the text to the *sender* and the *receiver*. The *sender (originator-author-subject)* of the deeds of textual expression is one of the elements of its content. The coming into being of the notion of "author" constitutes the privileged moment of individualization in the history of ideas, knowledge, literature, philosophy, and the sciences [16, p. 101]. But the *author function* does not affect all discourses in the same way at all times and in all types of civilization; it is not defined by the spontaneous attribution of a discourse to its producer but rather by a series of specific and complex operations; it does not refer purely and simply to a real individual, since it can give rise simultaneously to several selves, to several subjects-positions that can be occupied by different classes of individuals [16, p.113]. Anyway, the author can no loger be regarded the only one determinig the meaning of a text [17, p.48].

Likewise, the *receiver (interpreter-analyst)* should also be kept in mind, since the sensorial data that is given has an infinite number of structures that will be limited only through the action of a receiver who plays a more active role than that which could be deduced from the name itself. In themselves, sense data present an infinity of structures, which are limited (or attain a singular structure) only through action on the recipient's part [6, p.200]. The "aesthetic of reception" took up the idea of assigning the meaning of a text to the reader and, at least in a certain formulation, radicalized it by shifting the meaning-production completely to the side of the reader [17, p.49]. Anyway, some special qualities must combine this receiver, because the process of communication may become distorted or restricted when those who are defined as appropiate persons to create and control information retrieval systems are not themselves actively engaged in the intelectual development of the ideas they are attempting to structure [3, p.272].The aporias the aesthetic of reception got entangled in are no reason to return to a theoretical model with the author as generator of meaning. Both conceptions have at least two things in common: they

both analyze works of art or text with regard to their implicit meaning and both regard the subject (either the reader or the author) as the instance of meaning production [17, p.49].

However, the communicative value of the textual units are modified substantially by the so-called *communication factors*, or *context of the data* or, in other words, by its contextual insertion, so that the action of the different types of context gives, removes and changes the meaning of the messages. In fact, the situational or contextual aspects correspond with the pragmatic perspective that we have established as the prevailing dimension in research into the text, in such a way that a textual theory implies a theory of context. Any content analysis must be performed relative to and justified in terms of the context of the data. Really "the vicarious nature of symbolic communications is what forces a receiver to make specific inferences from sensory data to portions of his empirical environment. This empirical environment is what we refer to as the *context of the data*" [5, p.23]. As can be seen in the graph (figure no. 1), the textual nucleus is surrounded by various membranes, or contextual skins, which represent the *scientific, linguistic and documentary contexts*, each one of which conditions the strategies to be followed on the road towards obtaining the *textual content*. Anyway, from a documentary point of view, it is as well to distinguish among de *production* context and the *reception* context. On the whole, we can say that the *scientific* context establishes a characteristic area of knowledge and sub-language; the *linguistic* context imposes a characteristic "straight jacket" to which all the symbols used in the text must yield; and the *documentary* context, without doubt one of the most important at this point in time, defines, among other extremes, the user category and his/her possible demands, along with the working conditions of the analysts.

Since written texts vary greatly in intention and style, and since the problems of language analysis are evident, we shall restrict our attention to a type of text that covers the greater part of scientific production: the scientific text. This has certainly differentiating qualities, such as its *standardised argumentational structure PMRC (purpose, methodology, results and conclusions)*, together with a particular and highly formal style; the use of a *scientific sub-language* compared to normal or everyday language; and above all the *priority given to what is implicit*, that already-known "old" information, accumulated throughout the centuries by humanity thanks to documentary tradition and, consequently, to science itself as a human creation. To this effect a determined scientific *perspective* entail a shared understandings which are often in part implied and not accesible to an external observer (3, p.281). The fact is that in scientific texts a large quantity of information is usually presupposed, this *presupposition* being an important factor to be taken into account when carrying out analytical tasks. It is true that the text is a symbol in which what is omitted does not mean a lack of, or insufficient, information but rather a significant reference mark foreseen in the conditions themselves of its semantic existence.

But the textual production as communication (transference) channel have a different perspective depending on wheter is dealing with *natural sciences* (NS) or *social sciences and humanities* (SS-H). Owing to the big amount of specialities in NS, the scientists often deal with narrows ways of knowledge, for which there is an extremely accurate writing style. The content is almost completely determined by the author [18, p. 358], and the receiver's role as interpreter

is markedly diminished. Nevertheless, in SS-H the authors often raise broad and complex questions that are analyzed from differents points of view. Obviously, this kind of text have a formal structure much more changeable depending on the author's personality and the interpretation's margins are larger. In contrast with the "hard" vocabularies in most sciences, the "softer" vocabularies of the social sciences [19, p. 169]. The rhetorical structure's analysis of scientific documents based on the PMRC division has confirmed that differences among a standardised and highly framed writing style in the *natural sciences* and the more idiosyncratic in the case of *social sciences and humanities.*

We can conclude this section stating that, because of the huge divergence among different kind of text, there is a need for serious and thorough research into text structures [14, p. 21]. It would contribute to stablish a text's classification according with the three structures aforementioned (formal, content's and rhetoric). This taxonomy would allow a better knowledge of text's essence, and so its more accurate analysis.

2. WTDCA, INTERDISCIPLINARY WORK

From the already explained, we can suppose that *written text's documentary content analysis* (WTDCA) is an intellectual job which requires an exhaustive delimitation of its investigative territory and for which a high level of scientific training is necessary. If we analyse the large quantity of written papers dealing with the field of *content analysis* (CA) we can see certain main trends: on the one hand, those that presuppose which type of "content" they refer to, which are a majority, and which obviously evolve, or at least force the reader to do so, in an undesirable conceptual nebula; on the other, those that, like Bardin [20, p.7], in a line that is related to the journalistic origins of content analysis, talk about a CA based on the probabilistic and quantitatives "intuitions" that the analyst proposes from the text; finally, we find a minority of meticulous researchers who, from documentary surroundings and with exclusively documentary aims, are aware of the scope of the problem that the delimitation of the concept "documentary content" represents. It is true that, as Beghtol [21, p. 84] states, not a single definition of the term *content* has been established, nor a theory on its role in information recuperation systems. Needless to say, it is a question of informative content, directly linked to the transforming capability of the text. The "content" that we are searching for is described as documentary, its aims are documentary, and it should be governed by a truly documentary quality, effectiveness, and more precisely, documentary effectiveness. We do not mean that the essential stage of textual inference-interpretation carried out by the analyst should disappear, but rather that this process should be strictly submitted to the documentary imperatives of the moment. With the *ideal of documentary effectiveness* established as the basic starting point, we should know what content we are referring to, although the difficulty here is great by virtue of the vagueness and ambiguity of the term itself.

Using the dictionaries available, we can differentiate up to six different types depending on the point of view we adopt (*general, logical, grammatical, philosophical, psychological, or*

documentary), which will represent, as we shall see, a compendium of the types previously pointed out. Otlet differentiated three types of content: *intellectual* (truth, knowledge, science), *affective* (beauty, feeling, art) and *intentional* (good, action, morality). In any case, the epistemological points raised with regard to the concept of "content" vary between two universal trends: *subjectivism* (empiricism, pragmatism, idealism) based on the idea that sensorial perception is a prior and unavoidable step to knowledge, and *objectivism* (rationalism), which places any form of knowledge before sensorial perception. Our *documentary* (realistic) position, is based, as we shall see, upon the superposition or integration of both theories. In trying to overcome a theoretical conflict that shows all the signs of lasting for many years, and with an eminently documentary or, in other words, pragmatic spirit, we shall accept that "content" has two main components, one *theoretical* and the other *practical*, *ontological* and *instrumental*, *intrinsic* and *extrinsic*, *semantic* and *pragmatic*.

The term "content", however, is closely linked to the term "meaning" and our approaching it will allow us to confirm the aforementioned *conceptual duality*. According to Lyons [22, p. 15], the topic of *meaning* is of concern to so many of the social sciences and does not fall wholly within any single one of them. Of all the disciplines with a interest in meaning, *linguistic* -the scientific study of language- is perhaps the one to which it is of greatest concern. Meaningfulness is essential to languages as we know them; and it is arguable that the very notion of language without meaning is logically incoherent. Moreover, although many kinds of behaviour can be described as meningfull, the range, diversity and complexity of meaning expressed in language is unmatched in any other human or non-human communicative bahaviour. Just as occurs with "content", "meaning", one of the most controversial terms in our language, is subject to a large number of theories. Acording with Watson [3, p.278] we can't forget the social condition of any kind of "meaning", because "new knowledge, which start as personal experience, becomes objectified by framing it in a particular perspective and validating it according to the rules of that perspective.... The meaning of the terms in a document are mediated by the theoretical *perspective* in which they are embedded". In our case, keeping in mind the documentary objectives we are concerned with, we shall limit it to two main components: *transactional* (semantic, cognitive, conceptual, logical, referential), an integral part of the essential workings of language, according to which the meaning of an expression is what it refers to or represents; and *interactional* (pragmatic, expressive), more in the line of stylistics or pragmatics, closely related to the sender's communicative intention. Both are complementary and allow us an overall view of said communicative intention, since we should not forget the unanimity that exists among philosophers, psychologists and linguists when recognising an intrinsic relationship between meaning and communication.

The *complementarity* of the two components to which we have limited the complex and polemic term "content" is a reality which has not been insisted upon enough, at least in documentary circles. We should like to do so through this paper, in pursuit of a much-needed sorting out of ideas. In short, everything that may be explained with regard to this *basic* concept will be the obligatory starting point for subsequent research in the sector. Once accepted the such a

complementarity between the two basic components of "textual content" (theoretical-practical, transactional-interactional, semantic-pragmatic, ontological-instrumental, intrinsic-extrinsic), and taking for granted a certain informative (transforming) capacity in both elements, we must recognise, however, the situation of unsteady balance among both components depending on the type of text. Along similar lines of research, and taking into account the closeness between "content" and "subject", Hjorland [23, pp. 181-185] identifies "subject" with *epistemological potential*. But a potential is an objective possibility which is determined by the level of development of society or, in other words, by the social context, and the person who is carrying out the job of analysing obviously forms part of this context with very specific qualifications and interests. In this way, *inference* or *interpretation* is assumed and justified as one of the basic ingredients of informative content analysis of documents: its grade or level will depend on the interests of the users. In fact, determining the content implies the evaluating and attributing of priorities to the properties of the document with a view to its description. "Content" is not an aprioristic function of the properties of documents, but rather the whole context in which such content is described, the *context of documentary description*, is what determines this function. Even though pragmatic subject theory has its limitations, it makes an important contribution to perception of central properties of the concept of the subject by pointing out its means-goal nature (and thus repudiating the view of subject as "inherent qualities"; subjects are no more inherent qualities than is the *value* of a thing). A subject thus is always a subject for someone or for something [24, p.181]. In fact, the empiric interest towards symbolic facts can no longer be applied to the study of the messages in an isolated manner, nor can it limit communication to a psychological process or consider the linguistic interpretations of a message as the basis for an explanation. And so "the changed fabric of society calls for a *structural* definition of content, one that can take note of channels and constraints on information flows, communication processes and their functions and effects in society, and systems involving advanced technology and modern social institutions" [5, p.10].

In any case, the problem of *"content"* affects all linguistic units, and becomes more complicated as the units themselves become more complex. But the comprehension of complicated mechanisms demands first of all the comprehension of its primitive units. According to Salton [24, p. 379], the first step in any language analysis system has to be the recognition and identification of the individual words in the text. This is a question of minimal units which, on a formal basis and in a traditional manner, are endowed with said meaning: we find ourselves at an essential first level of *phonological or elementary analysis* aimed at understanding the corresponding meanings of the different words, although, here, the problem is incredibly complex and can be summed up by affirming that the correspondence between "forms" and "meanings" of the different language units is not sufficiently accurate. *Vocabulary control*, the definition of *grammatical types*, and *lexical fields* help to ease this apparently unsolvable difficulty. After this severe defect at the outset, we are faced with the *sentence* in the next stratum up, the meaning of which not only depends on the meaning of the words that make it up, but also on its grammatical structure, this being independent of the context and dependent on the characteristic

use of the type of part of speech it belongs to due to said structure. This is a second analytical stage, *structural analysis*, in which we pursue the transformation of the text into a structural representation that aids the design of the inferences with regard to the intention of said text [25, p. 89]. However, in *enunciations*, apart from the verbal meaning, there is a non-verbal meaning (prosodic and paralinguistic) of great importance and difficult to evaluate. Having arrived at this point, and given that language appears in both oral and written forms, a problem of transferences arises, and almost all written enunciations correspond to significantly different spoken enunciations.

At the apex of our analytical periplus we find textual analysis, directly aimed at obtaining *textual content*: a text is a *discrete significant singularity in which a system of significations is underlying*, the units of which, discrete also (words or enunciations) can be developed as a process. This development is usually called *updating of meaning* [10, p.45]. Supported by previous analyses (*elementary*-phonological, and *structural*-syntactic-semantic) we are better placed to carry out the analysis of textual content as such. In order to do this, and keeping in mind the condition of semiotic integrality associated with the textual unit, we shall commence with an approach which is also integral. This means that we must keep in mind the two types of content (meanings) that, as we have already mentioned, are complementary: *semantic* (transactional) content and *pragmatic* (interactional) content.

Semantic content refers to the *transactional-conceptual-cognitive-logical-referential meaning* of the text. But the content of the textual material is also related to the phenomenological mental content [26, p. 99]. Consequently, and having taken phenomenology to be the science of intentionality, we shall identify the pragmatic aspect of textual content with its *interactional-expressive-intentional meaning*. In any case, the fundamental problem of WTDCA consists of extracting both deep structures (*semantic and pragmatic*) from a specific surface structure.

After these brief but essential details with regard to textual *content*, identified with *both aspects of the textual meaning (transactional and interactional, conceptual and intentional, semantic and pragmatic)*, we shall be better placed to carry out the study of the analytic-synthetic textual processes. It is now time we define WTDCA in a strictly scientific manner. According to Krippendorff [5, p.21], content analysis "is a research technique for making replicable and valid inferences from data to their context". But inference, or inductive reasoning, is a probabilistic method and obviously can never give absolute certainties. Restricted to our documentary field, written text's documentary content analysis (WTDCA) is a *research technique for the intersubjective (verifiable and reliable) and systematic description of the content structure of texts*. As is shown in the graph that we have produced (Figure 2), it is basically a *cognitive inductive process* (bottom-up) *of controlled omission, although deductive insertion* (top-down) *is essential* to textual comprehension of a specific *base knowledge plot* which is extralinguistic and conventional (21, p.92). It thus occurs thanks to the intercrossing of flows derived from *textual analysis* as such and from the *base knowledge* of the analyst. The documentary interests and *base knowledge* of the analyst allow the determination of the context (*content is sensitive to context*) which is necessary in order to carry out the intellectual operations of inference-

interpretation which are essential in any analytical process. As can be deduced we are faced with a unique case of *cognitive transition* between two totally different worlds. The depth of the problem is such that its solution in terms of absolute accuracy would imply using a "magic" touch and, consequently, we consider it to be virtually unreachable.

However we can come close to this ideal of accuracy that we have expressed as utopian. In order to do this we must resort to the best resources that current science can offer us, among which we find the three basic pillars that support WTDCA: *cognitive psychology*, since we are faced with a mainly cognitive process; *logic*, since we must rationalise all the operations involved; and *linguistics*, for obvious reasons. Given that the superposition and integration of knowledge from these three fields is a constant in the development of WTDCA operations, any attempt at internal classification would seem rather absurd. However, in purely pedagogic terms, we shall establish a chronological sequence for said analytical-synthetic-textual phenomenon, differentiating three basic *stages* which are developed consecutively (Figure 3):

1) *reading and comprehension*, in which the contributions of cognitive psychology are of prime importance; 2) *inference-interpretation*, in which at least all three of the aforementioned disciplines come together. But this interpretative moment has little to do with that which occurred during the stage of textual comprehension, since on that occasion the documentary participation was minimal. Now, on the other hand, is the moment to develop this whole *informative capacity* that we have associated with documentary "content", adapting it to the documentary interests of the moment. Said development will be mainly conditioned by a series of documentary factors (documentary context), among which we shall point out the production conditions, the analyst's qualities and the demands of the system users. Needless to say, we are faced with the most documentary and creative phase of the whole process. It is, without a doubt, a delicate and difficult step, since if, up to this point, we have been able to submit the analytical activities to techniques that are more or less successful, it is practically impossible to establish inductive-interpretative mechanisms that are valid for all types of documents and demands; 3) finally, the *synthesis describes the content interpreted* in accordance with the documentary demands. This is the creative moment for the analyst, in which he/she must "expand" the content structure with regard to certain documentary objectives.

The areas that affect textual singularity most (*Cognitive Psychology, Logic and Linguistics*) are neither watertight nor independent, there being a noticeable mutual influence, thanks to which the advances recently experienced in linguistics and psychology may mean that between the two of them, along with logic, a supply cycle is established that can but benefit the three aforementioned sciences: any operation that is carried out on the text must be studied from, at least, the three points of view, *cognitive, logical and linguistic.*

3. COGNITIVE PSYCHOLOGY

The scientific study of human conduct, or *psychology*, is of interest to the WTDCA processes. In our century the psychology had fluctuated between two main tendencies: the paradigm of

behaviourism, according to which the human behaviour can be explained exclusively by means of physical stimulations and motive answers; and the *cognitivism*, that conceive the human organism as capable of produce information in terms that go beyond the purely physic to get superior levels. This *cognitive psychology* is a relatively new area that deals with the *mental processes and structures* implied in the acquisition, processing and use of knowledge or information, among which we should differentiate between the *basic mental processes* (memory and attention), the *mental representations* (imagination, formulation of propositions and establishment of categories), and the *complex mental processes* (comprehension, reasoning and problem solving), since almost all of them are involved in the content analysis we are concerned with. According to Cremmins [27, p. 28], information processors use their cognitive skill by adding value and meaning to the representational information they process or produce. This skill is based on the following mental activities: thinking, learning, communicating (reading, writing, listening, talking) and problem solving or decision taking (reasoning, judging, deducing, inducing, extrapolating, analysing and synthesizing). On the whole, mental phenomena are characterized by being *inaccessible to public information, extremely rapid*, producing a deceptive impression of simplicity, and *interactive*. Cognitive psychology is an empiric science based on the reality of Man as a communicative being. One of his variant is that of the information processing, considering the human mind as a system that processes information [28, p. 8], and in which the same language of the computer systems is used: that of *information processing*, by virtue of the similarity between computers and the human mind, a similarity which is not physical but, rather, functional, and which has been exploited in both directions, carrying out the function of metapostulate for cognitive psychology [29, p. 33]. The functional aspects of information processing, and not the physical ones are, therefore, what interest this new and promising discipline.

Although we are conscious of the disparity and complexity of psychological-cognitive problems, we shall centre our attention on some of them, beginning with two closely related phenomena: *reading* and *comprehension*. Reading, the only possible way of gaining access to the content of textual documents, is a concurrent process, and not simply a symmetrical one with regard to writing, of an interactive nature, which depends both on the text and the reader, consisting of a series of coordinated procedures that include perceptual, linguistic and conceptual operations. It is developed by means of the continuous and simultaneous application of two types of inverse and complementary information processing: *ascending, data lead, inductive, bottom-up*, in which the reading is linear from the parts to the whole of the text; and *descending, conceptually orientated, deductive, top-down*, in which we proceed inversely, taking advantage of the reader's *base knowledge* [30, p. 141]. This is a double action (visual reception-perception, comprehension) involving the frequencies of *vision, memory* and *reason*, the strategies of which depend not only on the reader and the text but also on the documentary objectives, and which can be *cognitive*, comprising interpretative behaviour which is automatic and unconscious, and *metacognitive*, or disautomated. Good reading, far from being a spontaneous action, should be organised and follow a method. The expert reader does not perform a linear reading, but rather

knows how to read past the implicit and make predictions about the following enunciations, confirming that prediction and reading are two inseparable realities.

As we have already mentioned, reading needs the use of memory, or Man's ability to recognise and remember. We can distinguish between *short-term memory*, which accumulates short-lived information with a superficial structure; and *long-term memory*, the aim of which are semantic structures, also known as semantic or conceptual memory. In most theories of semantic memory, concepts, not individual instances, are the basic content elements [31, p.12]. This is much more easily recognised if we identify it with *prior knowledge* or *base knowledge*, a type of personalized intellectual heritage based on outlines (frames), stereotyped conceptual structures which are present in everyone and which, as we shall see, play an active role in the analytic-synthetic process that concerns us.

The falseness of the traditional approaches to *comprehension*, based on the fact that it occurred automatically when one knew how to structurally process a text, has given way to new theories that define *textual comprehension* as the *creative process of a mental model with which to interpret described facts*. Comprehension processes are generally assumed to combine information from two sources: explicit statements from the text being read and general knowledge already known to the reader. Interactions of information from these two sources produce the representation of a text that is encoded into memory, ..., this interactions lead automatically to the encoding of minimalist inferences from which more goal-directed, purposeful inferences are constructed [32, p. 440]. The following ideas must to be highlighted: to understand is to integrate and interpret, to create meaning; in comprehension, the ascending and descending processes are involved; the depth of textual processing increases comprehension and the perspective adopted by the reader noticeably conditions it. In short, *comprehension* makes the *meanings* emerge after receiving the graphic symbols. It is known that it is not linear, since we conceptualize segments of discourse that are constantly re-shaped by the conceptualization of the following segments. Semantic is incessantly transformed into conceptual, and pragmatic into intentional. Van Dijk [33, p. 178] bases the comprehension of linguistic enunciations on four principles: *segmentation* of the signals from the constant flow of language; *categorization*, a process which refers to the syntactic categories of words, in the paradigmatic aspect of language; *combination*, because said categories are juxtaposed, generating syntagmatic structures; and *interpretation*, which is only possible as a consequence of prior mental operations, among which we can highlight *presupposition*, or the establishment of hypotheses on the text itself. The comprehension of the sequences of the parts of speech must possess a type of cyclical nature, corresponding to the *cyclical principle of the textual elaboration* of information [34, p.194], which joins old (in other words, already-known) and new information together by overlapping the different cycles. In any case it should remain clear that comprehension is a creative and constructive phenomenon.

Needless to say, the *inference-interpretation* stage represents the most subjective, rather intersubjective, moment in the whole of WTDCA, since certain extratextual (documentary) factors are involved, among which it is worthwhile pointing out the *base knowledge* of the

13

analyst, the *context*, in the broadest sense of the word, and the *documentary objectives*. Charaudeau assumes that the linguistic act lies in an asymmetry between the *production* and *interpretation* processes, there being two subjects with different competences (the communicative and the interpretative), implying a *semiolinguistic competence* in which we can differentiate the *linguistic* component, made up of the different organisational orders of language in its various conceptual elements; the *situational* component, made up of the sociolinguistic situations representative of the different social practices; and the *discursive* component, which stems from the combination of the different conceptual elements, giving rise to multiple discursive effects, in accordance with the communicative intentions [34, p. 39]. In a deeper point of view Petöfi [11, pp. 194-195] say that "one usually investigates the *system-immanent construction* of a text and/or the *functional setting* of a text. We cal this investigation "interpretation". Both can be investigated as a *static* or as a *dynamic* entity. In the first case we speak about *structural* interpretation, in he second case about *procedural* interpretation. We think that the two kind of interpretation (structural/static and procedural/dynamic) are necessary in the process of WTDCA. This double personality of text as static-dynamic reality must be pointed out.

4. GENERAL AND FORMAL LOGIC

Compared with psychology, logic represents the other side of the coin, since if the former is an eminently empiric science (*science of real thought*), the latter is an exclusively formal science (*science of abstract thought*). If one of them is necessary in order to meticulously carry out WTDCA operations, likewise the other for similar reasons. Logic and psychology proceed separately but in close collaboration. It is a fact that human knowledge materialized in scientific texts is a "world" of logical facts (universal concepts) arranged in a logical structure [26, p.97]. If we add the contrastable reality that any analytical process is, above all, a logical process, then any doubts about the possible contributions of logic to WTDCA are fully dissipated.

 Language, a mere go-between in our communicative intentions, creates certain severe shortcomings which are unacceptable from a logical point of view: *homonymy* (various things called by the same name); *synonymy* (various names for one thing); *extensional indetermination*; and *indistinction among levels*. In trying to solve these obstacles, we traditionally fall back on *logic*, a discipline that studies the structure, fundament and use of cognitive expressions, allowing, in short, a meticulous analysis of thought. The *reasoning* on which this important branch of philosophy is based is a mental operation by means of which new proposals are generated from proposals that already exist, a device that allows the organisation and widening of the different levels of human knowledge, especially scientific knowledge. Its help is essential in textual analysis processes, above all on a structural, or syntactic-semantic, level, for which two types of reasoning are used: *deductive* (Aristotelian syllogism), which is the most used due to its complete reliability, and which allows conclusions to be made from the textual data (conclusion and premises are necessarily and inexorably linked); and *inductive*, or

probabilistic *inference*, the reliability of which is relative, since it is based on the probability that something is true from that same textual data, there being a risk factor in its use (conclusion and premises are probably linked). In any case, *rationality*, understood to be the intellectual ability to formulate and solve problems [35, p. 238], is a fundamental quality in WTDCA processes which occurs and develops by means of the following intellectual operations: *classifying*, or ordering in variable degrees of relationship; *clarifying*, avoiding confused arrangements, vagueness, or lack of accuracy (very common in ordinary language) and ambiguity, a phenomenon which occurs when an expression has various meanings or when a grammatical structure is incorrect; *inducing*, or *inferring*, the free and creative need in any reasoning process; *establishing analogies*, since, either consciously or subconsciously, we entrust a large part of our thought to them; *weighing up evidence*, a factor that differentiates more than any other what we could call sensitive people from those that are not; and *deducing*, a talent that preferentially stems from our general sense of language. In order to reason well certain qualities are necessary, such as the command of language, the sense of what is pertinent, worldly knowledge acquired through experience and skill in exercising logic (formal), the command of which has the same fortifying effect on our conduct in informal reasoning as gymnastics has on conduct in sports.

Formal logic, so called by the fact that it deals exclusively with the *form* (significant) of enunciations, is a variation of general logic which has acquired particular importance in recent decades. As the name itself indicates, it limits itself to a system of symbols, relating the exercise of logical reasoning to an algebraic calculation. The importance of formal logic applied to operations of content analysis comes from the fact that logical symbols, unlike linguistic ones, have a perfectly accurate meaning. One of the most important discoveries of contemporary methodology is having realised that, using language in its *syntactic* plane (and therefore disregarding the other two) makes the intellectual work much easier. This method can be considered as forming part of the field of what we understand as *formalism* (conceptography), and consists of leaving aside the meaning of the symbols used and considering them exclusively as graphic symbols; it is a *formalised language* (metalanguage), an extension of a very old method, the *calculation method*. It is a fact that in ordinary language we come across expressions of a logical-formal nature. What the logical-formal does is to extract, isolate and integrate them into a calculation structure which will consist of three basic ingredients: 1) a collection of primitive elements, or *elementary symbols*; 2) a collection of *rules (formation or construction)* which establish the possible correct combinations of those elementary symbols; 3) and a collection of *transformation rules*. The calculations are deductive and autarchic, their importance lying in their exclusively formal and purely syntactic nature. They do not form a language, a means of communication, but rather a truly syntactic framework: their elements are opaque entities, though there is always the possibility of transforming a calculation into a language by interpreting its symbols and giving them a meaning [36, p. 31]. Let us finally say that, thanks to logic, linguists have been able to undertake the task of elaborating a generative-transformative theory of language. Present linguists believe that, due specifically to its complexity, ordinary language demands a formal logic, a formal reconstruction that is more and

more refined. Needless to say, *formal logic* is a science with great possibilities of application with regard to documentary content analysis, since it gives the possibility of operating logically with different units of signification (functors and arguments) through the unique device of assigning them a specific syntactic category. The transformations that stem from these *logical operations* with *semantic-pragmatic* units will guarantee the integrity of the text as a unit of content and, furthermore, will allow an accurate and meticulous logical-semantic-pragmatic approach to the analytic problem previously unheard of in the linguistic-documentary sphere [[37], pp. 113-116].

5. TEXTUAL SEMIOLINGUISTICS

Gardin wonder if the linguistic has played any significant part in the development of the intellectual tools used by documentalists [[38], p. 141]. The answer is emphatically yes. From Saussure (father of *structuralism*) to Chomsky (leader of *transformational* grammar), different linguistic schools having contributed to develop documentary methodologies. The presence of linguistic in the documentary field is unending.

   The *word*, the smaller linguistic entity capable of forming concepts, is the compulsory starting point: the greatest theoretical contribution to the understanding of said significant cell or elementary linguistic symbol has come from European structural-functional linguistics and its theories on *lexic fields*, begun by Trier [[39], p. 119], based on the fact that "articulation is the most general and profound essential characteristic of any language". According to Weisgerber [[40]], "the most important aspect of the field idea is that it has become the central methodological concept of applied research into linguistic content". Along this same line of research into content analysis we find Coseriu, Greimas and Pottier: they begin with the principle of linguistic economy, promoted by Hjelmslev through his "figures" (the construction of an unlimited number of linguistic symbols by means of a limited number of non-symbols, called "figures"). The number of these "figures", or *semantic primitives* (atoms-indivisible), is limited, but by linking said primitives we can build a potentially infinite number of symbols. Transferred to the next analytical step up, it would be unfair not to recognise that we owe *transformative-generative* linguistics a much more definite approach in order to reach a true knowledge of the sentence. Let us remember that transformational analysis begins, in the syntactic plane, with the so-called *grammar of sentence structure*, as the way to studying the transformations from one type of speech into another, and starting from certain *components* and their *composition rules*, generates sentences by proposing two levels, *deep (logical) structure* (DS), latent in the *competence*, and *surface (logical-psychological) structure* (SS), patent in the *action*. Every sentence has this double configuration, and on the whole, although there are nuances, we can affirm that *the deep structure generates the surface structure*, and that the complicated processes of *transformation* are underlying in both of them. In any case, the component that has a generative capacity is the *syntactic* one (the other two are interpretative components), made up of the *basis*, or set of rules that generate deep structures, and the *transformations*, or rules that turn the deep structures into surface structures. It should be pointed out that *the*

transformations are contextual rules, since there are certain restrictions or conditions imposed by the context in their formulation. In any case, the meaning of a sentence depends on its syntactic structure and the specific meanings of its elements.

The attractive category of the *double structure*, which is found in the basis of this new grammar, can be effectively applied to texts, the more extensive and complicated the better. All textual theory is based on the structural dichotomy (deep-surface). The *text* reveals, as does any linguistic entity, this double source of statement: as a *structure with a terminal-linear expression*, a *surface* structure, which we previously referred to as *formal* structure; and as a *deep* structure, or *content's* structure. The most operative type of textual analysis, at least for the present, is that which is *ruled by syntaxis*, the degree of universality of which is high. However, any syntactic analysis will be influenced by certain unavoidable forces: *context, semantics and pragmatics*. This is due to the fact that textual vagueness and ambiguity cannot be eliminated by purely structural considerations. We will, therefore, be forced to study entirely the syntactic, semantic, pragmatic and contextual phenomena. *Textual linguistics* is based on the well-known *principle of linguistic isomorphy*, according to which the organisation of our human communicative-verbal products is carried out by means of a process of quantitative expansion which, however, strictly respects the structure of the elementary linguistic cell. With regard to our analytical aims we are faced with a transcendental discovery: our methodological interest in discovering the *content* structure by virtue of an elementary principle of textual economy or entropy comes from the fact that both textual structures, the difference in size or extension of which is considerable, are *documentary equivalent*. The *transformation rules* that we shall use in this reducing period, very similar to the rules of sentence genesis in transformative grammar, are due to the aforementioned principle of *isomorphy*, allowing us the necessary informative *filtration*. Said transformations will affect the *morpho-phonological*, or expressive, *logical-semantic*, or conceptual, and *actuative*, or pragmatic, structural strata, leading to a global *content* structure, the obligatory starting point in the subsequent interpretation process.

6. CONCLUSIONS

A series of conclusions arise from what has been said up to this point which affect, above all, the *text* and its new semiotic conception, and the interdisciplinary contributions that favour the analysis of its *content*. Our aim is to encourage the analyst to reflect and become aware of the scientific importance of the textual object as the raw material in the complex analytical processes, and above all of the need to establish a *methodological territory* for these processes. As can be seen in the graph (Figure 4), there are five basic ingredients (necessary and sufficient) used to meticulously carry out textual content analysis: text, context, analyst's base knowledge, documentary objectives and a results validation method. Any investigative process begun should bear in mind each and every one of these factors, about which we should like to highlight certain ideas:

1)   The *text*, a new paradigm in textology, does not lend itself to naïve or simplistic

approaches, needing an appropriate systematization of its various structures (especially formal, content's and rhetoric). The scientific text serve a transactional (transference) function.

2) Each kind of text (according with an inexisting, by the moment, classification) require a kind of analysis. To point up the clear differences between NS (natural sciences) and SS-H (social sciences and humanities) texts.

3) The extent of textual content is reduced thanks to the concepts of perspective. Taking into account that each text entail a determined *perspective*, the concepts of "scripts", or causal chains representing stereotypic actions in a given setting and, in a superior stage, (MOP) "Memory Organization Packets", representations of knowledge common to many different situations [19, p. 163], decisively contribute to simplify the WTDCA's process. Presuppositions are essentially tied to that perspective.

4) The *context* has a similar analytical category, the importance of which and the need to define it beforehand have still not been expressed explicitly enough. The content is determined greatly by the context of documentary description. The analyst must to know not only the *production* context but the *reception/interpretation* context. Both are conditioning the analytical results.

5) A WTDCA is unviable without the contributions of certain disciplines which are fully established on the scientific horizon, among which we should point out *linguistics* (textual), *logic* (formal), and *psychology* (cognitive). There is a preferential link between psychology and textual surface structure and between logic and textual deep structure, as two sides of the same coin. The relationships between the three disciplines considered are bidirectional, with numerous mutual contributions.

6) Cognitive psychology, and specifically *information prosessing psychology*, open good prospects to WTDCA. Logic, and more exactly *formal logic*, allow the more than necessary sintactical analysis of texts. *Textual linguistic*, founded on transformative grammar, contribute decisively to the dificult transition between form and content. In whichever of this scientific sections there is room for a specific research applied to WTDCA.

7) Nor can we ignore the creative competence of the analyst, whose level of basic knowledge (general and specialised) is fundamental. Effective retrieval systems depend on an active and on-going partnership between information scientists and subject experts (12, p.281). Precisely for this reason we have mentioned *base knowledge* as one of the active ingredients in content analysis processes. As Allen and Reser state [4, p.259], there is substantial room for improvement in the quality of the content analysis being done in library and information science, and it is to be regretted that more researchers who use content analysis do not use more than une coder. It would be more than advisable the participation of several (at least two) analyst in the preparation of each analytical product.

8) The *documentary objectives* also constitute a key element when obtaining a content

that, let us not forget, should be documentary, in other words, endowed with the appropriate informative capacity in accordance with the previously established demands of the user.

9)   Conditioned by these four basic factors (text, context, documentary objectives and one's own prior knowledge), the content analyst is forced to carry out an extremely difficult job such as the *inference*, or *interpretation*, of the text, keeping in mind the rest of the determinants mentioned and the importance of which on this occasion is comparable to the importance of the text itself.

10)   Needless to say, a methodology on written text's documentary content analysis (WTDCA) would not be complete if we did not previously establish and define the system of *validation of the results* obtained, which is the only way of also validating the analytical process itself. However, this last point will be the subject of another publication, the research into which we have already begun.

Lastly we must tell that the purpose of this paper is not to dogmatize or present definitive answers -which, of course, we have not- but to contribute with our modest opinion to that necessary reflection, starting point on the problem of WTDCA.

REFERENCES

1.   LANCASTER, F.W. *Indexing and abstracting in theory and practice*. London: Library Association, 1991.

2.   HIRSCH, E.D. *Validity in interpretation.* New Haven/London: Yale University Pens, 1967.

3.   WATSON, L.E., et al. Sociology and Information Science. *Journal of Librarianship*, 5 (4), 1973, 270-283.

4.   ALLEN, B., RESER, D.: Content analysis in library and information science research. *LISR*, 12, 1990, 251-262.

5.   KRIPPENDORFF, K. *Content analysis. An introduction to its methodology.* Beverly Hills: Sage publication, 1985.

6.   BELKIN, N., ROBERTSON, S. Information Science and the phenomenon of information. *Journal of the American Society for Information Science*, July-August, 1976, 197-204.

7.   DERRIDA, J. *De la grammatologie*. Paris: Les editions de minut, 1967. English edition: Of Grammatology. Baltimore: Johns Hopkins University Press, 1987.

8.   NOEL, J. Sintax, semantics and pragmatics in the automatic analysis of texts. *In: New trends in documentation and information*. London: Aslib/FID, 1980.

9. BROWN,G., YULE,G. *Discourse analysis*. Cambridge: Cambridge University Press, 1983.

10. NUÑEZ LADEVEZE, L. *La construcción del texto*. Madrid: EUDEMA, 1991.

11. PETÖFI, J.S., OLIVI, T. Understanding literary texts. A semiotic textological approach. In: MEUTSCH, D., VIEHOLF, R. *Comprehension of literary discourse*. Berlin, De Gruyter,   1989.

12. BARTHES, R. *El susurro del lenguaje. Mas allá de la palabra y de la escritura*. Paidos, Barcelona, 1987. English Edition: The Rustle of language. Oxford: Basil Blackewell,   1986

13. HERRERO BLANCO, A.L. *Signo/Texto (De gramática a retórica).* Alicante: Universidad, 1986. Unpublished dissertation.

14. PAICE, C.D.   The rhetorical structure of expository text. *In:* JONES, K. ed. *Proceeding of 11th Informatics Conference*. London: Aslib, 1-25.

15. We do not wholly agree with the terms "microstructure", "macrostructure" and "superstructure", created by Van Dijk, because the prefix "micro", "macro" and "super", at least in Spanish language, cause a certain degree of confusion by their connotations.

16. Rabinow, P., ed. What is an author?. *In: The Foucault reader*. London: Penguin Books, 1986, 101-120.

17. BÜRGER, P. Interpretation after Duchamp. *In*: MEUTSCH, D., VIEHOFF, R., ed. *Comprehension of literary discourse*. Berlin: de Gruyter, 1989, 47-55.

18. KIRCZ, J.G. Rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval. *Journal of Documentation*, 47, (4), 1991, pp. 354-372.

19. HUTCHINS, J. Summarization: some problems and methods. In: *Informatics* 9: meaning, the   frontier of informatics; proceedings, 1987.

20. BARDIN, L. *Análisis de contenido*. Madrid: Akal, 1986. The author refers us to the fact that "while attempting interpretation, content analysis swings between two extremes : that of the strictness of objectivity and that of the fertility of subjectivity".

21. BEGHTOL, C. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42 (2), 1986, 84-109.

22. LYONS, J. *Language, meaning and context*. Bungay, Suffolk: Fontana Paperbacks, 1981.

23. HJORLAND, B. The concept of "subject" in information science. *Journal of Documentation*, 48 (2), 1992, 172-199.

24. SALTON, G. *Automatic text processing.* Cornell: Addison-Wesley, 1988.

25. HICKS, C., RUSH, J., STRONG, S.: Content analysis. *In: Encyclopedia of computer science and technology.* New York: Jack Belzer, 1977, V.6.

26. EVANS, A.: Basics of concept classification. *In*: JONES, K., ed. *Proceeding of 11th Informatics Conference.* London: Aslib, 1991.

27. CREMMINS, E. Value-added processing of representational and speculative information using cognitive skills. *Journal of Information Science*, 18, 1992, 27-37.

28. BOURNE, L.E., DOMINOWSKI, R.L., LOFTUS, E.F. *Cognitive processes*. New

Jersey; Prentice-Hall, 1979.

29. VEGA, M. *Introducción a la psicología cognitiva*. Madrid: Alianza, 1984.

30. ANTONINI, H.M., PINO, J.A. Modelos del proceso de lectura: descripción, evaluación e implicaciones pedagógicas. *In*: PUENTE, A., dir. *Comprensión de la lectura y acción docente*. Madrid: Fundación Germán Sánchez Ruipérez, 1991, 137-160.

31. BOURNE, L.E., DOMINOWSKI, R.L., LOFTUS, E.F. *Cognitive Processes*. New Jersey: Prentice- Hall, 1979.

32. MCKOON, G., RATCLIFF, R. Inference during reading. *Psychological review*, 99 (3), 1992, 440-466.

33. VAN DIJK, T. *La ciencia del texto*. Barcelona: Paidós, 1978.

34. CHARAUDEAU, P. *Langage et discours: elements de sémio-linguistique*. Paris: Hachette Université, 1983. Apud KOBASHI, N.: Análise documentária e tipologias discursivas. *In:* Grupo TEMMA. *Análise Documentária: consideraçoes teoricas e experimentaçoes.* Sao Paolo: FEBAB, 1989, 31-44.

35. KUPPERMAN, J., McGRADE, A.S. *Fundamentos de lógica.* Barcelona: Labor, 1973.

36. DEAÑO, A. *Introducción a la lógica formal*. Madrid: Alianza, 1989.

37. PINTO MOLINA, M. *El resumen documental: principios y métodos*. Madrid: Fundación Germán Sánchez Ruipérez, 1992.

38. GARDIN, J.C. Document analysis and linguistic theory. *Journal of Documentation*, 29 (2), 1973, 137-168.

39. TRIER, J.: *Der deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes*. I: Von den Anfängen bis zum Beginn des 13. Jh. Heidelberg, 1931, p.1 and following. Cfr. GECKELER, H. Semántica estructural y teoría del campo léxico. Madrid: Gredos, 1976.

40. WEISGERBER, L. *Das Menscheitsgesetz der Sprache als Grundlage der Sprachwissenschaft.* Heidelberg, 1964, p.71. Apud GECKELER, H. Op. cit., 127.