

# A THEMATIC ANALYSIS OF TECHNICAL DOCUMENTS: THE COLLECTION AND FORMALIZATION OF INFORMATION RELATING TO THE NEEDS OF PERSONS WITH DISABILITIES

Dr. Mabrouka El Hachani, Elico, Université Jean Moulin, Lyon, France

E-mail : [mabrouka.el-hachani@univ-lyon3.fr](mailto:mabrouka.el-hachani@univ-lyon3.fr)

Dr. Omar Larouk, Elico, Ecole Nationale des Sciences de l'Information et des Bibliothèques, Lyon, France

E-mail: [omar.larouk@enssib.fr](mailto:omar.larouk@enssib.fr)

## ARTICLE INFO

Studies and Articles

Received: September, 07.2017.

Revised: November, 30.2017.

Accepted: December, 08.2017.

doi:[10.5937/IJCRSEE1702099E](https://doi.org/10.5937/IJCRSEE1702099E)

UDK

376:004.91

### Keywords:

*technical document,*

*methodology,*

*discourse analysis,*

*linguistic*

*framework,*

*referentiality,*

*information,*

*cognition.*

## ABSTRACT

This article addresses the issue of technical documents and the impact their structuring has on professionals working in the field of disability. A model was developed to accompany users with disabilities when entering information; it was based on the linguistic analysis of an assortment of technical documents (institutional forms). We studied the problems related to technical documents by using a methodological approach which was based on a combination of complementary studies. The first was a terminological study, which aimed to look for relevant themes. It was complemented by a second one, which analyzed the discourse that was transmitted by the texts and which was centered on exploring the logical links between propositions and paragraphs. We completed these two studies with a final one, the logico-cognitive analysis. This which allowed us to look for contextual clues, which were produced by the texts in the documentary environment. It was therefore possible to formalize information and communication operations thanks to the reference discourse's targeted environment in view of designing an automated information system.

© 2017 IJCRSEE. All rights reserved.

## 1. INTRODUCTION

The main mission of the MDPH (*Maison Départementale des Personnes en situation de Handicap* – a local agency for people with disabilities) is to individually assist people with disabilities and to meet their specific needs. The goal is to present the chronological process in which a composite file (identity, region, contributing factors, etc.) is completed in order for the assessing team to make a quick decision for each situation. This composite folder (file) is a technical document with a specific structure. The linguistic content of these

forms includes textual information referring to the user's situation. It takes into account the information and needs of persons with disabilities. Professionals and social workers participate in providing guidance for the applicants. Our collection of documents was composed of the technical files of the MDPH-France, which are decisive in the context of a disabled person's life project. This study delineates the formal frameworks of the documents and the way they are exploited in order to allow the optimal reorganization of the way information is entered.

A methodological approach was used to analyze the problems related to this technical documentation. A terminological study was first implemented to search for relevant themes. It was then followed by a linguistic analysis of the discourse that was conveyed by the texts, and by an exploration of the links between the intra-sentence (intra-phrases), the intra-paragraph and the inter-paragraphs. The approach was centered on cognitive reference, in which the documentary environment

Corresponding Author

Dr. Mabrouka El Hachani, Elico, Université Jean

Moulin, Lyon, France,

E-mail : [mabrouka.el-hachani@univ-lyon3.fr](mailto:mabrouka.el-hachani@univ-lyon3.fr)



This work is licensed under a [Creative Commons Attribution - NonCommercial - NoDerivs 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/). The article is published with Open Access at [www.ijcrsee.com](http://www.ijcrsee.com)

conveys information and communication operations. It consists in extracting information from a set of heterogeneous technical documents. Such heterogeneity stems from the composite nature of the data (framework, texts, acronyms, symbols, etc.) which must be analyzed in order to offer recommendations to improve the structure (form, order, etc.) of MDPH documents.

## 2. STATE OF THE ART

### 2.1. Structuring technical documents as working documents

Otlet (1934) and Briet (1951) feature among the founders of document theory. Such thinkers were later reviewed by researchers in the late 1990s, such as Buckland (1997), who reflected on the nature of documents. Lund and Buckland (2008), Lund (2009) and Bachimont (2009) followed as they interrogated the transformations of the document in the age of digitization. In 2015, André Tricot defined a document in the following way, based on previous works by Buckland and Briet:

*“A document is an object that may or may not contain inscriptions, and which was conceived as a document and is perceived as such: it bears communicative and mnemonic intentions which are recognized by the users of the document (p. 17)“* (translated from French: Un document est un objet qui porte des inscriptions ou non, qui a été conçu comme un document et qui est perçu comme tel: il porte une intention communicative et une intention mnésique qui sont reconnues comme telles par les usagers du document (p.17)

Other researchers, such as Maurel, Mas, (2015), have tried to identify the characteristics of documents in organizations. He states:

*“The types of documents can therefore be characterized by their information structure (form and content) and by working practices shared by organizational partners. They are rooted in a context that stimulates their creation, their use and their preservation, thus providing a basis for mutual understanding among members of a given organization. The concept of the document type brings a new dimension to the typology of documents, in that it approaches it within a broader context and from different angles. (p. 3)“* (translated from French : « Les genres de documents sont donc caractérisés par une structure d’information (forme et contenu) et par des pratiques de travail que partagent des acteurs organisationnels. Ils sont ainsi ancrés dans un contexte

qui motive leur création, leur utilisation et leur conservation, et qui fournit une base de compréhension commune entre les membres de cette organisation. La notion de genre de documents apporte une nouvelle dimension à la typologie des documents, en ce qu’elle l’aborde dans un contexte plus large et sous des angles différents » (Maurel, Mas, 2015: 3).

The working document or the document type therefore reinforces a specific organizational reality. The field in which the working document exists is filled with other considerations, the latter of which allow us to gain perspective on the stakes that are at play in the documentary world and in informational and communicational transactions (Maurel, Mas, (2015)).

In the case of this research, we focused on the specific arena of the medico-social field. Verdier and Flory (2009) explain that *“For a long time, medical information was, in the worst of cases, limited to a paper medical record. In the best of cases, it was structured in relational formalism. Today, medical information is one of the keys to quality medical care”*. (translated from French: Longtemps cantonnée à un traditionnel « dossier médical » en papier, dans le pire cas, ou structurée dans un formalisme relationnel dans le meilleur cas, l’information médicale est aujourd’hui une des clés favorisant une médecine de qualité) (Verdier, Flory, 2009).

The development of structured documents in the 1990s and the 2000s led to health information being strategically repositioned in medical organizational structures (Verdier, Flory, 2009). The long history of patient records meant that these records could not really be implemented on a large scale due to technical and legal issues related to the protection of health data. The health sector has therefore made considerable progress on issues related to the management of health information. The medico-social sector, however, has just barely started the process with its “user” files. An initiative set up by an association called “Ob-Serveur”, (“The observatory of the needs of people with disabilities”) explains the value of centralizing information in a clear and lasting way both for the proper follow-up of users with disabilities and for those whose situation changes with time and health status. Due to the wide spectrum of disabilities and the variability of difficulties encountered by people in all aspects of their lives, the Piveteau report (2014) found that a number of disabled users did not find solutions for compensation, assistance, accommodation or social support,

despite the enforcement of the 2005 law for equality and despite the creation of a dedicated agency for people with disabilities known as the *Maison Départementale des Personnes en situation de Handicap – MDPH*. The purpose of such structures is to collect the people's applications for help and compensation. The Piveteau report warns that the situation is critical. Response times are too long and can sometimes be inappropriate, as the needs of the population are not fully taken into account, particularly in the application forms. These forms, in fact, require individuals to fill an empty frame without considering the difficulties of expression people with disabilities may have.

It is in this context that we wished to study a collection of three types of forms from three regions of France (Isère, Rhône, Calvados). These forms are viewed as being the working documents of the agents who collect the information. These documents then become the memory of the situation of the people who, in turn, reply. Such forms are an official notification of help. They are therefore vital documents in the lives of people with disabilities.

## ***2.2. Textual and argumentative analytical models in technical documents***

In this state of the art, which focuses on content analysis and on extracting knowledge from documents, we examined specialized texts that contained form-type spaces and procedures, in an effort to observe the document's technical organization and procedural content. However, linguistic and logical methods had already provided theoretical steps for content analysis.

At its origin, textual analysis focused on documents that could be referred to as "technical" in so far as they contained presentation rules and a specialized discourse (administrative, institutional, scientific, political, etc.). In his work, Pêcheux (Pêcheux, 1969) dealt with documents such as parliamentary speeches, trade union texts and others. According to him, in Pêcheux's sense, discourse analysis is a multidisciplinary (qualitative and / or quantitative) approach that studies the content of discourse. This approach is concerned with organizing the oral and written discourses that are under study. Pêcheux was inspired by the work of Roman Jakobson, especially regarding the functions of language. Jakobson distinguishes six functions in language:

*1.) the referential or representative function, where the statement gives the state of things (also called semiotic or symbolic); 2.) the expressive function, in which the subject expresses his own attitude towards what he is talking about; 3.) the conative function, where the statement aims to act on the recipient; 4.) the phatic function, where the utterance reveals the links and maintains the contacts between the speaker and the recipient; 5.) the metalinguistic or meta-communicative function, which refers to the linguistic code itself; 6.) the poetic function, where the utterance is endowed with a value as such, a value that brings creative power.*

The discourse also calls upon the phatic function, in which the utterance reveals links and maintains contacts between the speaker and the recipient; and *the meta-communicative function*, which refers to the linguistic code itself as understood by Jakobson, and which leads to verbal interactions. This approach gave rise to a new pragmatic analysis of interactions, thus allowing us to examine modern-day forms of discourses.

## **2.3. A basic approach to an argumentative analysis in statements**

Breakthroughs in logical analysis came from linguists such as Bally, and logicians such as Quine, Russell and Frege, who treated problems of representation (Larouk, 1993) with predicative relations. Predication can be separated into theme and predicate thanks to a certain relation of order. Lyons suggests the terms 'topical' and 'commentary'. "Expressive" and "referential" functions are thus found in the works of Charles Bally, who theorized a stylistics of discourse by observing this textual arrangement. A document therefore bears linguistic inscriptions that convey acts of discourse. An act of discourse is an act of expression and an act of representation as Charles Bally specified by focusing on enunciation. A predicate is what is affirmed about a referent as designated by the subject. Bally defines the predicative relation (page 101) in the following terms:

*"Every utterance logically includes two terms, the thing that is spoken of and what is said about it. What is said about it is the statement or predicate (in the broad sense). The term, which is the purpose's occasion, is the theme or the subject (in the broad sense)".*

(translated from French: Toute énonciation comprend logiquement deux termes, la chose dont on parle et ce qu'on en dit; ce qu'on en dit est le propos ou prédicat (au sens large), le terme, qui est l'occasion du propos, est le thème ou sujet (dans le sens large)).

Language thus provides various types of predicative relationships, as in this verbal statement:

<1> My disability rights [lower] /  
Theme [predicate] = Term + [quote] =  
Theme + [rhetoric]  
A + B = Topic + [comment].

We find that the predicative structure is always binary:

{P = [ATerme + Bpredicate]} with  
{Phrase = A [My disability rights] + B[drop]}

The relation theme/predicate makes it possible to "locate the dynamic aspect of the information contained in the enunciation". Therefore, the relative predicate enriches the subject.

Questions were raised regarding the elements that describe a technical document, based on observations made on textual documents provided by the MDPH: more specifically, what are the salient features that represent the intra-textual discourse, which is itself targeted by linguistic content. The work of indexing starts from the textual data in its original form in order to end with a representative form of meaning, summarized by Michel Le Guern's proposition (Le Guern, 1984: 2): "The descriptor is not a word of the language that belongs to the lexicon (language), but it is an updated word. Such updating entails the construction of the syntagma. The descriptor is not a word of language but rather a syntagma of discourse". (translated from French: Le descripteur n'est pas le mot de la langue en tant qu'élément du lexique (langue), mais le mot actualisé. Cette actualisation passe par la construction du syntagme. Le descripteur n'est pas mot de langue mais syntagme du discours).

In the sense of Le Guern, the content of the document is represented by the list of extracted Nominal Syntagmas (SN) or Noun phrases (NP) which refer to extra-linguistic discourse. Thus, the SNs are found in the following example:

<2> / The economic policy of *France*  
must respect the recommendations of *Europe* /  
<21> / France / <23> / The economic

policy of France /  
<22> / Europe / <24> / Recommendations of Europe

The "simple" referential terms represented by NP = <France, Europe> are, to a certain extent, items of knowledge that are extracted from the textual document in a terminological way. The entire list <2<sub>1</sub>>, <2<sub>2</sub>>, <2<sub>3</sub>>, <2<sub>4</sub>> creates the corpus's referential environment.

Michel LeGuern integrated C. S. Peirce's semiotic contribution to his theoretical approach in order to understand the referential mechanism [pp. 23-24].

"The word in language, unlike its occurrences in discourse, has no extra-linguistic reference: the sign / object relation, in Peirce's sense, broadly corresponds to the Saussurian signifier / signified relation; the signified aspects are also a part of the language's structure. In order for the descriptor to fulfill its function, which is to relate an object - an extralinguistic entity - to a document that will provide information on this object, the descriptor must be an index sign. (translated from French: Le descripteur n'est pas le mot de la langue en tant qu'élément du lexique (langue), mais le mot actualisé. Cette actualisation passe par la construction du syntagme. Le descripteur n'est pas mot de langue mais syntagme du discours). According to Le Guern, this operation makes it possible for predicative properties of textual structures to be updated by relating the properties with the objects of extra-linguistic reality in which the discourse environment is contained. The desired linguistic property is therefore the part of the discourse that bears reference to extra-linguistic reality, and the Nominal Syntagma (Noun Phrase) is the smallest part of the discourse with such a property.

### 3. WHAT TYPES OF STUDIES FOR SPECIALIZED DOCUMENTS?

#### 3.1. Characteristics of the corpus

Our corpus consists of technical documents produced by the administration of social affairs (Rhône, Calvados, Isère) and intended for an audience of people with disabilities. The latter individuals have motor, physical, auditory, visual, and / or psychic impairments. In order to improve the users' social status and to expedite their requests, analyzing the content of the documents becomes necessary in an ef-

fort to suggest a simpler process to access digital administrative information. We note that the MDPH technical documents are composed of **linguistic data** and **graphic data** (images, frames, initials, symbols, etc.). Data is contingent on the producers of the documents in paper form and is intended for aid applicants. Mixed structuration is observed in textual and/or iconographic blocks. These documents are

composed of linguistic signs (titles, legends, summaries, acronyms, etc.) and iconographic signs (figures, tables, etc.). This bimodal information is sometimes numbered. These documents are designed and intended for the general public. They are consulted and interpreted within the context of form-filling transactions in order to have access to state aid.

Figure 1. Example of a fragment of the technical corpus (Block with title)

Thus, this corpus is made up of statements formed through sequences of discourse that are produced by the administration, via documents for requests that are formulated by the disabled users. These textual documents reveal informational situations that highlight ties with the exclusion or inclusion of a specific status. Referential links serve to classify disabled persons in targeted categories, such as the following propositions:

<3> / If the person concerned is under 18 years of age, his / her parents are invited to answer for him/her./

We found that our documents presented a structuring of the zones (*text blocks in the form of consecutive paragraphs*) with an indication of the territory location (*region/department/tow/commune/*), but also through other logical and linguistic elements intra/inter-document.

We are interested in the medium as well as the content in the aim perspective of reorganizing the linearity of information to facilitate access to the semantics of the questions and expressions that are conveyed in these working documents (institutional or administrative).

### 3.2. Towards a thematic and logico-cognitive analysis of the contents

According to the framework of the MDPH documents, we describe the work we undertook to delineate the formal frameworks of the documents and their exploitation in order to reorganize information entries. To design a computer tool, we started with the corpus of technical documents, which respects a certain homogeneity of content that is linked to the producers of institutional and administrative documents. We attempted to define the notion of linguistic content as related to the discourse that is used in technical documents. We explored the concept of textual enunciation which refers to referential objects. This information communication analysis calls upon the references of discourse and is based on the semiotics of the document. We asked the following question, paying particular attention to textual data in technical documents: *What theme is spoken of and how is it delimited in relation to cognitive elements?*

#### 3.2.1. A thematic analysis of block titles

The terminology aims to locate the rel-

evant information that is contained in a text of the technical genre. The overall meaning of utterances often results from the meaning of the terminological units in the utterances, while entering into a process of constructing meaning throughout the enunciation (and thus of texts).

<4> / *The MDPH, This is the Local Agency for persons with disabilities. It studies your situation to meet the needs related to your disability* /

The desired linguistic property is therefore the part of the discourse that bears reference to extra-linguistic reality. We can show that the Noun Phrase (NP) is the smallest part of the discourse in this introductory paragraph.

<4<sub>1</sub>> / *The MDPH* / <4<sub>3</sub>> / *the Departmental House of Persons with Disabilities* /  
<4<sub>2</sub>> / *your situation* / <4<sub>4</sub>> / *your handicap* /

<5> / *You already had a file at the MDPH* / or <5> / *You already had a folder at the MDPH* /

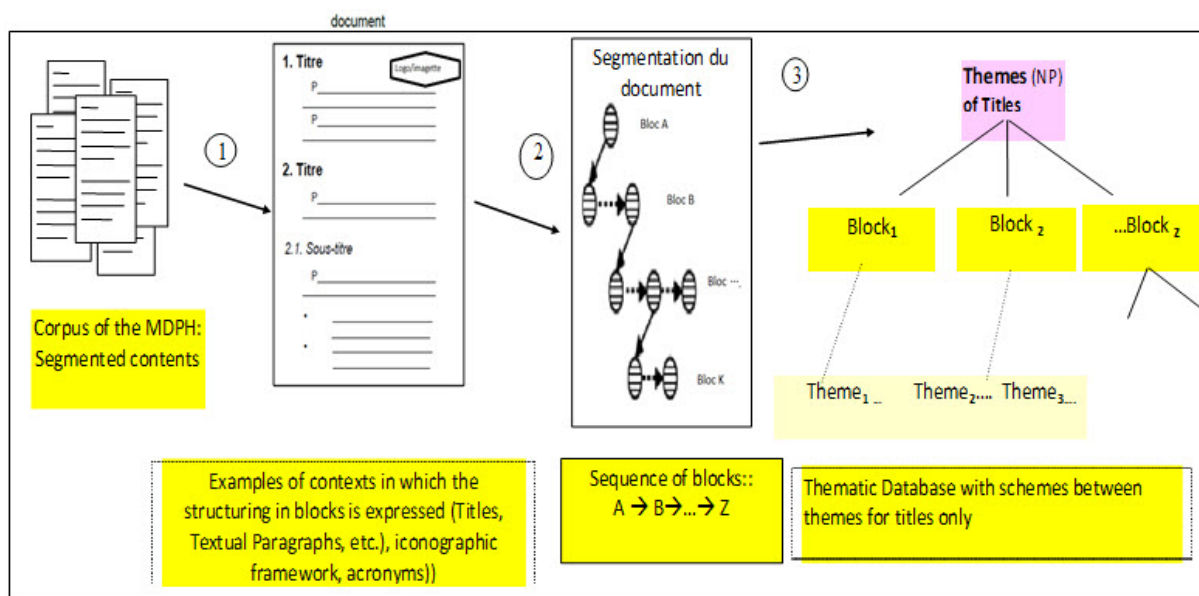
<4<sub>1</sub>> / *The MDPH* / <4<sub>3</sub>> / *a folder* ; / *a file* /

The discourse updates meaning that is stabilized and allows the dynamic construction of new meanings. The different relations

between thought and language that occur in enunciation lead to a progressive construction of meaning. However, this progressive construction of meaning in the statement relativizes the role of linguistic categories and categories of thought, since the relations between thought and language can be formulated in terms of an exploration of the utterance. The subject is therefore the referential construct elaborated by the producer (in this case, the administrator) on the basis of recurring textual elements inscribed in the text (the theme: what we are talking about) followed by a descriptive function in the discourse.

### 3.2.2. Method for describing the thematic base of block titles: An inventory of title themes

In our corpus, a document links a producer (administrator, author), and an applicant (consumer, disabled) via two elements that are a discourse (an argumentative and iconographic content) and a medium (paper). For the formalization of the database architecture, we begin with the salient parts of the technical document that evoke referential objects. Titles are viewed as prominent traces that require prioritization. This extraction makes it possible to make the inventory of the themes (SN) in the communication of the information.



**Figure 2.** Content description method for the creation of a thematic base (SN) with inclusive diagrams of themes.

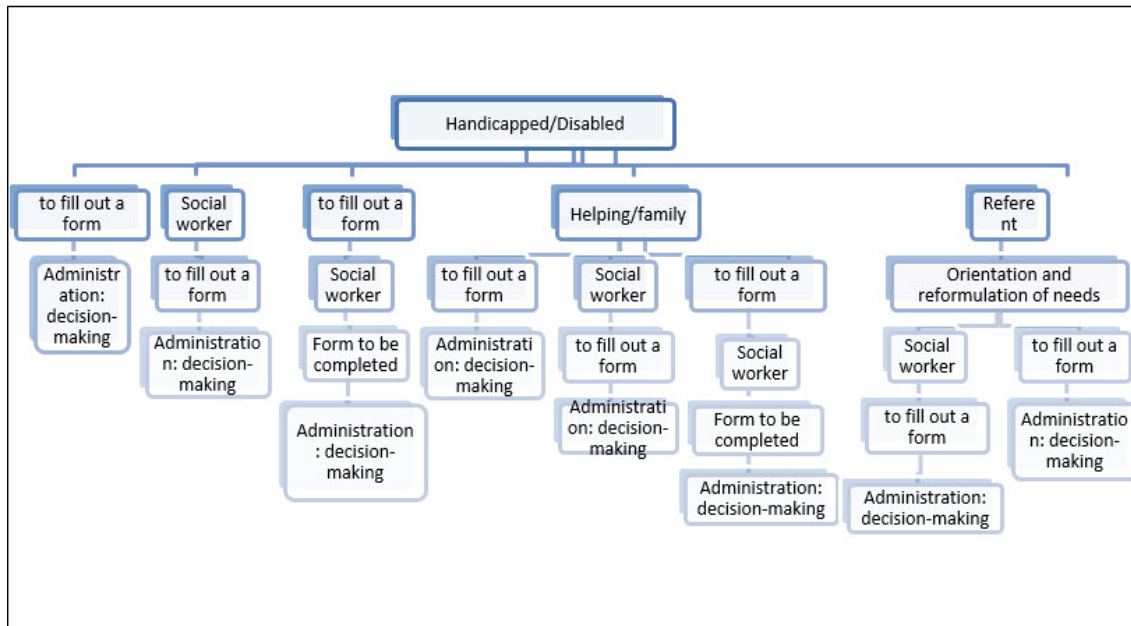
- i. The textual data in the technical documents take the form of a list of statements (1).
- ii. The study of these propositions allows us to observe block structures with frames for the titles and paragraphs with specialized

- content (2).
- iii. The identification of topics aggregates these referential NPs to conceive a thematic base (3).

### 3.2.3. Access path to the content of a help form

We observed that an MDPH document connects a producer (administrator, author), with an applicant (person with a disability) via two elements that are a discourse (argumentative and iconographic content) and a medium (paper). The communicational function of

documents and description situations can be found in this space. This fits well within the context of situation theory, since all of the information is supported by a situation such as territorial localization (urban / suburban), targeted state funding (state / department / metropolitan), the age to determine the recipient's orientation (specialized institutions, hospitals, schools, high schools, universities, etc.). All these signs characterize the type of discourse.



**Figure 3.** The formalization of an applicant's information activities: access routes to the content of a help document, Help System for Entering Information (HSEI)

The tree structure diagram is a summary of our recommendations based on the results from our previous studies. It reproduces the procedures needed to accompany a worker in his search for a suitable and optimized path. We are immediately able to recognize the effects of such a tree diagram for programming. This diagram refines the process of data entry for any requests for assistance that are solicited by the disabled persons, the caregivers, the referents and the social assistants.

## 4. CONCLUSION

The MDPH encounters difficulties in handling the users' requests for help and disability compensation. One of the main difficulties lies in understanding the administrative form itself, in view of the diverse situations of users and the inability for some people to express their needs and wishes, depending on their situation and their disability.

Our analysis of technical documents consisted in extracting salient elements and suggesting a way to reorganize these adminis-

trative forms so that they may be adapted and made into digital versions. This seems to be a prerequisite to improve the situations of these people. A dynamic document model would therefore need to be imagined for more interaction with the applicant. The Help System for Entering Information (HSEI) data capturing process shows the importance of support when gathering information on the needs of individuals.

Our analytical framework allowed us to identify associated themes (titles and intra-discourses), and to realize that the aggregation with the intra-stated and inter-propositional referential themes led to the logical-cognitive referential schemes of the discourse. The inter-proposal themes made it possible to fine-tune a document's information content (co-referencing, implication, causality, etc.) through its composite structure. In order to identify specificities in texts with a view to create a tree thematic construction, we suggested a knowledge representation model to guide the discursive analysis. This took the form of axes that structured a block or of crossed themes that allowed intra-documentary navigation be-

tween text blocks.

## ACKNOWLEDGMENT

This research project is part of a project entitled ACAPELHA (Support for People with Disabilities Adults and Children (2015-2018)) and funded by the Regional Health Agency (Agence Régionale de Santé Auvergne-Rhône-Alpes). The Elico research laboratory participates in the SHS component of the project coordinated by Ressourcial organization (<http://www.ressourcial.fr/acapelha/>).

### Conflict of interests

The authors declare no conflict of interest.

## REFERENCES

- Bachimont, B. (2009). Archivage audiovisuel et numérique: les enjeux de la longue durée. *Archivage et stockage pérennes. Paris: Hermès Science Publications: Lavoisier*. pp. 195-222.
- Bally, C., (1944). *Linguistique générale et linguistique française*. Édition PUF.
- Briet, S. (1951). *Qu'est-ce que la documentation?*, Paris, EDIT -Éditions documentaires, et techniques.
- Buckland, M. K. (1997). What is a "document"? *Journal of the American Society for Information Science* (1986-1998), 48(9), 804. <http://www.columbia.edu/cu/libraries/inside/units/bibcontrol/osmc/bucklandwhat.pdf>
- Frege, G. (1879), *Idéographie*, traduction Corine Besson 1999, Paris : Librairie philosophique J. Vrin, 1999
- Jakobson, R., (1960). *Closing statements: Linguistics and Poetics*, Style in langage, traduction du livre de T.A. Sebeok, New-York.
- Larouk, O., (1993), «Extraction des connaissances à partir de documents textuels : Traitement automatique de la coordination (Connecteurs et Ponctuation)», Ph.D, Diplôme de Doctorat, spécialité : Sciences de l'information et de la communication, et Informatique, (Doctoral dissertation, Lyon 1). Université Lyon 1, p. 365.
- Le Guern, M. (1984). Les descripteurs d'un système documentaire: essai de définition. *Condenser, suppl*, 1, 163-169.
- Le Guern, M. (1991). Un analyseur morpho-syntaxique pour l'indexation automatique. *Le français moderne*, 59(1), 22-35.
- Lund, N. (2009) Document theory. *Annual Review of Information Science and Technology* 43:399–432 . DOI: 10.1002/aris.2009.1440430116
- Lund, N. & Buckland.M. (2008) "Document, Documentation, and the Document Academy". *Archival Science* 8: 161-164.
- Lyons, J., (1970). *Linguistique générale: Introduction à la linguistique théorique*. traduit. F. Dubois-Charlier et D. Robinson dans Langue et langage; Larousse.
- Maurel, D., Mas, S., (2015). Genres de documents et coordination des activités dans les organisations: éléments théoriques (p. 69-87). In. *Québec: Presses de l'Université du Québec*. p. 224.
- Otlet, P (1934) *Traité de documentation*. Mundaneum, Brussels
- Pêcheux, M., (1969). *Analyse automatique du discours*, Dunod, Paris.
- Peirce, C. S. (1978). *Écrits sur le signe: rassemblés, traduits et commentés par Gérard Deledalle*. Seuil.
- Piveteau, D, "Zéro sans solution" : Le devoir collectif de permettre un parcours de vie sans rupture, pour les personnes en situation de handicap et pour leurs proches, Rapport officiel, juin 2014 online: [http://handicap.gouv.fr/IMG/pdf/rapport\\_zero\\_sans\\_solution.pdf](http://handicap.gouv.fr/IMG/pdf/rapport_zero_sans_solution.pdf)
- Quine, W. (1962), *Mathematical Logic*, New York : Harper & Row edition,
- Russell, B. (1905). Toward "Principia mathematica", London; edited by Gregory H. Moore (first Edition), recent edition: "The collected paper of Bertrand Russell" London. New York Routledge (2014)
- Tricot, A., Sahut, G., & Lemarié, J. (2016). *Le document: communication et mémoire*. De Boeck Supérieur. <https://goo.gl/BCYxQV>
- Verdier, C., Flory, A., (2009). « Éditorial », *Document numérique*, 2009/3 (Vol. 12), p. 7-8.