Construction and Evaluation of Hierarchical Structures of Indexing Languages for Online Catalogs of Libraries: An Experience of the São Paulo State University (UNESP)†

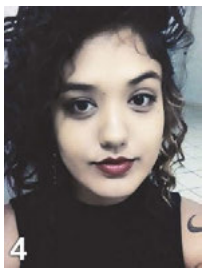
Mariângela Spotti Lopes Fujita*, Walter Moreira**,
Luciana Beatriz Piovezan dos Santos***, Maria Carolina Andrade e Cruz****,
Rosane Rodrigues de Barros Ribas*****

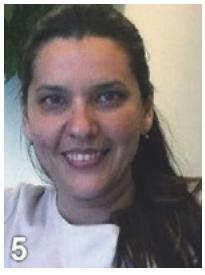
*Researcher of Information Science Academic Program, São Paulo State University—UNESP, Marília Campus, Av. Hygino Muzzi Filho, 737, Marília, SP, Brasil, 17525-300, <fujita@marilia.unesp.br>
**São Paulo State University—UNESP, Information Science Department, Marília Campus, <walter.moreira@marilia.unesp.br>
***São Paulo State University—UNESP, Doctorate Student in Information Science Academic Program, Marília Campus, <lbpiovezan@gmail.com>
**** São Paulo State University—UNESP, Master Student in Information Science Academic Program, Marília Campus, <maria.andradeecruz@gmail.com>
*****São Paulo State University—UNESP, Librarian of the Documentation Informations Group, Rua Quirino de Andrade, 215—Centro—São Paulo/SP—CEP 01049-010, <rosane@reitoria.unesp.br>

Mariângela Spotti Lopes Fujita is a researcher. She has a research grant from the National Research Council for Science and Technology in Brazil to further studies on indexing and indexing languages in the field of information science. She holds a master's and PhD in communication sciences from São Paulo University, and she graduated in library science. She teaches postgraduate courses in production and organization of information at Sao Paulo State University (UNESP).

Walter Moreira holds a graduate degree in librarianship and a PhD in information science from the University of São Paulo (USP). He is a professor in the information science department of São Paulo State University (UNESP), working in undergraduate courses in archival science and library science and graduate course in information science. He develops research on knowledge organization systems, documentary languages, terminology, ontologies, and classification theory.

Luciana Beatriz Piovezan is a PhD candidate in information science at São Paulo State University in Brasil, studying index language building and management. She holds a master's degree in information science and a bachelor's degree in library science from the same institution. Her others research interests are subject indexing, indexing policy, and knowledge organization in university libraries.





Maria Carolina Andrade e Cruz received a degree in library science in 2016 and currently is a master's student in São Paulo State University (UNESP). She develops research activities in the line of production and organization of information specifically focused on the topics of information organization policies and indexing languages.

Rosane Rodrigues de Barros Ribas has a graduate degree in library science. She is a librarian at the Documentation Informations Group of São Paulo State University (UNESP) where she works in the legal area, especially in doctrine, jurisprudence, and legislation research and in cataloging. She is a member of the Indexing Policy Group of the UNESP Libraries Network and the UNESP Language Group where she collaborates in the revision and correction of authority records and in the creation of new terms for the UNESP Language.

Fujita, Mariângela Spotti Lopes, Walter Moreira, Luciana Beatriz Piovezan dos Santos, Maria Carolina Andrade e Cruz and Rosane Rodrigues de Barros Ribas. 2018. "Construction and Evaluation of Hierarchical Structures of Indexing Languages for Online Catalogs of Libraries: An Experience of the São Paulo State University (UNESP)." *Knowledge Organization* 45(3): 220-231. 26 references. DOI:10.5771/0943-7444-2018-3-220.

Abstract: The construction and updating of indexing languages depend on the organization of their hierarchical structures in order to determine the classification of related terms and, above all, to allow a constant updating of vocabulary, a condition for knowledge evolution. The elaboration of an indexing language for online catalogs of libraries' networks is important considering the diversity and specificity of knowledge areas. From this perspective, the present paper reports on the work of a team of catalogers and researchers engaged in the construction of a hierarchical structure of an indexing language for an online catalog of a university library's network. The work on hierarchical structures began by defining the categories and subcategories that form the indexing language macrostructure by using the parameters of the *Library of Congress Subject Headings*, the National Library Terminology and the Vocabulary of the University of São Paulo Library's system. Throughout the stages of the elaboration process of the macrostructure, difficulties and improvements were observed and discussed. The results enabled the assessment of the hierarchical structures of the languages used in the organization of the superordinate and subordinate terms, which has contributed to the systematization of operational procedures contained in an indexing language manual for online catalogs of libraries.

Received: 17 October 2017; Revised: 22 February 2018; Accepted 28 March 2018

Keywords: indexing languages, hierarchical, engineering, subject heading lists

† The research within the project "Indexing Language for Libraries in the Perspective of Indexing Policy" has received funding from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

1.0 Introduction

Indexing languages, such as thesauri and subject heading lists, used in online library catalogs for vocabulary control, have fundamentals associated with bibliographic classification systems by the hierarchical organization of their descriptors and subject headings. For Cleveland and Cleveland (2001, 38), "the indexing languages with vocabulary control devices such as subject heading lists, thesauri and classification schemes are assigned-term systems." Hjørland (2012, 304) contests the distinction between classification schemes and indexing languages "because the act of labeling a document (say by assigning a term from a controlled vocabulary to a document) is at the same time to assign that document to the class of documents indexed by that term (all documents indexed or classified as X belong to the same class of documents)." The terminological issue, as it can be seen with the small sample cited, is still not resolved. Recently, the term KOS has received increasing attention in the literature. In the Encyclopedia of Knowledge Organization, the entry by Mazzocchi (2018) on knowledge organization systems (KOS) relates the term indexing language as a part of the knowledge organization (KO) terminology because it is "still employed in a LIS/KO environment," despite of considering that the terms referring to the notion of "language" has been replaced by "system," following the inclination of the NKOS (networked knowledge organization systems) community.

The construction and updating of indexing languages depend on the organization of their hierarchical structures formed by categories and subcategories aimed to determine the classification of related terms and, above all, to allow a constant updating of vocabularies, a fundamental condition for knowledge evolution. For this reason, a consistent methodology and support are required from other indexing languages that have proven useful and representative of knowledge areas. The consultation of vocabulary resources is recommended in ISO 25964-1 (2011), which provides a list of resources for the construction of thesauri, such as existing thesauri or classification schemes, collections of terminology, indexes of existing databases, transaction logs of relevant websites and standard reference works (89-90). Thesauri or classification schemes as well as other tools that will serve as bases for the development of the thesaurus need to be supported by warrants to ensure their best use. Thus, the choice of terms should reflect the literary, organizational and user warrants as specified in NISO Z39.19 (2005).

The elaboration of an indexing language manual containing systematized procedures for online catalogs of libraries' networks is important considering the diversity and specificity of knowledge areas whose organization and representation need a hierarchical structure in an indexing language. Library catalogs in an online environment are very useful for various purposes: users can search, access and retrieve information by interacting with websites, which helps library catalogers in the construction, updating and insertion of new bibliographic records.

Access and retrieval by subject is undoubtedly the major advantage of online catalogs and, for this reason, they depend on the organization of their contents by subject cataloging. Bates (1989), who brought out the discussion on

the need for introducing online catalogs into libraries, recognized that they would open up new and impressive possibilities for retrieval and ease of use. The librarians' task would be to design the intellectual content and arrangement of catalogs so as to take full advantage of these new technical capabilities of subject cataloging.

The construction of online catalogs has advanced significantly with the creation of many applications resulting from technological evolution. On the one hand, they refer to the descriptive and thematic representation of the information resources available in the bibliographic records built according to international standards. On the other hand, they are related to the construction of authority records for vocabulary control. Bibliographic records and their contents are continuously checked and corrected by authority records that warrant the consistency and correctness of access points either by authority or by subjects.

Sharing bibliographic and authority records is another advantage of libraries' online catalogs, whose main function is to avoid duplication of effort and, in this sense, norms and standards play an essential role for cooperation among libraries. Today, cooperative cataloging is a world-wide bibliographic record-sharing system pioneered by Charles Coffin Jewett in 1850 from the systematization of cataloging rules elaborated in 1840 by Anthony Panizzi et al. and published in 1841 by the British Museum (Campello 2006). In 1901, the Library of Congress began printing catalogs for sale, and, with the information technologies available today, it operates the Program for Cooperative Cataloging (PCC). Cooperative cataloging has advanced with the creation of cooperative cataloging centers in a large number of countries as well as in many library systems in various institutional settings, such as universities, corporations and governments.

In Brazil, cooperative cataloging began in 1942 with the creation of the Cataloging Exchange Service (SIC) by the Administrative Department of Public Service (DASP). In 1954, the SIC was transferred to the Brazilian Institute for Bibliography and Documentation (IBBD) and, in 1976, to the Getúlio Vargas Foundation (FGV) that created the BIBLIODATA/CALCO Network, a cooperative center for cataloging processes based on MARC II (Machine-Readable Cataloging), a format of data description managed by the Library of Congress (*Histórico* 2016). BIBLIODATA is a cooperative cataloging network among Brazilian libraries that provides bibliographic records, authority records and the controlled vocabulary "Subject Headings List of the BIBLIODATA Network" (LCARB).

The development of the LCARB began around 1977 as part of the BIBLIODATA/CALCO (Computer-readable Cataloging) project, a joint effort of FGV and the Brazilian Institute of Bibliography and Documentation (IBBD), currently named the Brazilian Institute of Information for Science and Technology (IBICT), aimed to standardize the use of subject headings (*Histórico* 2016). From 2013 until the present day, the BIBLIODATA network and the LCARB language have been managed by the IBICT. Because the discussion of the construction of the UNESP Language hierarchies is a major priority of the present study, it is important to highlight the significant contribution of the BIBLIODATA network in the development and maintenance of the UNESP online catalog, which currently enables bibliographic and authority records to be imported.

The UNESP (Universidade Estadual Paulista "Júlio de Mesquita Filho") is a relatively young public higher education institution (founded forty years ago), whose mission is to provide access to free and quality education. It is one of the largest universities in Brazil and the most successful model of a multi-campus university comprising thirty-four campuses and thirty-two libraries spread out in twenty-four strategically distributed cities in the State of São Paulo, the most developed state in the country. All the libraries are served by the UNESP Library Network, which, in turn, is managed by the General Coordination of Libraries, created in 1977 with the purpose of delivering support to the information needs in teaching activities, research and community extension for an effective interaction in the academic environment. Together, the thirty-two libraries give support to 37,770 students enrolled in 155 undergraduate courses and 13,541 students enrolled in 255 graduate programs (masters and PhD) according to the Statistical Yearbook 2016. UNESP also has 1,287 research groups working on various knowledge areas. These groups are registered by the National Council for Scientific and Technological Development (CNPq) and certified by the university (*Anuário Estatístico* 2016).

The automation of the UNESP Library Network began in 1997 with the acquisition of ALEPH software for the improvement of efficiency in the performance of computerizing tasks and routines as well as to optimize the use of its collection and resources. In order to create a bibliographic database, the UNESP Library Network has adopted cooperative cataloging for the conversion of bibliographic records of national and international databases. With the inauguration of the ATHENA catalog in June 1999, the bibliographic collection data already converted into machine language were made available on the internet.

Standards for the cataloging of bibliographic records of the UNESP Library Network were developed and then published in the book *Padrão de qualidade de registros bibliográficos da UNESP* (UNESP. CGB 2002), which establishes the procedures for ensuring the quality of the online catalog's bibliographic records shared through cooperation among the libraries. However, subject indexing was performed at UNESP libraries without a manual of procedures. Thus, library catalogers imported many bibliographic records from the BIBLIODATA network without

verifying whether the terms of this language were needed and, therefore, there was rarely a request for the inclusion of new ones in the LCARB.

When the creation of a bibliographic record for a nonexistent document in the BIBLIODATA network was required, the most pertinent heading of the document in the LCARB was adopted. Consequently, UNESP catalogers had no participation in the construction of the indexing language carried out by the BIBLIODATA system team. This problem was further aggravated by the use of several other languages in other subject fields in an attempt to find a solution to outdated terms. Thus, there was no vocabulary control of a single indexing language, an important condition to enable consistency between indexing and information retrieval. The LCARB, whose terms are arranged alphabetically, is a controlled and pre-coordinated language composed by translating and adapting the *Library of Congress Subject Headings* (*LCSH*) to Portuguese. According to the BIBLIODATA network subject headings manual, the option for the *LCSH* was based on its multidisciplinary characteristics and its reliability (*Fundação Getúlio Vargas* 1995). IBICT is responsible for both the BIBLIODATA network and its LCARB language.

Due to the problem of LCARB's outdatedness and the difficulty to obtain new terms from the maintainer institution, librarians had to create "local subjects" with no synonym control for document indexing, among other problems. Without the possibility of updating this language because the network cooperation service had been discontinued, it was necessary to substitute the LCARB terms for the National Library Terminology (TBN), *Library of Congress Subject Headings* (*LCSH*) and *Medical Subject Headings* (*MeSH*). Studies on the inadequacy of the LCARB for UNESP were carried out before the elaboration of a proposal for its library network indexing policy, thus resulting in the *Manual de política de indexação para as bibliotecas universitárias da UNESP* (*Universidade Estadual Paulista Júlio de Mesquita Filho* 2014) to be used by the UNESP libraries in particular. In addition to systematizing elements, variables and indexing processes for subject cataloging, the manual contains guidelines and procedures on the use of an indexing language that must be continually corrected and updated with new terms.

In this way, the LCARB would be gradually replaced by a new and more updated indexing language adapted to the reality of the UNESP libraries. This process of replacing one language with another was accomplished by matching the authority records provided by the LCARB with updated information from the indexing languages of the National Library Terminology (TBN), *Library of Congress Subject Headings* (*LCSH*) and *Medical Subject Headings* (*MeSH*) for records that had not been translated in the first one. After studies and comparisons that proved to be the best replacement option for the LCARB, the TBN was chosen, because it offers a more up-to-date translation of the *LCSH*, a visible hierarchical structure and thesaurus associative relationships that were unavailable in LCARB.

The internalization of the LCARB's authority records database in the ATHENA online catalog enabled the construction of the UNESP language based on the compatibility between its authority records and the three other indexing languages. To this end, catalogers began to import authority records with the new terms contained in the *LCSH*, National Library and *MeSH* languages. The need to update and to ensure compatibility with the records imported from the LCARB was also taken into account. However, the new terms that did not exist in the other languages had to be requested or created by the catalogers themselves, who often had to resort to dictionaries, the only reliable data sources. This practice gave rise to problems, such as duplication of authority records with similar terms and construction of authority records for a particular topic without identifying the relationships between generic and specific terms.

The UNESP Language Group, composed of catalogers and researchers, was created and trained to deal with the inclusion of new terms into the UNESP indexing language by gradually replacing the LCARB language previously adopted. The group began by focusing on meeting the catalogers' demands regarding the updating of authority records for topical and geographical terms that already existed in the language, as well as the inclusion of new terms that were nonexistent in the *LCSH*, TBN and *MeSH* languages. The new language, called the UNESP Language is a multidisciplinary thesaurus in Portuguese and English that "arose from the need to use an updated and hierarchical language that would represent the different thematic areas present in the library collection of the University" (Ribas and Parra 2016, 67).

As new terms were requested by the UNESP Library Network catalogers, the Language Group began to face some problems concerning the hierarchy of terms in the different languages adopted for compatibility of authority records and inclusion of new terms in the UNESP Language. Often, there were doubts as to which term would be more generic in relation to the one that would be included in the language, or even doubts regarding the interdisciplinarity of the term. Many issues such as these have been resolved while others are still under analysis; in any case, however, the context in which the problems occur is always taken into consideration. This procedure aims to ensure effective cultural warrant to vocabulary terms, since the relationships among concepts should be adequate to the users' cultural context and vocabulary. According to Beghtol (2002), cultural warrant "means that any kind of knowledge representation and/or organization system can

be maximally appropriate and useful for the individuals in some culture only if it is based on the assumptions, values and predispositions of that same culture."

The creation of new terms required the definition of a single macrostructure of thematic categories in the UNESP Language to guide the inclusion and classification of terms in order to compose the hierarchical structure of superordinate and subordinate terms. From this perspective, the present paper reports on the experience of the UNESP Language Group in the area of knowledge organization and representation aimed to the construction of the hierarchical structure of an indexing language for the ATHENA online catalog of the UNESP libraries.

The construction of hierarchical structures began by defining the categories and subcategories that form the indexing language macrostructure by using the macrostructure parameters of the *Library of Congress Subject Headings* (*LCSH*), the Brazilian National Library Terminology as well as the Vocabulary of the University of São Paulo libraries' system (VocaUSP). Although VocaUSP and *LCSH* have different terminological foundations, as it is also the case of the LCARB and the TBN, VocaUSP was selected, because its development was based on the terminology of the knowledge areas offered at the University. It was possible to access the *LCSH* macrostructure via Classification Web, a software available through the Library of Congress Web-based subscription service. Moreover, the USP and the National Library websites provided free access to the two other languages. The results enabled the assessment of the hierarchical structures of the languages used in the organization of the superordinate and subordinate terms, which contributed to the systematization of operational procedures of the indexing language manual for online catalogs of libraries.

2.0 Theoretical framework

In relation to indexing languages, the documentary representation, obtained through the documentary analysis process by using theoretical resources from documentary linguistics, function as a commutative code among the different language perspectives involved in the documentary system: user, indexer and system. Documentary representation can be observed from two perspectives that are defined according to the methodology and the object to which they apply: on the one hand, the descriptive representation, which aims to promote physical access to the document; on the other hand, subject representation, whose purpose is to enable access to the informational content of the document. The latter process is generically called "subject approach of information" (Fosket 1996; Guimarães 2009) and can be understood from different theoretical perspectives (Table 1).

Framework	Origin	Assumption	Precursors
Analyse Documentaire	French Early 60's	Focused on the development of theoretical-methodological referentials for the "subject approach to information"	J.C. Gardin M. Coyaud J. Chaumier
Indexing	English-UK Mid 20th. century	Development of tools, such as thesauri.	A. C. Foskett F. W. Lancaster B. Vickery
Subject Cataloging	American-USA End of the 19th. century	Development of products, such as catalogues.	C.A. Cutter E. J. Coates

Table 1. Theoretical frameworks of the subject approach to information (elaborated by the authors following Guimarães 2009).

The broader term "the subject approach to information" encompasses more specific concepts that give rise to distinct theoretical conceptions, namely, "documentary analysis," "subject indexing" and "subject cataloging." According to Table 1, organized from the systematization proposed by Guimarães (2009), these concepts, not rarely taken as synonyms, have different origins, assumptions and precursors and, therefore, are related to different theoretical frameworks. Nevertheless, they have the same purpose: representation for information retrieval.

Studies on subject cataloging, classification, subject indexing and documentary analysis form the theoretical and practical foundations of the subject approach to information and, consequently, they are a substantial basis for organizing information (Café and Sales 2010). From the point of view of their macrostructure, most indexing languages are organized into hierarchical structures, which comply with both culturally and ontologically determined concepts.

The organization of classes in classification systems, such as the *Dewey Decimal Classification* (*DDC*), is governed by the "ordering" principles of science division according to the social organization at the time it came to being and in the various moments in which they were updated. The greater the hierarchical structural rigidity of the system, the greater its probability is of admitting only concepts in consonance with its ordering structural principle, something that Svenonius (2000, 136) called "structural warrant. In thesauri, concepts, particularly in terms of categories, are also structured on a hierarchical basis. In this case, hi-

erarchies are constructed in accordance with the "theoretical-conceptual framework of specific domains to determine sets of terms of the nuclear domain, represented by the specialization area, and by the peripheral domains or complementary areas in order to meet the objective needs of the system at issue" (Cintra et al. 2002, 41).

As for the distinction between "alphabetic" and "classificatory" languages, Svenonius (2000) points out that notation is the most obvious differential; while the former uses verbal expressions for the construction of representations, the latter uses notations. Other distinctions indicated by this author refer to the different vocabulary tools used (thesaurus and subject authority lists vs. classification schemes) and to how concepts are ordered and presented (alphabetical ordering vs. systematic ordering). The distinction has more didactic than really enlightening effects, since both instruments are presented alphabetically as well as systematically. As Svenonius (2000, 128) summarizes, "classification schemes require alphabetical indexes, and alphabetical thesauri can be enhanced by displays of terms in hierarchical order."

Indexing languages are approached from a systemic perspective underlying the idea of the semantic structuring that will define the relationships between concepts. It necessarily implies the notion of categorization *lato sensu*. Categorization defines more specific domains in which the meaning of the terms tends to be more precise. Thus, in a dialectic process, the location of a defined term makes it possible, by understanding its categorial structure, to comprehend in which facet its application occurs. On the other hand, the location of a given category allows one to understand by means of classes, concepts and facets that it groups their coverage extension in the notional system.

Chan (2007) discusses the rapid changes brought about by the recent growth of networks, and stresses the imperative need for categorization to ensure an efficient information retrieval through indexing languages (31):

> The reason for the turn toward more systematic subject control lies in the fact that subject categorization defines narrower domains within which term searching can be carried out more efficiently and thus enables the retrieval of more relevant results. In fact, combining subject categorization with term searching has proven to be an effective and efficient approach in resource discovery and data mining. In this regard, classification or subject categorizing schemes function as information filters, used 'to quickly eliminate large segments of a database from consideration of a query.' Furthermore, classification schemes can also serve as switching mechanisms across different languages and different controlled vocabularies.

Taking into account representation, organization and information retrieval actions in a continuum, regardless of the categorization level, and considering the mediating aspect of indexing languages, a theory is necessary, particularly in the sense that "a theory implies a set of concepts and their relations" (Hjørland 2015, 124). A set of mutually related concepts reveals both their identity individually and, above all, the theoretical position that gives support to their relations. Thus, following the concept of "cultural warrant" (Beghtol 2002), the group of catalogers and researchers believe that the tasks involved in the UNESP language construction require discussions on the identification of concepts and their inter-relations based on theoretical and epistemological aspects.

According to Shera (1965 cited in Jacob 2004, 517), an effective information retrieval "requires an accord between the cognitive organization imposed on information by the individual and the formal organization imposed upon representations by the system." In other words, and considering the ontological foundations of knowledge organization (Gnoli 2011), it is necessary to establish Quine's (1948) "ontological commitment" criterion in broader discussions related to ontologies (Branquinho 2006). In documentary information systems, this accord or commitment rests on three basic assumptions: "that there are certain cognitive structures that can be identified and described;" that "it can be demonstrated that these structures are shared across individuals;" and "that identification of these shared structures will provide the basis for a theory of organization" (Shera 1965 cited in Jacob 2004, 517).

Studies on the indexing languages used in libraries reveal that, in principle, the adoption of subject headings lists associated with cooperative cataloging has disseminated subject cataloging principles. With the theoretical-methodological and normative studies on the construction of indexing languages, the development of these lists has also occurred. Thus, they evolved from originally pre-coordinated instruments to languages with more complex structures. In this evolutionary line, it is worth observing the incorporation of thesaurus characteristics, such as the displacement of the focus from the subject to the term/concept arranged in a structure of relationships that include equivalence, hierarchical and associative relations. It is also worth noting the adoption of the standardization of international norms and a greater concern with vocabulary control.

This project included a questionnaire survey of Brazilian university libraries to gather information on their use of indexing languages. Data were obtained from forty-six libraries, out of which twenty reported using indexing languages while the other libraries adopt unspecified vocabularies. Analysis of the responses indicated that the TBN, *LCSH* and LCARB languages exhibit the three highest frequencies and are used by eight, eight and six libraries, re-

spectively (Fujita and Santos 2016a). It is worth mentioning that both TBN and LCARB are versions of *LCSH*. This result indicates that the use of the *LCSH* language and the principles underlying its construction, the development of its concepts and relationships have been widely disseminated in Brazilian university libraries.

Besides the languages that use the *LCSH* as a primary source**,** the Controlled Vocabulary of the Library System of the University of São Paulo-SIBI/USP (VocaUSP) is used as a language constructed according to an orientation that differs from the other languages, as it can be observed from its hierarchical structure. This structure takes into account the integrated systems context of university libraries and the Brazilian academic culture in its system. The SIBI/USP (VocaUSP) is available for external consultation by users and librarians; however, it does not offer the possibility of importing terms automatically, which hinders its adoption by other libraries.

3.0 The development of hierarchical structure

Hierarchical structures are developed in three stages: identification of the knowledge areas of the university; identification and selection of the indexing language macrostructures; and formation of the hierarchical structures of subject areas with the use of the languages.

The construction of the macrostructure began with a discussion about the division of the work among the members of the Language Group and the criteria to be followed. Macrostructure is understood as a set of hierarchical structures responsible for the vocabulary classification that covers the knowledge areas of UNESP undergraduate and graduate courses as well as teaching, research and community extension activities. Each hierarchical structure is formed by categories and subcategories that correspond to a knowledge area and the vocabulary terms; their syntactic and semantic relations represent knowledge objects.

Ordering categories by the names of course disciplines was one of the criteria established to ease the identification of the knowledge area and its corresponding vocabulary by users and catalogers, as well as to make it possible to visualize the knowledge organization in university libraries. In this way, it will be possible to consult the UNESP Language by its macrostructure, by a systematic list as well as by an alphabetical list. The UNESP offers undergraduate and graduate programs in fifty-six knowledge areas that have been organized according to the classification adopted by Brazilian government agencies, such as the National Council for Scientific and Technological Development (CNPq) and Coordination for the Improvement of Higher Education Personnel (CAPES). These agencies, whose main role is to promote scientific and technological research, were used as parameters to define the knowledge areas that constitute the hierarchical structure of the UNESP language, as shown in Table 2.

Large knowledge areas (7)	**Knowledge areas (56)**
Engineering	Engineering, Architecture, Industrial Design, Cartography (4)
Humanities	Languages & Literature, Pedagogy, History, Geography, Philosophy, Psychology, Translation and Interpreting, Journalism, Broadcasting, Public Relations, Visual Arts, Performing Arts and Music (13)
Applied Social Sciences	Social Sciences, Law, Social Work, Librarianship, Archivology, Administration, Economics, International Relations and Tourism (9)
Biological and Environmental Sciences	Biology, Ecology, Biotechnology, Marine Biology, Coastal Management, Environmental Engineering and Environmental Chemistry (7)
Pure Sciences	Chemistry, Physics, Mathematics, Statistics, Computer Science, Systems Analysis (6)
Agronomical Sciences	Agronomy, Zootechny, Veterinary Medicine, Forestry Engineering, Wood Industry Engineering and Geology (6)
Health Sciences	Medicine, Dentistry, Biomedicine, Medical Physics, Speech Therapy, Physiotherapy, Occupational Therapy, Nursing, Pharmacy, Physical Education and Nutrition

Table 2. Division by knowledge areas.

The work was divided among catalogers and researchers and it was based on their experience and familiarity with knowledge areas and proximity to facilitate contacts with specialists whenever it was necessary in order to resolve doubts about the development of hierarchical structures. Thus, catalogers of libraries with engineering collections, for example, were in charge of developing the hierarchical structures of this knowledge area.

The second stage consisted of two phases: first, search for the identification of the hierarchical structures of the indexing languages: Subject Terminology of the National Library (TBN), *Library of Congress Subject Headings* (*LCSH*) and Controlled Vocabulary of the University of São Paulo (VocaUSP); and the second phase, selection of indexing languages and elaboration of the construction methodology.

The search for the macrostructure identification of the categories started with TBN, and *LCSH* was used later on. The TBN indexing language is provided by the National Library of Brazil and, like LCARB, it adopts the *LCSH* terminology in the construction and updating of the lan-

guage (Grings 2016). In a meeting with the head of the team responsible for the TBN construction, it was confirmed that this language does not systematically have a separate hierarchical structure, and that it was necessary to search and identify it by using its descriptors. The search for the *LCSH* macrostructure indicated that it is not apparent in the vocabulary, just as in the TBN; moreover, a classificatory structure is adopted for the definition of its hierarchical structure available on the National Library of Brazil website through the Library of Congress classification web service to which a subscription was entered for the development of the work.

For the selection of indexing languages and elaboration of a methodology for the construction of the hierarchical structures of the subject areas, the UNESP Language Group carried out an exploratory research on the hierarchical structure of indexing languages. The bottom-up committee approach provided by the ANSI/NISO Z39.19 Guidelines (American National Standard / National Information Standards Organization 2005) was adopted for the compilation of controlled vocabularies using the integrated methodological model. The hierarchies of *LCSH*, TBN, "Subject Headings List of the BIBLIODATA (LCARB) Network" and VocaUSP languages in the areas of physics and mathematics were then compared (Fujita and Santos 2016b).

The results confirmed that the use of the LCARB was not feasible, because it lacks hierarchical relationships and the viability of the other languages regarding the use of its hierarchical structures, and no procedures for the development of hierarchical structures of the subject areas. The research also defined the use of the hierarchical structure of categories and subcategories of the VocaUSP, since it is more similar to the structure of the knowledge organization in UNESP libraries. Although VocaUSP lacks the same terminological basis of the *LCSH*, as it is the case of LCARB and TBN, it was adopted mainly because of the cultural warrant of its knowledge areas. Furthermore, VocaUSP macrostructure is visible and available for consultation on the indexing language website.

A complementary consultation of the specialized indexing languages *Medical Subject Headings* (*MeSH*) and Health Sciences Descriptors (DeCS), which have the same linguistic and structural basis, was maintained for the Health Sciences areas. DeCS is financially supported by the Ministry of Health in Brazil and is annually updated by the Virtual Health Library (BVS) through MeSH for the indexing and retrieval of Brazilian scientific literature.

The third stage consisted of building hierarchical structures of the subject areas by using the languages. This process began by defining the categories and subcategories that form the indexing language macrostructure using those of *LCSH*, TBN and VocaUSPas parameters.

A summarized manual for using the Library Congress Classification web service was developed following the tutorial available on https://classificationweb.net/. This manual contains procedures for making queries about categories (classes and subclasses), hierarchical relationships between terms, subdivisions, notes, classification numbers, association of terms with *DDC* numbers as well as authority record search for importing from the ALEPH software used by the university libraries' network in the construction of the online catalog.

The creation and construction of the hierarchical structures were performed by using the "grouping" feature in the Excel spreadsheet with the following procedures:

- To use the feature "grouping:"
  - Include line (number of lines referring to the number of subordinate terms);
  - Paste the terms in the next column (next hierarchical level);
  - In the Excel header, use the "grouping" feature in the "Data" tab.
- To create the hierarchical structure for each area in an Excel worksheet using the "grouping" feature. Each column will contain a hierarchical level: column B will be the first level (categories), column C the second level, and column D the third level, and so on.
- To create hierarchy of terms, use the indexing languages hierarchical structures in the following sequence: first, VocaUSP; secondly, the analysis of the *LCSH* and, if necessary, the TBN.
- Use color-coding to identify the indexing languages: maintain the same color standard for each language (blue for *LCSH*, red for TBN and green for VocaUSP); when the term appears in more than one language, place it side by side in the same cell, each term with its respective color, according to Figure 1.

SPACE
LAW

BRAZILIAN TELECOMMUNICATIONS CODE

MILITARY SPACE LAW

SPACE LAW—Space law—Space Law

Figure 1. Example of using color-coding.

4.0 Results and discussion

Throughout the development of the macrostructure construction process, the improvements and difficulties observed by monitoring the stages were discussed among the members of the Language Group. To this end, the Language Group has held regular meetings to discuss doubts

and solutions as the hierarchical structures were elaborated. An archive was created for sharing these occurrences and it was accessible to all members of the group.

So far, the creation of eleven hierarchical structures among the fifty-six areas of knowledge shown in Table 1 has been accomplished. They are: geography and cartography, design, architecture, engineering, dentistry, physics, history, international relations, public relations, tourism and librarianship. Currently, eight hierarchical structures in the areas of social work, law, chemistry, literature, arts, psychology, medicine and agronomy are in progress.

Some of these are large areas, as it is the case of engineering, which includes thirteen subareas: civil engineering, aircraft engineering, mining and petroleum engineering, production engineering, electrical engineering, geothermal engineering, mechanical engineering, metallurgical engineering, naval and ocean engineering, chemical engineering, environmental engineering, aerospace engineering and fishing engineering. Each of these subareas has its hierarchical specificities and subdivisions, as it can be seen in the example of chemical engineering, which is considered as an interdisciplinary area, because it is composed of two disciplines, engineering and chemistry. The same occurs in fishing engineering and environmental engineering.

Still in the area of engineering, *LCSH* is used in seven subareas in combination with VocaUSP at the first hierarchical level (civil engineering, production engineering, electrical engineering, mechanical engineering, chemical engineering, environmental engineering and aerospace engineering) and one subarea without combination (geothermal engineering). VocaUSP does not have a combination at the first hierarchical level with *LCSH* in five areas (aircraft engineering, mining and petroleum engineering, metallurgical engineering, naval and ocean engineering and fishing engineering). However, combinations of the three languages are observed at the second and third hierarchical levels.

When a combination of two or three languages could not be made for an area or subarea, it was decided that the hierarchical structure of the language in which the area exists would prevail and would be complemented with glossaries, Wikipedia and consultation with specialists. The combination of geography and cartography into a single hierarchical structure is an example of the kinds of decision-making required throughout the Language Group's task: to maintain geography as an autonomous subject area, because at UNESP it is an undergraduate course that does not require a geography course; or to obey the hierarchical subordination of cartography to geography, as in the VocaUSP language, or to choose mathematical geography subordinated to geography according to the *LCSH* language.

Mining and petroleum engineering

- Mining Engineering / Mining engineering
  - Borehole mining
  - Boring
  - Blasting / Blasting
  - Electricity in mining
  - Petroleum engineering / Petroleum engineering
  - Gas engineering
  - Ground control (Mining)
  - Hydraulic mining
  - Mine lighting
  - In situ processing (Mining)
  - Lasers in mining
- Petroleum Engineering / Petroleum engineering
  - Formation assessment
  - Petroleum classification
  - Petroleum completion
  - Petroleum composition
  - Petroleum constituents
  - Electricity in petroleum engineering
  - Well elevation

Figure 2. Example of the hierarchical structures of mining and petroleum engineering, which only corresponds to the second and third levels of the languages.

Hierarchical subordination of languages with preference for using VocaUSP was considered the best option, because it is the language that most closely approximates to the terminological contents of the UNESP language semantic field. This decision defines an important operational procedure in relation to the hierarchical subordination options made available by the languages. In other words, preserving the autonomy of the subject area was a preferable option for academic reasons. However, subordination should be maintained if, in the analysis of the hi-

erarchical structures of the languages, the area is subordinate in more than one language, mainly in VocaUSP, since it is built according to the academic culture.

For the health sciences areas, *Medical Subject Headings* (*MeSH*) and Health Sciences Descriptors (DeCS) were adopted as the fourth languages for hierarchical structures combination. This decision was made in order to deal with the specificity of other areas, such as medicine and dentistry that also have several subareas. Therefore, *MeSH* has been used to combine the hierarchical structures of medicine (still in progress). On the other hand, DeCS allowed the determination of hierarchical structures in sixty-seven subareas.

As for names of hierarchies that could not be combined or correlated with VocaUSP and *LCSH*, for instance "equipment industry" and "cruise lines," which are related to tourism, it was necessary to do a research on Wikipedia, in glossaries, thesauri, etc., in order to check whether those terms could be embedded in the existing levels. Other terms, such as "vehicle rental business" can also be placed in another hierarchical structure. This term is used in the VocaUSP language in the category "tourism" and subcategory "tourism industry." However, the preferred term in the *LCSH* is "leasing and rental services" that is placed in "service industries" category. The choice was for the term "vehicle rental business" in the VocaUSP, because it is more adequate to the knowledge area of tourism at UNESP.

TOURISM

TOURISM INDUSTRY
TOURISM SERVICES COMPANIES
TOURISM EQUIPMENT RENTAL COMPANIES
VEHICLE RENTAL COMPANIES
EVENT ORGANIZATION COMPANIES
ENTERTAINMENT SERVICES
TOURISM INFORMATION SERVICES
Equipment industry / Equipment industry (included by proximity)
Cruise lines / Cruise lines (included by proximity)

Figure 3. Example of the hierarchical structure of tourism.

It was also decided that the names of categories and subcategories without any correlations would be evaluated considering their meaningful proximity to names that existed in other categories and subcategories the area of tourism, for example, the terms "garden tours," "dark tourism," "lake tourism," which are at the first level of the *LCSH*, should be included at the VocaUSP level "types of tourism." However, very specific names, such as "indian tourism" or composite names linking two or more subjects, for example, "architecture and tourism," should not be included in the macrostructure, but they should be placed in a separate list of "non-used names." Other composite names that have qualifiers, such as "woman in tourism," "woman in physics" should not be included in the macrostructure either.

Non-used terms
Architecture and tourism / Architecture and tourism
Culture and tourism / Culture and tourism
Holocaust memorial tours
Music and tourism / Music and tourism
Sports and tourism / Sports and tourism
Women in tourism
Indian tourism
Customs administration and tourism

Figure 4. Example of a list of non-used terms in the tourism area.

During the combination process of indexing languages for hierarchical structures construction, the occurrence of polyhierarchical terms, that is, terms subordinated to more than one generic term, was observed. Table 3 illustrates this case: the term "courts" is subordinated to four hierarchies in *LCSH* and in TBN, respectively:

Courts (LCSH)	**Courts *(TBN)***	**Courts (VocaUSP)**
BT Dispute resolution (Law) Judicial districts Law Procedure (Law)	TG Districts TG Law TG Procedural Law TG Dispute resolution (Law)	TG Procedural Law

Table 3. Polyhierarchical subordination.

Since VocaUSP does not deal with polyhierarchical subordination, the Language Group decided to construct the hierarchical structures by combining the structures of the three languages starting with VocaUSP, which does not rule out the acceptance of *LCSH* and TBN polyhierarchies. Thus, the term courts, subordinated to four other generic terms in *LCSH* and TBN, was included as a sub-

ordinate term in the four hierarchies. This example (Figure 5) demonstrates the group's decision to accept polyhierarchy in order to reduce semantic restrictions and, consequently, to enlarge the more diversified and favorable semantic field.

Figure 5. Polyhierarchy of the term courts.

5.0 Conclusions

The catalogers and researchers of the UNESP Language Group believe that the construction of hierarchical structures is a valid and relevant approach to a great variety of subject areas. Throughout the development of this challenging work, the team had the opportunity to confirm that an immense body of knowledge can be organized by using indexing languages to combine hierarchical structures. Several decisions had to be made in order to ensure the systematization of procedures, as illustrated in the previous section. The construction of the UNESP language is in progress and, given the diversity of knowledge areas, future issues will be analyzed for the proposition of other systematized decisions.

The work is a complex enterprise, not only due to the great variety of subject areas, but also to the fact that it is a teamwork. In this sense, it is recommended that the members be experienced in various knowledge areas and, above all, be in contact with experts that can contribute with the history and significant peculiarities of the specialized fields. The example of the subordination of cartography to geography gives an idea of how important this recommendation is in many other ways: the cataloger in charge of this task had to be very tactful when he consulted with specialists of these areas. They told the cataloger about their efforts to make cartography an autonomous academic discipline that would be relevant to other areas, such as engineering and architecture, not only to geography.

The fusion of different indexing languages implies matching different contexts that are reflected in their hierarchical and conceptual structures. As observed, the case of the hierarchies in geography and cartography may be problematic in the compatibilization process. In order to deal with demands such as this one, specialized vocabulary studies are being conducted with the participation of experts. It is also recommended that the UNESP Language Group carry out discussions on theories, epistemology and cultural warrant to find a possible solution for these problems.

Another equally relevant decision was related to the choice of the VocaUSP as the initial language for the first level subareas that would be combined with the other languages; VocaUSP is the closest to and most familiar language in knowledge organization in the São Paulo state universities, which warrants the cultural aspect of knowledge areas. However, the combination with the other indexing languages should favor visibility of the knowledge area in international environments, which is indispensable for the dissemination of the scientific literature produced in Brazil.

In addition to this, the combination of more than one language in the construction of hierarchies was viewed by the group as an advantage rather than a challenge. As an example, it was possible to create the hierarchy of a knowledge area using just one of the languages without correlation with the others. In this case, the recommendation is to construct hierarchical structures by combining several languages even if they do not have the same hierarchical and vocabulary parameter, as in VocaUSP.

The results that have been accomplished so far can have a broader reach in Brazil if the UNESP and USP language groups work together with the same purpose, that is, to make their indexing languages available to other systems of Brazilian university libraries with the elaboration of an indexing language manual containing the final systematization of operational procedures.

References

Bates, Marcia J. 1989. "Rethinking Subject Cataloguing in the Online Environment." *Library Resources and Technical Services* 33: 400-12.

Beghtol, Clare. 2002. "A Proposed Ethical Warrant for Global Knowledge Representation and Organization Systems." *Journal of Documentation* 58: 507-32.

Branquinho, João. 2006. "Compromisso ontológico." In *Enciclopédia de termos lógico-filosóficos*, ed. João Branquinho, Desidério Murcho, and Nelson Gonçalves Gomes. São Paulo: Martins Fontes, 149-55.

Café, Lígia Maria Arruda and Rodrigo Sales. 2010. "Organização da informação: Conceitos básicos e breve fundamentação teórica." In *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento—eroic,* org. Jaime Robredo and Marisa Brascher. Brasília: IBICT, 115-29.

Campello, Bernadete Sampos. 2006. *Introdução ao controle bibliográfico,* 2nd ed. Brasília: Briquet de Lemos.

Chan, Lois Mai. 2007. *Cataloging and Classification: An Introduction,* 3rd ed. Lanham, MD: Scarecrow.

Cintra, Ana Maria Marques, Maria de Fátima Gonçalves Moreira Tálamo Marilda, Lopes Ginez de Lara, and Nair Yumiko Kobashi. 2002. *Para entender as linguagens documentárias*, 2. ed. rev. e ampl. Coleção Palavra-chave 4. São Paulo: Polis.

Cleveland, Donald B. and Ana D. Cleveland. 2001. *Introduction to Indexing and Abstracting.* 3rd ed. Englewood, Colo.: Libraries Unlimited.

Foskett, A. C. 1969. *The Subject Approach to Information*. London: Bingley.

Fujita, Mariângela Spotti Lopes and Luciana Beatriz Piovezan dos Santos. 2016a. "A estrutura lógico-hierárquica de linguagens de indexação utilizadas por bibliotecas universitárias." *Scire* 22, no. 2: 37-46.

Fujita, Mariângela Spotti Lopes and Luciana Beatriz Piovezan dos Santos. 2016b. Política de indexação em bibliotecas universitárias: Estudo diagnóstico e analítico com pesquisa participante." *Transinformação* 28, no. 1: 59-76.

Fundação Getúlio Vargas. 1995. *Manual de cabeçalhos de assunto: Normas e procedimentos; v. 1.0*. Rio de Janeiro: FGV.

Gnoli, Claudio. 2011. "Ontological Foundations in Knowledge Organization: The Theory of Integrative Levels Applied in Citation Order." *Scire* 17, no. 1: 29-34.

Grings, Luciana. 2016. "Controle de autoridades na Biblioteca Nacional do Brasil: Breve histórico e práticas atuais." *Revista brasileira de biblioteconomia e documentação* 11: 139-54.

Guimarães, José Augusto Chaves. 2009. "Abordagens teóricas de tratamento temático da informação (TTI): Catalogação de assunto, indexação e análise documental." In *Avances y perspectivas en sistemas de información y documentación,* ed. Francisco Javier García Marco. Zaragoza: Ibersid, 105-17.

Hjørland, Birger. 2012. "Is Classification Necessary after Google?" *Journal of Documentation* 68: 299-317.

Hjørland, Birger. 2015. "Theories are Knowledge Organizing Systems (KOS)." *Knowledge Organization* 42: 113-28.

International Organization for Standardization. 2011. *Information and Documentation: Part 1: Thesauri and Interoperability with other Vocabularies*. ISO 25964-1. Geneva: International Organization for Standardization.

Jacob, Elin K. 2004. "Classification and Categorization: A Difference that Makes a Difference." *Library Trends* 52, no. 3: 515-40.

Mazzocchi, Fulvio, 2018. "Knowledge Organization Systems (KOS)." *Knowledge Organization* 45: 54-78. doi:10.5771/0943-7444-2018-1-54

National Information Standards Organization. 2005. *ANSI/NISO Z39.19: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, MD: National Information Standards Organization.

Quine, Willard V. 1948. "On What There Is." *Review of Metaphysics* 2, no. 5: 21-38.

Ribas, Rosane Rodrigues de Barros and Lúcia Silva Parra. 2016. "A construção e atualização da linguagem UNESP." In *Política de indexação para bibliotecas: Elaboração, avaliação e implementação*, org. Mariângela Spotti Lopes Fujita. Marília: Oficina Universitária, 67-87.

Svenonius, Elaine. 2000. *The Intellectual Foundation of Information Organization*. Cambridge, MA: MIT Press.

Universidade Estadual Paulista. Coordenadoria Geral de Bibliotecas UNESP. CGB). 2002. *Padrão de qualidade de registros bibliográficos da UNESP, ed.* Cássia Adriana de Sant'ana Gatti [et al.]. Publicações técnicas 4-5. Marília: UNESP; Coordenadoria Geral de Bibliotecas.

Universidade Estadual Paulista "Júlio de Mesquita Filho." 2014. *Manual de política de indexação para as bibliotecas universitárias da UNESP*. São Paulo: UNESP. http://www.biblioteca.UNESP.br/portal/arquivos/manual-politica-indexacao