Suellen Oliveira Milani  (São Paulo State University, Marília)  
José Augusto Chaves Guimarães  (São Paulo State University, Marília)  

Bias in the indexing languages  
Theoretical approaches about feminine issues  

Abstract  
The process of knowledge representation as well as its procedures or tools and its products are not neutral in terms of values; instead they imply moral values. In this context, bias in representation related to prejudice and discrimination, to gender issues, to dicotomic categorization in classification systems or in thesauri and to lack of cultural warrant may arise. Concerning the problem of bias in indexing languages, starting from the initial theoretical reflexions of Brey (1999), Berman (1993), Olson (1998; 2002), López-Huertas Pérez & Torres Ramírez (2005), Guimarães (2006), Hjørland (2008) and Milani et al. (2009), the proposal is to present a preliminary categorization aiming at facilitating the identification of bias concerning feminine issues in indexing languages, to offer a contribution to the theoretical universe of the specific questions of knowledge organization and to present a theme to be discussed by educators and professionals in the areas of cataloging, classification and indexing. If in a society which intends to be politically correct, social attitudes towards stigmatized citizens should be modified, then, the universe of indexing languages, taken as tools of knowledge representation, is a fertile field to sow this reflexion.  

1: Aim and scope of the study  
Knowledge representation is a process of constant decision making and it must search for reliability in order to avoid favoring certain interests and values rather than others, which would result in biased representation in which the surrogate of knowledge is built on bias. Taking this issue into account, methodologies should be developed aiming not only at avoiding bias but also at detecting it in knowledge representation, once the unavoidable subjective influence of the information professional makes it impossible to be a neutral process.  

Considering that bias in knowledge representation “besides violating individual rights, also inhibits the users, pushing them aside, due to an absolute lack of reliability and even to an identity problem towards the representation criteria” (Guimarães 2006, 238), it’s expected that the information professional reflect on the non-neutral nature of the representation process, as well as on the five dimensions of ethical commitments related to it, as follows: commitment with the user, the organization, the information, the profession and with himself/herself as a professional and a citizen (Guimarães 2000). Specifically related to minority groups, the aim of the study was to elaborate a preliminary categorization of bias concerning issues related to women as a contribution to the theoretical universe of social questions in knowledge organization (KO).  

2: Method  
The present documentary research starts from studies on bias in knowledge representation carried out by Brey (1999), Berman (1993), Olson (1998; 2002), López-Huertas Pérez & Torres Ramírez (2005), Guimarães (2006), Hjørland (2008) and Milani et al. (2009) with the aim of proposing a preliminary categorization of bias which may motivate discussions and thoughts in the contexts of education and professional practice of cataloging, classification and indexing.
3: Outline

The international literature of the last decade points out that bias concerning feminine issues in knowledge representation is related to problems such as: prejudice and discrimination (Berman 1993); gender questions (Olson 2002; López-Huertas Pérez & Torres Ramírez 2005); dicotomic categorizations in classification systems or thesauri (Guimarães 2006); and lack of cultural warrant (Beghtol 2002). In this way, “when naming is biassed, when it leaves out diverse features, it disenfranchises groups and topics outside of an accepted norm. Therefore, I expect the greatest problems of subject access to occur in the representation of groups and topics that are marginalized from that norm or that seek to change it.” (Olson 2002, 15).

Bias can be conceived as “a negatively loaded word, as something to be avoided or minimized” (Hjørland 2008, 256), and can be observed in knowledge representation usually related to gender, sexuality, race, age, ability, ethnicity, language and religion matters, “described as limits to the representation of diversity and to effective library service for diverse populations” (Olson 2002, 7).

In this context, there are some specific studies on gender and discrimination of minority groups in the Universal Decimal Classification (Santos et al. 1999); on woman representation in the Universal Decimal Classification (Morán Suárez & Rodríguez Bravo 2001); on KO and representation in gender issues in Internet (López-Huertas & Barité Roqueta 2002); on analysis of terminological representation in four thesauri on women’s issues (López-Huertas et al. 2004); on comparison of the development of the subject field of gender studies in Spain and Uruguay (López-Huertas 2008); and on the feminine issue in lists of subject headings and in thesauri (Rodríguez Bravo 2007). Such studies highlight and exemplify gender discrimination and point out that it is necessary to think of an ideal model of KO for each domain (López-Huertas et al. 2004, 38), once different domains in different cultures have specific informational needs and require different kinds of organization and different ways to get information (Beghtol 2002).

Another problem discussed by the authors is the non-uniformity of thematic representation and the emphasis on representation of gender aspects only related to sexuality (specially to homosexuality), family life (including maternity) and personal relations.

Concerning indexing languages, Caro Castro & San Segundo Manuel (1999, 105) suggest the inclusion of minority groups in accordance to its condition, be it “ethical, linguistic, political, religious, national, ideological, social or economic condition”, emphasizing also the need of adequacy (place, society, user and moment), of up to dating and of defining the elaboration criteria. In this sense, Olson (2002, 6) points out that the limits of the system (and consequently its inclusions and exclusions) are defined by the selection of concepts to be named and “the selection of terms of these concepts often introduces blatant biases or, more commonly, subtle, insidious marginalizations”.

This problem of bias in the feminine representation in traditional indexing languages has its origins, as Rodríguez Bravo (2007) points out, in the rare presence of women in such languages, due to the use of masculine as universals, or to the fact that the “knowledge representation discourse have maintained feminine sexist stereotypes, presenting an anacronic image of women”.

Specially concerning questions of gender, Olson (2002) believes that in aspects related to the representation of women and for women, standardizations present three
frequent problems: women have been treated as exceptions to masculine norms; women issues have been treated isolated, separated from knowledge as a whole; or these issues have been completely omitted.

López-Huertas Pérez & Torres Ramírez (2005) highlight and specify the types of knowledge representation bias as reflects of the own human language bias: the use of general terms in the masculine form; consider woman as the subject of anticonception; the use of masculine terms in the plural form to refer to both men and women; the use of nominals has shown a tendency to create feminine descriptors and to omit their correspondent masculine descriptors.

Problems which involve aspects of moral nature are pointed out by Vieira Júnior (2005, 11) taking into account that society, aiming at establishing social norms for its citizens to live by, makes use of some controlling mechanisms such as: the state; the family; the Church; the school; the library; and the media.

In thesauri, as more specialized indexing languages, “there is more simetry in the treatment of feminine and masculine. The lists of headings dedicate a special space to feminine issues and present a great number of headings as to allow representing a feminine theme which otherwise might be hidden in the vast vocabulary of the list”. (Rodríguez Bravo 2007)

There is a need of cultural warrant in the representation of any specific domain (and in this case of the domain of feminine issues), which implies the need of carrying out studies adopting the methodological approach of domain analysis (Hjørland 2002) in feminine issues.

Thus, in the case of translations, there are many terminological blanks caused, many times, by cultural blanks, by the lack of knowledge of this domain in different cultures. “The degree of terminological precision can be a reliable indicator of the level of development of this theme in different cultures” (López-Huertas Pérez & Torres Ramírez 2005).

On the process of creating an indexing language to base studies on women, one has to take into account how to represent the masculine. López-Huertas Pérez & Torres Ramírez (2005) believe that “it’s an important issue because the way it’s done may be guided by ideologies that may have negative effects on the desired objectivity”. There is a tendency to avoid descriptors which use the word woman/women as the nuclear element. Rodríguez Bravo (2007) presents some suggestions of actions that may minimize the problem:

• insertion of feminine forms without hiding the masculine ones – to make efforts to eliminate sexist stereotypes and masculine terms as false general terms from indexing languages, searching for a balance between the presence of both feminine and masculine;
• insertion of notes which allow the use of the feminine form and/or the use of headings/sintagmatic descriptors to distinguish the masculine form the generic, when there is no other option; and
• insertion of gender qualifiers along with the headings/descriptors, being “(W)” and “(M)”, when the theme does not represent humanity in general.

In this case of the feminine issue, there is a need of constant reflexion, of critical studies of epistemological nature on specific domains of knowledge, to “provide guidelines for selection, organisation and retrieval of information and provide the
highest level of generality about information needs and relevance criteria that can be obtained” (Hjørland 2002, 440).

According to Milani et al. (2009), once an information system reveals a type of articulation of a certain domain, special attention should be given to resources such as authorized/non-authorized descriptors/subject headings (USE, UF, “see”), related terms (RT, “see also”) and scope notes (SN), because they transcend the technical dimension and play an important social role: since knowledge representation, as an activity which aims at making information available to all and any kind of user, it must ensure that every user, without distinction, is able to recognize himself/herself represented in the system. If this does not occur, the system will have failed; the system will not have fulfilled the expectations of the user; it will have violated the user’s rights and moral values, pushing him/her aside.

As a consequence, bestowing a naming to information is a quite complex representation knowledge activity since the act of labeling leads to creating an identity (Olson 2002). In this sense, information professionals should consider – and be careful with – their effective “power to name” (Olson 2002) when they decide what to represent and what to leave unnamed, because a broad and deep set of moral values (as well as problems derived from their negation) are effectively (and sometimes dangerously!) involved (Guimarães et al. 2008).

Referring to moral questions, Guimarães et al. (2008) presented, categorized and discussed a set of moral values (e.g.: respect to privacy, copyright, accessibility, freedom, safety, equity, diversity, risk minimization, competence-skill, efficiency, flexibility, reliability, professional recognition, up-to-dating, autonomy, power awareness, cooperation, precision, recall, cultural warrant, exhaustivity, consistency, usability and hospitality) and problems (e.g.: digital segregation, pornography, discharge of electronic garbage, professionals being replaced by technology, violence, surveillance, censorship, lack of cultural warrant, negligence, informational directness, professional inefficiency, misrepresentation, racism, ambiguity, marginalization, impartiality or neutrality belief, idiosyncrasy, inaccessibility to information, biased terminology and inadequate translations) which deeply affect KO and representation activities and can contribute to avoiding (or to maintaining) biases.

Therefore, if in a society which intends to be politically correct, social attitudes towards stigmatized citizens should be modified, then, the universe of indexing languages, taken as tools of knowledge representation, is a fertile field to sow this reflexion.

References


Vieira Junior N.C., 2005, Uso de informação como forma de amenizar os preconceitos acerca da homossexualidade, Universidade Estadual de Londrina.