

●王知津 黄欣

主题标引过程的符号语言学分析

——作为无限指号过程的主题标引过程

摘要 主题标引过程由4个要素(文献、主题、主题描述和主题款目)和3个步骤(文献分析、主题描述和主题分析)组成。按照无限指号过程的思想,可以把主题标引过程中的每个要素都看成一个符号,每个步骤都起到解释行为的作用,而这个行为联系着连续中的符号。图4。参考文献15。

关键词 主题标引 符号语言学 无限指号过程

分类号 G254

ABSTRACT Subject indexing is composed of four elements (document, subject, subject description and subject entry) and three steps (document analysis, subject description and subject analysis). We can regard every element in the subject indexing as a symbol according to the theory of infinite semiosis. 4 figs. 15 refs.

KEY WORDS Subject indexing. Semiotics. Semiosis.

CLASS NUMBER G254

1 引言

为检索而表示文献主题内容的主要问题与含义和语言有关,更具体地讲,是怎样用一些单词或符号来表示一个语句。语言哲学涉及到如何确定和建立含义,以及语言如何表示实体。主题标引过程和语言哲学之间的理解似乎存在着有意义的交叉,因此,本文从语言哲学的观点,探讨主题标引过程。其他人也已经采用了类似的假设。例如,Fairthorne指出:“只有冒着无结果的风险,才能把特定论题看成是孤立的论题,因此,对语言和含义的一些问题有所了解是基本的”^[1]。Blair指出:信息检索研究的中心任务是理解为了有效地检索,应当如何表示文献。这基本上是一个语言和含义的问题。任何文献表示的理论必须以一个清晰的语言和含义的理论为基础”^[2]。在这一方面,本研究认为,主题标引过程由许多步骤组成,这些步骤应当被看做解释。Benediktsson提出了标引过程的解释的本质和需要认识解释的重要性的方针:“任何一种书目描述都被看成是描述性的。当到达解释时,问题在于,描述不应该遵循像教规一样的、使解释成为可能的方法或标准吗?”^[3]本项研究将探讨由以上3人及其他人提出的研究标引和图书情报学的方法。

2 标引过程的步骤划分

在有些文献中,标引过程经常被描绘成包括2个、3个甚至4个步骤的过程,我们分别称之为两步法、三步法和四步法。

两步法(如Benediktsson^[4]; Forhmann^[5])由以下两步组成:第一步确定主题内容;第二步把主题内容转换并表示成标引语言的形式。

三步法(例如Mikda^[6]; ISO^[7]; Farrow^[8]; Taylor^[9]; Petersen^[10])在该过程中增加了一个步骤。第一个步骤仍然是

确定主题,但第二步包括把从步骤一中找出的主题内容重新用较为正规(规范)的语言表示。然后,第三步把较为正规表述的主题进一步转换成某一标引语言的清晰术语。例如:(1)确定文献的主题内容;(2)重新用一种自然语言语句表示主题内容;(3)将主题内容转换为标引语言。

四步法(Langridge^[11]; Chu & O'Brien^[12])的前两步与三步法中的前两个相似。第一步或多或少有些非正式地确定文献的主题内容。在第二步中,标引员较为正式地概括文献的主题内容,通常是用标引员自己的词汇和比较浓缩的语句概括。此后的步骤与三步法不同。把主题内容翻译成标引语言不是由一个步骤而是由两个步骤组成。在第三步中,标引员将句子转化成标引语言使用的词汇。在第四步中,标引员用标引语言构建一个或多个主题款目(以标引词、分类号或标题的形式)并且考虑它们之间的句法和关系。例如:(1)确定文献主题内容;(2)用自然语言语句重新表示主题内容;(3)用标引语言词汇重新表示自然语言语句;(4)将主题内容翻译为标引语言。

应当指出,这里所说的“步骤”的概念主要是与标引过程中的逻辑有联系,而与脑力和体力工作的实际顺序没有必然联系。某些标引员,尤其是搞这项工作的初学者,“借助一系列数字”,在已完成的步骤上打勾,即可完成标引,这或许是个好办法。然而,当取得经验后就不能这样做了。实际上,有经验的标引员和编目员根本意识不到各种步骤的存在,不管假设多少个步骤是最正确的,所有步骤几乎会同时发生。简言之,一个有经验的标引员将在一次复杂活动中完成这个标引过程。Mai探讨了标引员从新手到专家的演变^[13]。但是当分析这个过程时,带着步骤的概念去工作还是有用的,因为将这个过程分成几个部分将允许人们更详细地进行检查。

两步法在主题标引过程中的概念过于简单。实际上,两步法似乎主要用来作为拆分主题标引过程中两个不同活动的手段:确定文献主题内容和把该主题转化成标引语言的术语。它很少用于讨论该过程本身的细节。相反,四步法似乎给该过程的最后部分(把文献主题转换成标引语言术语)带来了不必要的麻烦。四步法将该过程的最后部分分成两步,但这两步之间没有本质上的差别,而仅仅是一般的和特定活动的差别,所以,这种分法似乎没有必要。最后两步中的前者,即把文献主题转换成给定主题存取词汇的语言,而后者仅仅把上一步的结果转换成该系统的标引词或词串(即句法)。所以本文倾向于选用三步法。

3 “四要素,三步骤”

本文的重点不仅仅关注主题标引过程的步骤本身,而是提出这些步骤的强化表演。更确切地说,这里提出的标引过程的观点由 4 个要素和 3 个步骤组成,即一个要素由一个被作用的客体 and 作用于这个客体的一个步骤组成。要素和步骤的顺序如下:要素 1—步骤 1—要素 2—步骤 2—要素 3—步骤 3—要素 4。

第一个要素由被审阅中的文献组成。作为一个被作用的客体,这篇文献是给定的。它的存在使标引过程迅速启动。

第一个步骤也叫文献分析过程,作为这篇文献存在的反映而出现。为了确认它的主题内容,第一步由检查该篇文献的行为(即题名、目次、摘要,如果有的话,还包括书后索引、项目评论等)组成。

第二个要素是第一步骤的产物。它是由一些标引员对文献中主题内容的智力感觉组成。第二要素可以被称为文献的主题,就像它最初存在于标引员头脑中并包括了在该过程中选择出来的大量相对无序的印象、短语、术语等。这些想法是从构成第一步检查过程基础的来源上形成的。

第二个步骤是标引员对第二要素的回应,并命名为主题描述过程。它由试图用语言生成一个文献主题明确表述的行为组成。第一个步骤的产物是大量的相对无序的印象、短语、术语等,它们是在检查文献的过程中收集来的。第二个步骤的产物是给出不同的印象、短语、术语等协商工作的结果,即某种顺序和结构。

第三个要素是第二个步骤的产物。它由或多或少的用语言明确表述的文献主题组成——主题描述。第二个要素主要是由智力产物以及不同候选词、思想、概念等的连续智力标识等组成,这些标识是由标引员在检查文献时收集的。第三个要素表示把所有这些压缩成某些东西的尝试,这些东西以某种相对紧密的方式概括文献的主题。

第三个步骤受到第三个要素存在的提示,也就是说,出现了用语言相对紧密概括的文献主题。这一步骤在此处被命名为主题分析过程,并且由将第三个要素的产物转换成同一事物的规范表达的行为组成,只是此时要根据合适的主题存取系统语言的词语。这意味着,将一个人的语言表述转换成分

类号、主题标目或叙词。当然,在这一活动中,人们必须知道任何系统使用的所有的不同规则、条例、禁止等。

第四个要素,也是这个过程的终点,简单地是第三个步骤的产物。它由来自一个给定系统的完整主题款目组成,这是标引员最终选出并表示文献主题的款目。

现在重新措词,标引过程中的要素和步骤的顺序是:文献(要素 1)—文献分析过程(步骤 1)—主题(要素 2)—主题描述过程(步骤 2)—主题描述(要素 3)—主题分析过程(步骤 3)—主题款目(要素 4)。

4 符号语言学概述

符号语言学一般被定义为符号的研究。符号的研究分为欧洲和美国两大流派:欧洲流派以法国语言学家 Ferdinand de Saussure(1857~1913 年)为代表,通常被称为符号学(semiology)。美国流派以美国科学家和哲学家 Charles Sanders Peirce(1839~1914 年)为代表,通常被称为符号语言学(semiotics 或 semeiotic)。尽管人们试图定义一个统一的符号语言学,但这两大流派还是有区别的。前者的理论是一种如何从词中抽取含义的理论。而后的理论一般是符号如何表示含义,而符号不仅仅是词。

Peirce 把符号定义为 3 个实体之间的联系,这 3 个实体是符号本身、符号的指示物以及从符号得出的含义。他所关心的是如何从符号中得出含义和如何将之转换为其他符号。运用符号的三面的或者三元概念,他定义如下:符号或“表示法”是某一事物,它在某一方面或某一能力上支持某人做某事。也就是说,它指明某人在其头脑中创建了一个等价的符号,或许是一个更发展的符号。它创建的这个符号称之为第一个符号的“解释”。这个符号代表某一事物,即它的“客体”。它不是在各个方面代表这个客体,而是关于某种思想^[14]。

他区别了物理实体(如词)、这些词所指的思想和从这些词中取出的含义。如图 1 所示,把符号的概念表示成一个三角形。

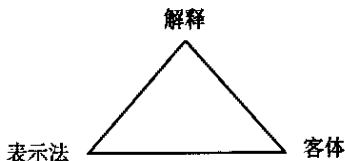


图 1 符号语言学三角形

表示法是表示符号的方法,通常以某种物理实体的形式表示,或者至少以某种形式出现。换句话说,表示法是符号关系的实体,符号关系被感知并通常表示成“符号”。

表示法表示某一客体。但在表示法和客体之间不存在一对一的关系。客体不是某种可辨认的独立于符号而存在的实体。Peirce 对客体做了如下说明:每个客体(一个符号可能会有许多客体)可能都是单一的已知的现存的事物,或

者是先前认为已经存在或期望存在的事物,或者是这类事物的集合,或者已知质量、关系或事实,那个单个客体可能是一个集合或部分的整体,或者它有某种存在的模式,如某一允许的行为,该行为的存在不妨碍它否定同样允许的存在,或者是期望的一般性质的事物,或在某种普通环境下总是不变的某种事物^[15]。

符号只能表示客体并告知客体的有关情况,它不能提供对客体的熟悉或者对客体的认识。因此,客体不是某些存在的客观实体,但可以通过符号知晓和认识。客体是“符号预示熟悉的东西,以便传递有关它的某些进一步信息”。应该把符号看成人们理解符号所必需的背景知识,或者是能被制作出来的关于符号的一系列可能有意义的表述。表示法可能是另外表述或代表某物的任何事项,这就是说,Peirce 的符号概念没有局限在词或语言。因此,正如后面所说的,可以把文献看做是一种符号。

表示法与其客体之间的联系由解释做出,它是符号关系中的第三个实体。解释不是解释符号的人,而是从表示法中产生出来的符号。换句话说,当把表示法看成一个符号时,在表示法的基础上,创造出一个新的和更具发展性的符号。解释这个符号的人在所见到的事物(即表示法)与其背景知识(即客体)之间建立一种联系,从而创造出一种对符号的理解或含义(即解释)。这一过程被称作指号过程(semiosis)(在逻辑学中,指语言或非语言起记号指代作用的过程),即解释符号的行为。

在图2的符号Y-腿模型中,强调了表示法与把解释作为一种指号过程而创造的客体之间的联系作为指号的一个过程。从表示法到客体的连线强调了初始符号(表示法)及其指示物(客体)之间的联系。这两个实体之间的联系是表示法的含义,即表示成解释。这种想法在Y-腿模型中得以强化,但在符号语言学三角形中就不特别清楚。本文使用这两个模型。

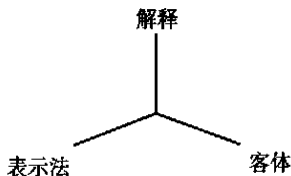


图2 Y-腿模型

Peirce 的符号语言学理论中的一个关键要素是“无限指号过程”的概念,这个概念被看成符号的关联或一个符号产生另一个符号的过程。无限指号过程以指号过程的基本思想为基础:一个符号(b)是在另一个符号(a)的基础上产生的。当一个新符号(c)在第二个符号(b)的基础上产生时,第二个指号过程也就发生了。由于新符号仍会产生更多的符号,这个过程将无限期地继续下去,因此是无限的,故使用“无限指号过程”这个词。

图3表示无限指号过程,该过程中的第一个符号的解释转变成第二个符号的表示法。它们之间存在着某种关系,但每一框中的客体保持了对表示法和解释的独立性。客体将会改变整个过程。无限指号过程中的每个客体关系对于这个符号关系是惟一的。无限指号过程中的单个客体是相互独立的。对于Peirce的符号语言学理论来说,后者至关重要。

图3中的三角形将继续产生新三角形。应当这样来理

解:在它们之前有三角形,在它们之后也有三角形。这意味着,对某一事物的理解通常是以对其他事物的理解为基础的,并且将会继续产生另一种理解。这里阐述两个重要思想:第一,不同的符号关系有不同的实体,每一实体依赖于创造这个解释的人,不同客体之间没有必然联系。第二,接下来,每一表示法以解释为基础,解释又以表示法为基础。

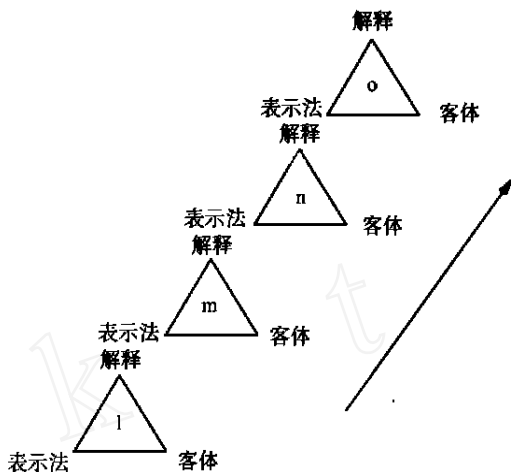


图3 无限指号过程

5 作为无限指号过程的主题标引过程

如上所述,主题标引过程由4个要素(文献、主题、主题描述和主题款目)和3个步骤(文献分析、主题描述和主题分析)组成。这些要素和步骤以如下方式相互联系起来:可按照Peirce的无限指号过程和符号的思想进行解释。下面将对标引过程进行符号语言学分析,特别是简要阐述无限指号过程如何适应上述情况。

按照Peirce的无限指号过程的思想,可对主题标引过程作如下表述:可以把主题标引过程中的每个要素都看成一个符号,每一步骤都起到解释行为的作用,而这个行为联系着连续过程中的符号。

这个过程以最初的符号(文献)开始。标引员最初做出解释行为(第一步),以便最初确定第一个符号(文献)是关于什么的。这个行为的产物是一个新的(第二个)符号(主题)。然后做出一个新的解释行为(第二步),以便把标引员作为主题提出来的东西转变成对于标引更为简明、可以管理的某些东西。这个行为的产物是产生另一个新的(第三个)符号(主题描述)。最后,做出另一个解释行为(第三步),以便使主题描述适合于给定的主题标引系统词汇。接着这个行为又发展到另一个新的(第四个)符号(主题款目)。

当然上述过程可以进一步延伸。例如,用户将会使用索引查看主题款目(符号),在解释行为中,用户把它看做是关于文献的相关内容的表述,尽管在这种情况下相关内容很可能以某种形式与用户首先检索出来的信息有关。用户关于主题款目是什么的结论将会继续产生另一个符号。如此等等。

整个过程如图4所示。应当首先注意的是,图中三角形

被称作 m 、 n 、 o 和 p (而不是 a 、 b 、 c 和 d) , 这是为了强调这样一个事实: 实际上, 主题标引是大量解释过程中的一部分。

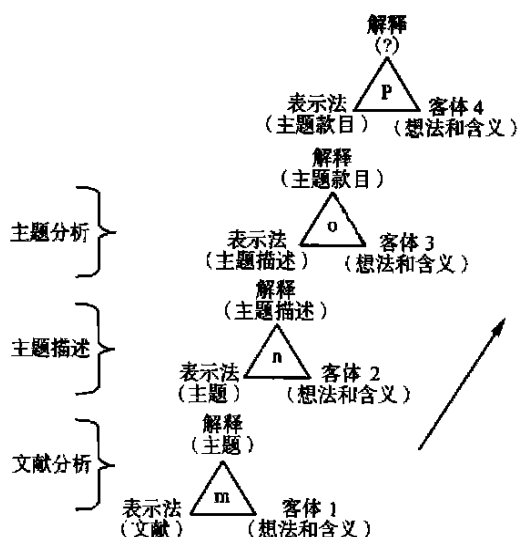


图 4 标引的符号语言学模型

图 4 中要注意的第一个问题是, 在标引员开始主题标引过程之前, 文献以某种论文共同体的形式或科学论文共同体的形式产生。文献的真正创新之处是那部分文献作者和文献作者所参考的那部分文献作者的许多解释行为的结果。文献一旦被完成和出版, 以及对可得文献进行主题标引过程之后, 该文献将有可能被大量的信息用户检索和使用, 而某些文献则在论文共同体之外。那些信息用户查阅目录和关注标引系统中 (在其他文献之间) 表示文献的主题款目, 以及后来的作为整体或部分的文献用户的活动, 同样都是解释行为。无限指号过程 (这里主要限于主题标引过程) 开始于主题标引过程开始之前, 并将在其完成之后继续下去。图 4 简单地表示了一个大型过程的截断部分。

图 4 中要注意的第二个问题是它的三角形布局。每个三角形都是构成无限指号过程的一个要素的符号。符号被定义为三个实体之间的一种关系: 表示法、解释和客体。这种关系构成了符号, 并且其本身就是一个指号或解释过程。换句话说, 每个要素都是一个符号, 符号的解释是一个指号过程。同样地, 每个三角形表示一个指号过程, 三角形左下角是解释 (表示法) 的开始符号, 在它的顶端是从解释行为中新产生的符号, 与表示法 (客体) 相关的想法和含义范围位于右下角。

我们在前面曾经指出, 一个要素由被作用的客体和作用于这个客体的步骤组成。提出这个观点是为了将标引过程中的要素纳入考虑之中。以前的探讨只关注步骤, 忽略了要素在这个过程中的地位。但是, 根据图 4 所解释的观点, 在要素和步骤之间没有精确的分界线。甚至, 要素和步骤重叠成单一的解释行为——指号过程。当标引员作用于一个要素时, 他实际上已经进入了产生下一个要素的步骤。例如, 当标引员查看和作用于文献时, 这个行为实际上是主题标引过程中的第一步, 即文献分析。标引员不能查看或作用于文献, 那么就探讨第一步。不能把要素和步骤分成

两种不同的现象。然而, 为了较好地理解主题标引过程, 应当进一步分别分析要素和步骤, 思考连续指号过程中的各种解释行为的本质以及该过程特殊符号的本质。但应当清楚, 在现实工作中不能保持这种区分。

参考文献

- 1 Fairthorne, R. A. Content analysis, specification, and control. *Annual Review of Information Science and Technology*, 4(1969)
- 2 Blair, D. Language and representation in information, retrieval. New York: Elsevier Science, 1990
- 3,4 Benediktsson, D. Hermeneutics: dimensions toward LIS thinking. *Library and Information Science Research*, 11(1989)
- 5 Frohmann, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, 46:2(1990)
- 6 Miksa, F. The subject in the dictionary, catalog from Cutler to the present. Chicago: American Library Association, 1983
- 7 ISO. Documentation methods for examining documents, determining their subjects and selecting indexing terms. Geneva: International Organization for Standardization, 1985
- 8 Farrow, J. E. A cognitive process model of document indexing. *Journal of Documentation*, 47:2(1991)
- 9 Taylor, A. G. Books and other bibliographic material. In: Petersen, T, ed. Guide to indexing and cataloging with the Art and Architecture Thesaurus. New York: Oxford University Press, 1994
- 10 Petersen, T. Introduction. In: Petersen, T, ed. Guide to indexing and cataloging with the Art and Architecture Thesaurus. New York: Oxford University Press, 1994
- 11 Langridge, D. W. Subject analysis: principles and procedures. London: Bowker-Saur, 1989
- 12 Chu, C. M. & O'Brien, A. Subject analysis: the first critical stages in indexing. *Journal of Information Science*, 19(1993)
- 13 Mai, Jens-Erik. Deconstructing the indexing process. *Advances in Librarianship*, 23(1999)
- 14,15 Mai, Jens-Erik. Semiotics and indexing: an analysis of the subject indexing process. *Journal of Documentation*, 57:5(2001)

王知津 南开大学国际商学院图书馆学系主任、教授、企业管理专业博士生导师。通讯地址: 天津。邮编 300071。

黄欣 南开大学国际商学院图书馆学系 2000 级硕士研究生。通讯地址同上。(来稿时间: 2002-08-14)