

● 朱震远

文献情报检索语言障碍浅析

关于文献情报检索语言障碍(以下简称检索语言障碍),有人指出它是语言障碍的一种^[1],但可惜未能进一步揭示它同自然语言障碍、科学术语障碍的内在联系。把检索语言障碍排除在语言障碍以外,或仅仅将它视为检索系统内部障碍^[2,3]。更多的人过多地囿于检索语言的技术分析和微观分析,倾向于把检索语言仅仅视作检索系统的子系统,视作技术方法,而忽视了检索语言作为语言交流工具的一些本质性问题。

从语言交流角度对检索语言及其障碍进行探讨,有助于把检索语言研究从上述近乎纯技术观点的氛围中解放出来,有利于从宏观上、本质上更全面更深入地认识检索语言,从而为检索语言的设计、运用、改进和创新导航。

一、检索语言障碍的地位和重要性

文献情报检索是人类文献情报交流过程(以下简称交流过程)的特殊环节和特殊方面。检索语言障碍是文献情报交流障碍(以下简称交流障碍)的一种。人们对检索语言障碍的认识,首先表现在交流障碍种类的划分上。对交流障碍种类划分,可从多种角度进行。从交流障碍的性质划分,可分为语言障碍、社会文化障碍、心理障碍、管理障碍、学科障碍、技术障碍等;从社会交流过程的各环节划分,可

分为编辑系统障碍、印刷系统障碍、出版发行系统障碍、情报系统障碍、图书馆系统障碍等;从某具体系统的实际障碍划分,又可分为系统外障碍和系统内障碍。检索语言障碍从检索系统角度看属于系统内障碍,主要受制于由系统目标、功能、性能、服务对象、管理水平和技术设备等构成的子系统环境,在系统的开发运行过程中,可用管理的和技术的手段消除。它又是语言障碍,具体表现为检索语言各种特殊交流功能的障碍;它不仅由语言工具的本质属性所制约,还受到各种语言的、社会文化的、心理的各方面因素的支配。

在检索这一特殊交流过程中,检索语言障碍是最主要的障碍。这是因为:

1. 人类文献情报交流的最主要内容是语言信息。其他如图象信息、非语言音频信息等交流,不仅在规模上远比不上语言信息,而且在整个交流过程中必须依附于语言信息的交流^[4]。文献情报交流过程是一种语言交流过程。检索作为交流过程的特殊环境和方面也是语言的交流过程。检索语言在这个过程中承担着几乎是全部的语言交流功能。

2. 前苏联学者切尔内认为:“情报检索语言、标引规则和输出标准”构成的“逻辑语义工具”称为“抽象检索系统”。这个“抽象检索系统”构成了检索理论研究的“主要研究对象”。抽象检索系统直接承载着所有信息的交流全过程,而电脑设备、文档和人员操作等的“具体检索系统”只通过“抽象检索系统”间接作用于所交流的信息^[5]。检索过程即是检索

语言转换、运行的过程。

3. 在检索过程中,所有的语言障碍,包括自然语言障碍、科学术语障碍等都集中表现为检索语言障碍,而其他如社会文化的、心理的、管理的、技术的各种障碍,也主要表现为检索语言障碍。

二、检索语言障碍的本质

检索语言同任何自然、人工语言一样,是人类社会特定群体(各民族、国家、地区、行业、学科专业等)的交流和表达工具。国外布拉格学派将语言定义为“为特定的目的服务的表达手段的功能体系。”语言是“信息的载体”^[6]。人类的一切自然、人工的语言交流,都是以语义信息为内容,以传递语义信息为直接目的。从交流内容和直接目的来说,这种交流功能障碍是所交流内容的语义失真。信息的时间性,决定了障碍还表现为时间阻滞。故此,可将检索语言障碍的本质表述为“在检索这一交流过程中的语义失真和时间延滞。”在这里,语义失真有两种表现形式。一种形式是信息源——文献内容原来语义的丢失或部分丢失,另一形式是非原有语义信息作为“噪音”误传给接收者,在整体检索效果上前者表现为漏检,后者为误检。

张琪玉同志这样描述了检索过程(见图1)^[7]。

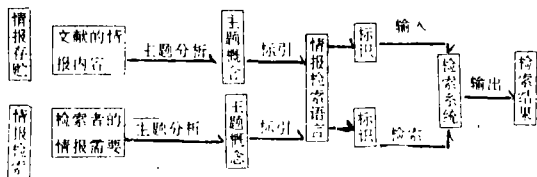


图1 检索过程

在理解图1检索过程时应注意两点:1. 不能将图中检索语言理解为检索过程的一个环节,图中虚线(虚线为笔者所加)的右边部分为检索语言的转换和运行。2. 用户需要和得到的结果是经过标引人员转换的文献作者的思想,检索语言沟通的是用户同作者(而不是标引人员)的思想。检索过程是作者思想信息向用户的单向传递过程。基于这两点,可将检索过程简化为图2。

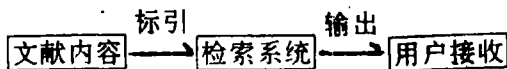


图2 简化的检索过程

由于检索过程就其内容说是一个信息过程,可将它抽象为图3。



图3 相应的信息过程

因检索过程又是一个语言交流过程,可将它表示为图4。

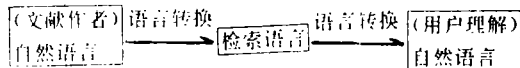


图4 相应的语言交流过程

从图4不难看出检索语言障碍产生于两次语言转换及其中间的检索语言运行过程。

通过图2、3、4的对比,可以对检索过程同相应的语言过程(抽象检索过程)、信息过程(检索交流的内容)之间的关系,有个比较

完整的认识。检索语言是其他两个过程的衔接体,实际检索过程通过检索语言的中介,使信息的交流得以实现。

三、检索语言障碍产生的内部原因

检索语言产生于自然语言,从自然语言中分化出来,以其特殊功能区别于自然语言。它与自然语言在本质上相同并有着千丝万缕的联系,和自然语言互相补充,共同实现交流。文献情报检索的实质是用户同文献作者的间接交流。这种间接交流环境有如下特点:1. 用户面对的是数量庞大,整体上无序的文献集合;2. 用户完成交流的前提是,在大量文献中进行选择,首先需要概括地了解文献内容;3. 用户的最终需要是获取文献中的基本语义信息;4. 用户和作者双方的所有发现创造(包括写作和阅读)活动都是以自然语言为基本工具的。特殊交流环境决定了检索语言的特殊功能。

检索语言必须具备一些特殊功能:1. 整体有序化功能。检索语言应在反映人类文化和科学知识的基础上,以便于掌握的方式,将文献情报在整体上有序化;信息内容的层次上有序化和检索语言标识的有序化。2. 简要表达功能。检索语言要用简要浓缩的形式表达文献或情报单元的内容。3. 语义匹配功能。检索语言要在基本语义的层次上将用户需求同文献内容进行相符性比较。要求语义上而不是语词符号上的相符性,一般不要求对情感语义等附加语义进行表达和匹配,只将基本语义作比较。4. 语言转换功能。要完成同自然语言的互相转换。将转换中的语义失真和转换的复杂性降到最低限度。四方面功能都包含着语义和时间的要求。

这些功能应视为一个整体。交流最终通过这个整体实现。功能之间是相互补充,相互

渗透的。比如:简要表达功能的实现有助于语义匹配,简要的表达形式更接近基本语义;整体有序化功能的实现,为语义匹配打下了可靠的基础,因为具体语义总是存在于特定语义场中,为特定关系所确定,而语义层次的整体有序化正提供了这种关系;整体有序化既然是检索语言的有序化,就必然建立在语言转换功能实现的基础上,因为文化和科学知识的一切成果,必然首先建立在自然语言基础上;等等。

四方面功能之间又是互相制约,并在一定程度上互不相容的。人类至今尚未有任何一种检索语言的具体形式和结构完美地实现上述所有功能。任何检索语言对某种功能来说可能比较完善,但必然对其他功能表现出不同程度的不适应。如关键词语言的语言转换功能较强,而整体有序化功能、语义匹配功能较弱;分类语言比主题语言更适于实现整体有序化功能,但在语言转换功能上略逊一筹;语义代码语言在语义匹配功能方面效果很好,却在语言转换上效果较差。现有的各种检索语言都从某种角度表现出检索语言功能间的互不相容。

检索语言功能间的互不相容是由功能的内在矛盾所决定的。检索语言功能既有语义内容的要求,又有时间的要求,既有对文献个体的语义要求,又有对文献集合整体的语义要求。这些不同方面的要求之间的矛盾决定了检索语言功能之间的不相容,甚至造成了同一功能中的内在冲突。比如,组配式语言与列举式语言的矛盾,主要表现为个体文献语义要求同整体文献语义要求的矛盾;控制语言和非控制语言的矛盾,在语言转换方面,表现为语义要求和时间要求的矛盾,而在整体有序化和简要表达方面又表现为文献整体语义要求和文献个体语义要求的矛盾,等等。计算机技术的应用虽然缓解了检索语言交流的矛盾,尤其对时间要求和语义要求的矛盾解决起了不小作用,但消除障碍的总体成效还

是非常有限的。

检索语言障碍还有更深的内部原因。检索语言具备语言交流工具的一些共有本质属性,这些属性限制了检索语言功能的实现:

1. “每一种语言都是一种自我包含的系统”⁽⁸⁾。语言符号的意义主要取决于系统内的关系结构。不同语言的互相转换,由于不同系统的关系结构的差异,往往导致语义传递的难以克服的障碍。从严格的意义上说,两种不同语言之间,不可能实现完全等价的转换。“语言转换功能”的实现只能是相对的。

2. 语言表达思想、反映客观现实的过程,是一个在内容上简化、抽象化,在时间上滞后的过程。要完整准确地反映对象间的复杂关系和及时反映对象变化是非常困难的。整体有序化功能和简要表达功能的实现也只能是相对的。

3. 语言工具的具体语用环境是开放的,不可控制的。作者思想的完整表达,除了靠词语的基本语义,往往还由特定社会历史背景和心理特征等构成的语境确定。用检索语言对文献语义的完整准确的重新表达往往也只能是相对的,这样就限制了简要表达功能和

语义匹配功能的实现。

4. 语言工具的音、形物质性决定了语言交流有必要的耗时量。口语音节的传递必须要一定的时间完成,而对书面语言字符的写作和阅读,同样要消耗一定的时间。语言交流的耗时特性限制了检索语言功能的实现。

至少在检索语言构建运行中,检索语言障碍的完全消除是不可能的。事实上,除了极特殊情况外,任何构造精良的检索系统,不可能同时达到100%的检全率和检准率,主要原因即在于此。

四、检索语言障碍产生的外部原因

由于检索语言障碍包括语言障碍和系统内部障碍,障碍的直接的外部原因也相应地分为两种,即语言因素和管理、技术的因素。这两种因素又分别受制于社会文化、心理、元语言、学科和母系统因素。检索语言障碍外部原因作用图示如图5。

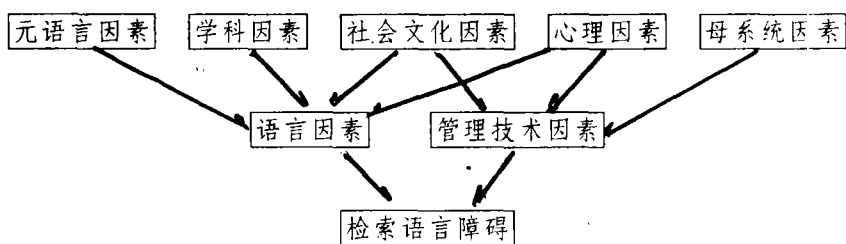


图5 检索语言障碍外部原因作用图

社会文化是形成检索语言障碍的重要原因。社会文化因含义较广,包括社会的文化特点、社会信息化程度、科技发展水平、管理水平、法制环境等。例如,社会信息化程度较低,可能影响系统管理人员、标引人员和用户对某种检索语言技术手段或相应管理方式的接受能力;不同国家民族或历史时期形成的社会文化环境,会带来语言风格的明显差异,表

现为同一专业用语的哲学化倾向、政治化倾向或实用化倾向等不同特点,往往对同一概念赋予明显不同的含义,由此而通过语言因素形成障碍。

心理因素同样间接地通过语言或管理技术的因素作用于检索语言。用户和系统设计、标引、操作人员,由于不同的年龄、职业或经历,会形成不同的心理特点,这些特点有时会

成为障碍产生的外部原因。有人曾介绍过国外的研究:在同一学院不同年级的大学生,对以字母为基础的字顺排序和以词为基础的字顺排序,其适应性有明显不同^[9]。一般社科研究人员比工程技术人员对机检语言的心理适应能力要差得多。

元语言因素指的是检索语言藉以构造和定义的自然语言因素。自然语言是任何人工语言,包括任何检索语言的元语言,即使是分类代码体系,同样须由自然语言给出定义。元语言的结构形式特点会通过语言因素形成某种特殊的检索语言障碍。最简单的例子就是汉语和英语检索语言的字顺排列具有完全不同的特性。

学科领域的因素对检索语言的作用不可忽视。它主要表现为检索语言中科学术语方面的障碍。罗春荣同志对此有较为详细的论述^[10]。

母系统指的是检索语言实体得以存在的检索系统。检索系统从管理和技术方面影响检索语言。检索系统的目标、功能、性能、服务对象、技术设备、管理方式等因素,构成了检索语言特定的运行环境,从外部的管理和技术方面作用于检索语言。

上述各外部原因的划分是相对的,互相交叉的。比如有些既是心理因素又是社会文化因素。还有一果多因的现象,如某种具体障碍同时受到多种外部因素的作用。还应注意到各外部因素之间的互相作用,如学科因素作为外部原因形成了障碍,可从这些学科因素的背后找出其他社会文化因素的作用。这类复杂关系对检索语言障碍的探讨过于间接,我们只需了解它的存在而不必细述。

大多数的检索语言障碍同时受到内外部原因的共同作用;而任何障碍在受到外部原因作用时,必然同时也有某种内部原因在起作用。亦即内部原因对障碍来说是必定存在的,外部原因必须通过内部原因起作用。

五、检索语言障碍的种类

根据障碍产生的内部原因,可分为功能不相容障碍、语言本质属性障碍等;

根据障碍产生的外部原因,可分为语言因素障碍、管理和技术因素障碍、还可再分为社会文化原因障碍、心理原因障碍、元语言原因障碍、学科原因障碍、母系统原因障碍等;

根据检索语言的功能,可分为整体有序化功能障碍、简要表达功能障碍、语义匹配功能障碍、语言转换功能障碍等;

根据检索语言构建和运行过程,可分为检索语言构建障碍、检索语言维护障碍、标引障碍、机检匹配障碍、检索需求表达障碍等;

根据检索语言的语言结构和交流,可分为字符障碍、词汇障碍、语法障碍、语义障碍、语用障碍等;

根据检索语言的文本结构,可分为主表障碍、辅助索引障碍、参照系统障碍、标引规则障碍、检索规则障碍等;

根据检索语言的标识排序,可分为分类排序障碍、字顺排序障碍等;

根据检索效果,可分为漏检障碍、误检障碍等;

根据障碍存在状态,可分为潜在障碍、实际障碍。

以上的划分是互有交叉的。由于划分是从不同角度进行的,而任何具体障碍是多维的,因此,某种具体障碍必然是某种原因障碍,某种功能障碍,同时还是某种过程障碍,某种效果障碍等。相对的、多维的划分只是为了便于从不同角度、不同方面分析检索语言障碍,更重要的是通过分析而全面了解各不同方面、不同因素、不同局部之间的互相关系、互相作用,从整体上全面、深入地认识检索语言障碍,并用以指导实践中的利弊权衡,方案选择,这才是分析的目的所在。

六、检索语言障碍的性质

检索语言障碍具有绝对性、相对性和统计规律性。

绝对性指的是障碍在任何检索语言中都是不可避免的。检索语言的本质属性限制了检索语言功能的实现,形成了障碍,这些障碍在检索语言构建运行过程中是无法消除的。当然,障碍的绝对性并不排除应用检索语言以外的手段可有效消除许多检索语言障碍。

相对性指的是具体障碍总是对特定系统、特定用户和特定需求而言。任一具体检索语言的结构和形式,对于不同系统、用户或需求,它的作用是不同的,对有的用户产生障碍,对其他用户可能不产生障碍;在某些条件下表现为障碍,在其他条件下可能不表现为障碍。例如,15篇特定文献的误检,对总数20篇文献的检索结果来说,是严重的障碍,而对总数为150篇以上的检索结果来说,可不算是障碍。汉语拼音音序排列中的同音现象,对熟悉音序的用户不成为障碍,而对不熟悉拼音的用户可能是严重障碍。检索程序的同样耗时量,对有些用户和需求是可接受的,对另一些用户或需求可能是障碍。

统计规律性(或随机性)指的是障碍的产生和作用,表现出随机性和不确定性。对这种随机的、不确定的现象,可用概率统计的方法予以科学描述。例如,某种检索语言的结构或形式特点,可能对系统A中90%的用户形成障碍,而对系统B,可能只有5%的用户产生障碍,这样,系统B中,这种检索语言障碍可忽略不计。又如,某种检索语言的特点,在某种检索环境中有90%的可能性会形成障碍,在另一环境中可以说有10%的可能产生障碍。这样,也可相应地把前一环境看成存在障碍,而把后一环境看成不存在这种障碍。大量的障碍都表现出上述统计规律性。统计规律性是绝对性和相对性的具体统一。

七、检索语言障碍消除的原则

检索语言障碍的消除是非常复杂的工作。这里仅就障碍消除的原则略述如下:

1. 整体性原则。检索语言的运行是个非常复杂的过程,它同时受到内外部各种原因的共同作用。消除障碍要注意各方面和各局部因素之间的互相关系,注意各种方法的协调运用,从待选方案的整体功效着眼。尤其要注意对任何改进和新技术新方法的引进,应充分考虑到新的障碍产生的可能性,审慎权衡,以求达到最优的整体功效。

2. 实践反馈原则。对检索语言的准确全面的认识,是个不断深入的过程。充分利用实验手段是一条捷径。目前,对检索语言各种具体结构、形式的互相关系以及在各种环境中的功能特点,人们一时还难以准确地描述和预测。即便如此,仍可通过实践的反馈,利用参数控制、数据分析等技术不断调节,使检索语言在具体环境中获得理想的效果。

3. 具体性原则。脱离了对具体环境的分析,盲目套用任何先进的方法手段,往往很难达到预期的总体效果。选择任何消除障碍的措施、方法,都必须从特定的具体环境出发,只有如此,检索语言的功能才会充分发挥出来。

4. 用户友好原则。用户是检索语言设计开发和运用中应考虑的最重要方面,用户的社会文化、心理、专业、语言等因素,是这些方面因素中的主要因素。检索系统和检索语言的成败关键在此。

5. 非检索语言手段原则。由于检索语言障碍的绝对性,仅限于从检索语言本身着手是不可能完全消除障碍的。必须充分运用非检索语言手段来帮助消除障碍。比如,计算机屏幕显示和程序的设计,可提供各种非检索

语言信息,如题名、刊名、刊期、标引频率、系统提示等信息,构成必要的语境,帮助消除障碍。运用检索反馈程序,配合检索方式的选用,以及加强管理或用户宣传教育等各种措施,都有助于障碍消除。

检索语言障碍问题复杂。笔者所见浅陋,愿能抛砖引玉,得专家同行指正,则不胜荣幸。

参考文献

- 1 严怡民主编.情报学概论.武汉:武汉大学出版社,1983:65,70~72.
- 2 米哈依洛夫 A. H. 等.科学交流与情报学.北京:科学技术文献出版社,1980:18~34.
- 3 周文骏.文献交流引论.北京:书目文献出版社,1986:140~145.

- 4 同[2]:90~92.
- 5 切尔内 A. H. 情报检索理论概述.北京:科学文献出版社,1980:16~17.
- 6 邢凯.语言是信息的载体——一个新的语言观.百科知识,1986,(11)
- 7 张琪玉.情报检索语言.武汉:武汉大学出版社,1983:2.
- 8 米曼. W. P. 描写语言学引论.上海:上海外语教育出版社,1986:19~21.
- 9 杰西卡. L. M. 主题检索系统设计选择.北京:书目文献出版社,1991:68.
- 10 罗春荣.主题检索语言的术语学基础.图书情报知识,1989,(3)

(来稿时间:1992-08-21. 编发者:黄文田。)

朱震远 1991年华东师大图书情报学系毕业。现为上海卢湾区教育学院资料室负责人,曾在2种刊物上发文2篇。通讯地址:上海市自忠路455号 邮编:200020

(接64页)

相统属,市馆难以起到这样的作用。专业性的图书馆群体可分为研究性及普及性两个层次,则作为网络建设的中心任务之一的文献资源协调也可分为两个层次,根据重点及易于实行两个因素,可先在同科技兴市同经济建设直接有关的层次中进行。如果科技文献馆是科委、文化局、科技情报所共同投资,科研单位和企业也参加了集资,就很自然地在他们的协作支持下把科研单位和企业的图书情报部门连结起来,建成以科技文献馆为中心的科技文献情报网络。如有可能还可建若干更专门化的分网。我们已建成了古籍馆,全市收藏古籍的单位不过十余家,已编有全市古籍善本书目,建一个以市古籍馆为中心的包括市县古籍收藏单位的古文献网络更易实行。市级少儿馆建网的目的和重点不是书刊资源的协调而是活动内容的协调,市少儿馆是少儿图书活动的组织者及示范者。建立专业性馆,网络建设的步子要比建综合性馆快得多。

不能认为有了星罗棋布的大中小型图书

馆,有了一些协作协调活动,网络已建成了。这只是建网的基础,不是真正意义上的网络,因为这是无形的,不稳定的。真正能发挥作用的网络必须有现代化手段,以可供检索利用的文献数据库为基础。这种网络是有形的,稳定的。市级公共图书馆率先建成专业性图书馆群体,要在网络建设中打破条块分割的界限,必须建立一个市级图书情报协调委员会,以规划指导全市图书情报事业的发展,使全市图书情报事业整体功能得到充分发挥。

以上所论是在一个市的范围内,以改革市馆的体制为契机,对市图书情报事业改革的总体设想。相信通过改革,图书馆情报事业为科学进步和经济建设服务的工作将会更上一层楼,焕发出新的光彩。

陆纲武 1971年吴县师范学校学英语。1984年学图书馆学专业,现为苏州市图书馆学会理事长、苏州市图书馆副馆长,曾在4种刊物上发文4篇。通讯地址:苏州市公园路2号 邮编:215006

(来稿时间:1992-08-12. 编发者:刘喜中。)