

《圣经》语词索引及其编制

贾玉文

(辽宁师范大学信息管理系 大连 116029)

文摘 本文对圣经语词索引的概念、种类、传统及现代的编制手段、编制工具,以及在现代语词索引编制过程中的意义进行了比较系统的阐述。作者认为,八十年代以来MAT主持的CD-ROM版和网络版圣经语词索引的变革具有划时代的意义。其自动化的编辑手段,系统化的索引数据库编制工具,对语词索引的自动化编制、不同语言文本的机器自动翻译,均具有非常重要的借鉴意义和参考价值。

关键词 语词索引 圣经语词索引 机编索引 机器翻译

语词索引在英文中写作Concordance,是一种以文本中的字词为标引对象的索引。《哈罗德图书馆员术语词典》(Harrods Librarian Glossary)将其定义为“一种按照字母顺序编排,以作品中的所有段落、关键词为标引对象,并通过这些段落或短语对与之相关的文句进行检索。在机器标引领域,则指一种按字母顺序编排的以文献中出现的每一词为标引对象的逐词索引。”语词索引在上个世纪二十年代被引入我国,译作“**堪靠灯**”,其索引理论和方法对后来的索引编制产生了极为深远的影响。

在西方,人们除了为著名作家的作品之外,数量最多的是《圣经》语词索引。近年来,随着机编索引技术和网络技术的发展,机编圣经语词索引及网络版圣经索引大量涌现,新的语词索引开发工具逐渐得以完善,集自动化标引、词语切分、文本自动纠错、多语言机器翻译及互注等技术为一体的索引编制工具为语词索引的编制带来了革命性的变化。先比之下,中文语词索引,特别是古籍索引的自动化编制发展相对滞后,我们期待着信息时代的“**堪靠灯**”能够引起一场新的索引革命。

1 圣经语词索引的种类

《圣经》语词索引即圣经经文的词语索引,这些词语包括圣经中的词汇、短语及相关专题,经常按字母顺序,对圣经词语在全书、章节或句子中的位置进行索引。

圣经语词索引有多种不同的形式,按照索引所标引的对象,有关键词索引和专题索引,以及由姓名和专有名词构成的名称索引;既有单行本的索引,也有附于圣经书后的附属索引;从规模的角度,圣经语词索引有详编和简编之分。1980年代以后,圣经语词索引又出现了新的载体形式,即CD-ROM版索引和网络版索引。

专题性的语词索引则将所有相关语句放在同一个主题之下,使整部圣经成为一个连贯的整体。关键词索引则是以经文文句中的关键性词汇,如人名、地名、专有名词等为标引对象,用以查考文句出处的一种索引。

详编圣经语词索引是对经文中的每一个字逐个进行标引的完全索引。这种索引对于语言计量或文本分析有重要的意义。然而,这些出版物规模庞大,许多词汇如各种连词、冠词的标引实际上对于非专业人士来说没有任何意义。一部详尽的圣经语词索引可以通过编制一部足本圣经的词表就可以完成。这些索引应用对象仅限于某些语言学家,而对那些普通读者则鲜有用处。

绝大多数圣经语词索引属于简编本索引,它可以单行,也可以附载于书后。相比之下,附于圣经后面的索引更为常见。实际上,大多数教徒虽然不购买或自己拥有语词索引,但经常使用。因此,小而精且好用的索引,如附于圣经书后的小型语词索引更乐于为人们所接受,它既可以方便阅读,也可以用来检索。

此外,圣经中的姓名及专有名词是圣经语词索引标引的重要内容。圣经中涉及到的

人物姓名或故事经常在圣经中一个部分或章节中出现,人们既需要查找人物,也需要相关事迹,所以单独编制姓名和专有名词索引,或两者合并在一起共同组成一个名称索引同样非常重要。

2 早期的圣经语词索引及其编制

世界上最早的语词索引是从马所拉本《圣经》(the Masorete)的书后词表开始的。马所拉本《圣经》大约出现在6-9世纪(一说11世纪),泰比里厄斯(Tiberias)和苏拉(Sura)的犹太教学者根据《圣经》文句和读音的口传传统,编成了《圣经译著集》,他们译注和鉴定的《圣经》后来被称作“马所拉本”《圣经》。为了查阅的方便,他们编制了一个词表和目录附在这部《圣经》的后面,这个词表就是圣经语词索引的雏形。

第一部给出完整参考文献的圣经语词索引是由口米尼克教派教士根据拉丁文本《圣经》编成的。该索引完成于1230年,主持者名字叫雨果(Ilugo de Sancto Charo),参编者多达500多人。与后来的语词索引不同的是,这部索引没有引文,所以称其为一部查找圣经文本中语词出处的索引更为合适。大约两百年以后,一个叫Isaac Nathan的犹太人受拉丁文《圣经》索引的影响,用十年的时间编成了既可以查找词语出处,又有引文的圣经语词索引,由于这部索引的初衷在于捍卫犹太教教义,且以希伯来文编写,在当时基督教世界影响较小,直到1523年才在威尼斯出版。

最早的英文圣经语词索引出现于十六世纪中叶,但其索引对象仅仅限于《新约全书》。首部针对整部《圣经》的英文本语词索引则出现在1550年,编者叫John Marbeck,他是温莎学院(Windsor College)的创建者,该索引亦因此而久负盛名。

在英语世界里影响最大的圣经语词索引当属克鲁登(Alexander Cruden)于1736年编辑的《新旧约全书语词索引》(Complete Concordance to The Old & New Testaments),该书第一版在1738年付梓印刷。这部书对新约全书和旧约全书中的所有词汇均进行了标引,范围甚至包括了介词和标引词的读音。虽然克鲁登患有严重的精神疾病,但在患病期间,他仍然治学不辍,分别在1761年和1769年完成了对第一版的修订,其中1769年版本还包括了圣经外传(Apocrypha)中的词汇。克鲁登的圣经索引家喻户晓,后来的许多同类著作均可以说是克氏著作的缩略本,其中以Rev. John Eadie《简编克鲁登圣经索引》(Condensed Cruden)较为流行。

此外,1879年罗伯特·杨(Robert Young)在爱丁堡出版的《圣经语词分析索引》(Analytical Concordance to the Bible)也是西方世界使用比较广泛的圣经索引,该书对英文本圣经中的语词增加了希伯来文或希腊文的词源释读。1894年,詹姆斯·斯特朗(James Strong)在纽约出版了《详版圣经语词索引》(Exhaustive Concordance of the Bible),其所收录的英文语词绝大部分标注了对应的希伯来语或希腊语字词,从而将英文本圣经索引的详尽标引传统与希腊语和希伯来语词典的权威解释有机地结合在一起。

中文圣经语词索引常以“圣经词类串”或“经文汇编”的形式出现,但品种少,数量也极为有限。

手工编制语词索引既费时又费力。1980年代以后,随着计算机和网络的普及,圣经索引的编制也出现了许多新的变化,许多新的非纸质载体的圣经语词索引应运而生。为了适应新的需求,圣经公会设立专门组织——机器辅助翻译小组(Machine Assistant Translation Team,简称MAT),专门负责不同语言的CD-ROM版和网络版圣经索引的开发。自八十年代以来,该组织已经开发出十多种不同类型的圣经语词索引,供研究人士和普通信众使用。利用MAT编制的索引,阅读或检索圣经会更为方便。

3 圣经公会(the Bible Society)的机编索引计划

圣经公会是基督教专门从事圣经出版和发行的组织,1712年由德国男爵康斯太因(Hildebrand von Canstein)创立。1802年,英国圣经公会成立,这是近代以来最大的

圣经会组织。在它的影响下, 尼德兰圣经公会(1814年), 美国圣经公会(1816年), 苏格兰的全国圣经公会(1826年)等区域性组织纷纷成立, 联合圣经公会就是在这些区域性组织的基础上成立的。两百多年来, 圣经公会为圣经的推广和基督教的传播起到了相当大的作用。1980年代初, 鉴于新型媒体和载体的出现, 英国圣经公会成立了专门负责开发新载体圣经索引的机构——机器辅助翻译(MAT)开发组, 在全球范围内施行其圣经语词索引编制计划。从1985年至今, MAT已经开发出十几种不同语言、不同规模的圣经语词索引数据库, 并完成了从语词索引自动标引、互译、检测, 到排版、印刷等一系列自动化语词索引数据库编制工具, 这些工具不仅对圣经索引编制, 而且也对其他领域语词索引数据库的编制, 同样具有重要的参考价值。

到目前为止, MAT计划中已完成的索引项目包括:

数据库名称	语种	版本形式	出版年
Good News Bible	英文	半详编	1983
Good News Bible	英文	书后索引	1987
New Testament	赞比亚汤加文	单行本	1990
Synodal Russian Bible	俄文	半详编	1995
Modern Hebrew concordance to the New Testament	希伯来文	近详编	1999
Reina\Velera (RVR) Spanish Bible	西班牙文	简编本	
Version Popular (VP) Spanish Bibles	西班牙文	简编通行本	
Swahili Common Language Bible	斯瓦西里语	简编本	
Swahili Common Language Bibl	斯瓦西里语	简编本	
the New Welsh Bible Concordance	威尔士语		1998

目前正在进行的项目有:

数据库名称	语种	版本形式
the Kikuyu Bible	肯尼亚基库尤语	简编本
Contemporary English Version (CEV)	英文	简编本
the Latvian Bible	拉脱维亚语	半详编
the Modern Hebrew New Testament	希伯来文	CD-ROM版
the New Welsh Bible	威尔士语	网络版

其中的《现代希伯来语新约全书》(the Modern Hebrew New Testament)是MAT为以色列圣经公会开发的一种基于屏幕的CD-ROM版详编可视检索数据库, 而网络版的《新威尔士圣经》(the New Welsh Bible)语词索引则是MAT与威尔士国家图书馆(the National Library of Wales)、斯旺西高等教育研究所(Swansea Institute of Higher Education)合作基础上完成的。

除了这些综合性的索引之外, MAT还利用自己研发的自动索引工具, 开发出种类繁多的专题性索引及工具书, 包括语词索引、名称索引、圣经选集索引、术语词典、双语词典等等。

4 MAT的自动索引编制工具

八十年代以后, MAT研发出了一系列的工具性软件, 对综合性和专题性的圣经索引, 尤其是商业性很强的小型索引进行自动编制, 这些程序共同组成了MAT程序家族。

MAT程序组包括两大部分, 其一是工具系统, 主要用于圣经索引的编制和检测, 包括文本审核程序(BARUCH) [注: 括弧内为程序名, 下同]、文件格式检测及转换程序(RHESUS)、文本编辑器(Philemon)、语词索引编辑器(SILAS)、圣经检索引擎(LUKE)、

圣经版本对照工具 (Mark)、索引检测程序 (Job)、名称索引编辑器 (Adam)、字符串自动切分工具 (Atropos) 及排版程序 (Aaron) 十种程序。

其二是示范系统, 这是MAT目前正在研究的示范程序, 具体包括多语言互注系统 (AUGUSTUS)、词素自动分析系统 (BARTHOLOMEW)、专有名词自动确认系统 (PERCIVA)、专有名词翻译引擎 (PTOLEMY) 四部分。

在MAT程序组中, 语词索引编辑器 (SILAS)、圣经检索引擎 (LUKE)、圣经版本对照工具 (Mark) 属于索引编制的基本工具, SILAS可以利用已有的标引词, 通过增减原模型中的词汇, 建立起一个新的索引编制模型, LUKE则可以用来建立主题索引或文句模块, MARK是一种各种版本间的词语、句子的比照引擎, 它可以利用先前的语词索引建立一个全新的语词索引, 方法就是通过比较两个版本中的共用元素 (词语或句子) 建立新的语库。

对索引文件 (包括文本或数据库) 的编辑是索引编制的重要环节, 文件格式的转换、多语言文件的分析与审核、字符的切分及专有名词的自动抽取, 均需要专门的工具来完成。

在MAT程序组中, BARUCH 和Philemon是多语言文件编辑、分析及审核程序。Philemon负责多语言文件的编辑, BARUCH则可以对文本内容进行统计, 也可以对文档中的非法字符或错误文字进行跟踪定位。对文本结构组织 (如词组) 的统计包括: 对全词表进行ASCII、频率及用例分类, 在ASCII、频率及用例分类基础上的两组、三组或四组字符的匹配分析, 词频及用例分析 (即对所有词汇出现次数列表), 对字符长度、首字母进行线性分析, 对字符长度、赋值、频率进行量化分析, 对字符使用频率进行组配分析等多种功能, 这些功能可以对圣经的研究工作提供强有力的支持。

文件格式转换由 RHESUS 专门负责, 它可以对 BCV、SFM、Ventura、RTF 格式各种新格式文件, 如 XML 文件进行转换, 不列颠及其他国家圣经研究会均可以利用该程序对任何格式的圣经文本进行转换。

Atropos 主要处理由 Silas, Luke 或 Mark 生成的语词索引文件的字符切分, 它可以通过对排版数据和语言信息的处理, 利用编辑程序提供的参数, 通过 Silas 编辑系统所生成的一系列索引文件, 自动将那些最适合语词索引的每一个可作为标引对象的文句挑选出来。筛选文句中的一部分作为索引的检索标目, 这是一项既费时又费力的工作。现在这项工作多数情况下都有计算机来完成。一般说来, 语词索引中大约 70-90% 的条目均可以由编辑程序按照设定规则自动进行切分挑选。切分的准确程度一般要视文本文件的语言及现有索引的印刷文本情况而定。Atropos 还可以与另一个利用字母大小写确认事物名称的标引工具 Adam 一起对文本中的各种名称进行完全标引而直接生成索引。

圣经文本多语种的翻译与互注主要通过示范系统的功能来实现。AUGUSTUS 语言互注系统实际上是一个双语词库, 在目标文本和对照文本作支持的基础上, 可以对曲折变化及词缀变化较多的语言的术语进行自动翻译和注释。作为 AUGUSTUS 奥古斯都系统的补充, BARTHOLOMEW 词素自动分析系统可以对既定文本进行词素的自动分析和分类, 两个系统的目标基本一致, 均是为既定文本语言的寓意结构和词素结构提供自动的知识平台, 但 BARTHOLOMEW 仅需要目标文本, 而不需要对照文本作支持。

对于专有名词的确认和翻译则需要另两种工具——PERCIVAL 专有名词自动确认系统和 PTOLEMY 专有名词翻译系统来完成。PERCIVAL 和 PTOLEMY 是对圣经文本中的专有名词进行自动分析的专用软件, PERCIVAL 由奥古斯都的部分组件及其他用于确认词汇中缀结构的程序组合而成, 用以指认圣经中的专有名词, 对于存在拼写差异、不仅一种拼写形式、且使用过程中拼写形式有变化的专有名词, PERCIVA 也可以进行有效识别。PTOLEMY 则可以以圣经文本为底本, 通过与另一已确认专有名词的文本的比照分析, 从圣经中提取专

有名词。

圣经语词索引的编辑工作完成以后,对索引文件的检测及分析仍是不可或缺的环节, JOB 专门用于检测由 STILAS, LUKE 及 MARK 生成的索引文件,它可以在排版之前对索引语词的逻辑错误及编辑审校过程中可能出现的错误进行检测,并同时对文件内容进行量化分析。

5 结语

作为一种专门索引,数百年来,圣经语词索引在载体形式、编制手段、编制工具等方面均出现了重大变革,尤其是上个世纪八十年代以来MAT主持的CD-ROM版和网络版圣经语词索引的变革,更具有划时代的意义。其自动化的编辑手段,系统化的索引数据库编制工具,不但对圣经索引的编制,而且对其他领域语词索引的自动化编制、不同语言文本的机器自动翻译,均具有非常重要的借鉴意义和参考价值。

参考文献

- 1 任继愈.《宗教词典》.上海辞书出版社,1981.12
- 2 詹卫东.词的语义分类在汉英机器翻译中所起的作用及难以解决的问题.《语言工程》.清华大学出版社,1997
- 3 Harrod L. M. ed. Harrod' s Libraryans' Glossary. 5th ed. Vermont:Gower Publishing Company, 1984
- 4 <http://www.bfbs.or.uk/index.html>
- 5 http://www.wikipedia.org/Main_page
- 6 <http://www.speedbibleconcordance.com>
- 7 <http://www.godsvieiw.com>
- 8 <http://www.bibleric.com>
- 9 <http://www.online-bible.org/index.html>