

加注标签软件与日语研究

本书是「方法工具与日语教学研究丛书」之一。主要介绍了日语语料库专用加注标签软件、偏误语料库专用加注标签软件和翻译语料库专用加注标签软件的内容、使用方法，以及加注标签软件在日语教学和日语研究中的应用。

于康 [日]田中良 [日]高山弘子 著

丛书主编 张威

方法工具与日语教学研究丛书



浙江工商大学出版社

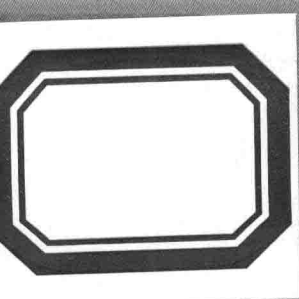
方法工具与日语教学研究丛书

丛书主编 张威

加注标签软件与日语研究

本书是『方法工具与日语教学研究丛书』之一。主要介绍了日语语料库专用加注标签软件、偏误语料库专用加注标签软件和翻译语料库专用加注标签软件的内容、使用方法，以及加注标签软件在日语教学和日语研究中的应用。

于康 [日]田中良 [日]高山弘子 著



浙江工商大学出版社

图书在版编目(CIP)数据

加注标签软件与日语研究 / 于康, (日) 田中良,
(日) 高山弘子著. —杭州: 浙江工商大学出版社,
2014.3

(方法工具与日语教学研究丛书)

ISBN 978-7-5178-0221-1

I. ①加… II. ①于… ②田… ③高… III. ①日语—
研究 IV. ①H36

中国版本图书馆 CIP 数据核字(2014)第 008780 号

加注标签软件与日语研究

于 康 [日]田中良 [日]高山弘子 著

责任编辑 罗丁瑞 姚 媛

责任校对 刘 颖

封面设计 王好驰

责任印制 包建辉

出版发行 浙江工商大学出版社

(杭州市教工路 198 号 邮政编码 310012)

(E-mail: zjgsupress@163.com)

(网址: <http://www.zjgsupress.com>)

电话: 0571-88904980, 88831806(传真)

排 版 杭州朝曦图文设计有限公司

印 刷 浙江云广印业有限公司

开 本 710mm×1000mm 1/16

印 张 15.25

字 数 298 千

版 次 2014 年 3 月第 1 版 2014 年 3 月第 1 次印刷

书 号 ISBN 978-7-5178-0221-1

定 价 36.00 元



版权所有 翻印必究 印装差错 负责调换

浙江工商大学出版社营销部邮购电话 0571-88804228

卷 首 语

近年来,我国的日语研究与日语教学研究取得了蓬勃发展,越来越多的青年学者和研究生投身于这个学术领域,使得研究的规模不断扩大、研究的内容不断深入、研究的方法不断改进、研究的水平也不断提高。目前全国有 466 所大学开设日语专业,具有日语语言文学硕士点的院校已经超过 100 所。尤其是自 2011 年以来,具有日语语言文学专业博士学位授予权的单位也已从过去的 7 家发展到了 30 余家,翻了两番多。这样一个前所未有的发展态势,实在是令人欣喜。在这样一个大的背景下,如何有效地指导青年学者和研究生掌握与时俱进的研究方法与工具,成为学界广泛瞩目的重要课题和艰巨任务。为了适应我国日语研究与日语教学研究日益发展的实际需求,我们筹划编辑出版了这套“方法工具与日语教学研究丛书”。

2013 年 3 月,本丛书的首部作品《语料库的制作与日语研究》问世,在社会上引起了很大的反响。从读者反馈的信息得知,像这样由浅入深地传授日语语料库的基础知识,手把手地讲解如何利用免费软件制作日语语料库的方法,详细介绍如何使用语料库进行日语研究和日语教学研究的专业指导书尚不多见,读者非常喜爱这样务实的学术作品。尤其是关于如何建设个性化加注标签语料库的论述和对于加注标签的详细步骤以及如何利用加注标签的语料库进行日语研究等内容的阐述循序渐进、通俗易懂。掌握了这种工具方法后,就能够熟练地将其应用到各自的研究中去。读者的评价与期待对于作者来说,既是莫大的鼓舞,也是一种激励,成为激发我们继续深入研究的动力。现在,本丛书的第 2 部作品《加注标签软件与日语研究》马上就要与读者见面了。本书是在上一本的基础上,对如何建设和使用不同用途的有标签日语语料库进行更加深入的研究与探索而取得的又一个具有代表性的成果。

随着家用电脑和计算机网络技术的普及与更新,在当代语言学研究中,语料库作为一种现代化的研究工具,正在逐渐取代传统的手工收集语料的方式。这是

一个与时俱进的历史性变化,这个趋势将继续发展下去。而语料库的出现与发展,大大减少了语言研究者在收集例句方面的困难。同时,我们也清楚地意识到,在语料库语言学领域,尤其是在日语语料库的研究与开发方面,有标签语料库无论是在数量、规模还是应用等方面,都还处于比较低的发展水平,这是一个亟待改善的局面。要想使语料库能够按照研究人员的目的和意愿有效地完成收集、检索、统计和分析语料的一条龙式操作,需要具备以下两个前提:一个是科学、合理地设计需加注的标签种类和内容;另一个是掌握方便、快捷地给语料加注各类不同标签的方法。

为此,本书的作者经过不懈的努力钻研,开发研制了以下3个专用的加注标签软件——日语语料库专用的加注标签软件 TNR_JapaneseCorpus、偏误语料库专用的加注标签软件 TNR_ErrorCorpus 和翻译语料库专用的加注标签软件 TNR_TranslationCorpus。这3个软件均属作者自行开发,在使用它们给语料加注标签时具有简明、实用、方便、快捷和用途范围广等优点,非常易于学习和掌握。为了使广大读者能够充分地利用这些软件,推动我国日语研究和日语教学研究的发展,作者声明决定放弃申请专利,在适当的时候向广大的读者和研究者公开这些工具软件。这种一心为学术和社会服务的奉献精神的确令人钦佩,也值得我们提倡和进一步推广。使用这3个加注标签的专用软件,可以制作出能够应用于日语研究、日语偏误研究以及日语翻译教学研究的专用语料库。本书坚持由浅入深、通俗易懂的一贯理念,对于上述3个专用的加注标签软件的相关内容以及使用的具体方法和步骤进行了详细、全面的讲解。正如本书的作者所强调的,语料库和加注标签的软件是两个非常有效的研究工具,配套使用这两个工具可以大大地提高研究的速度,增加研究的深度。因此,深刻了解和学会使用上述3个软件给语料加注标签,对于充分地利用语料库进行日语研究、日语教学研究及日语翻译研究,具有重要的现实意义,可以使我们在收集、检索、统计和分析语料方面事半功倍,收到“快刀斩乱麻”的效果。

本书还有一个重要的亮点,就是通过细致的描写和阐释,向读者展示了关于日语研究、日语教学研究以及日语翻译研究的理论方法和研究理念,其中有不少内容对于广大日语研究者而言具有较高的借鉴价值。尤其是在翻译语料库的设计方面,作者提出了不同于传统的翻译教学理念和衡量标准。我们认为,这些新的观点和视角可以为推动我国翻译教学领域的创新提出新的思路,引发人们进行更加深层次的思考与研究。

本书是作者多年来潜心研究的经验积累和成果结晶。俗话说“授人以鱼不如授人以渔”，这也代表了本丛书的基本理念。我们衷心地期待这部书能够成为有志于日语研究、日语教学研究和日语翻译研究的广大青年学者和研究生朋友们的良师益友，从而为推动我国的日语教学和日语研究事业做出应有的贡献。

张 威

2014 年 1 月

前 言

文科的研究大部分需要依靠材料来说话。也就是说,以材料为根据,从材料中抽取规则。日语研究也是如此。

日语研究收集语料经过了一个十分艰苦的阶段,即在相当长的一段时间里都是依靠笔记和卡片来收集例句。因此,大多数人在研究的准备阶段就已经耗费了很多的时间和精力,实在有点浪费。因为,研究的重点应该放在如何分析例句和解决问题上。

近年来语料库的开发研制发展很快,收集例句已经不是什么难事了。我们十多年来也一直致力于语料库的建设,并且,为普及如何制作语料库也做了一些工作。

现在,收集例句的确不再是难事,但是面对大量的例句该如何去分析,如何从大量的例句中发现规则,又成了一个新的令人头疼的问题。为了解决这个问题,我们开发研制了3个加注标签的软件。为了尽快满足读者的要求,我们放弃了申请专利,日夜兼程将其归纳成书奉献给各位读者,希望能对大家的日语研究和日语教学研究派上一点用场。

由于这3个软件都属于首创,无先贤的成果可以借鉴,所以,会存在各种不足之处,敬祈各位读者不吝赐教。我们还在不断地对3个软件进行改进和升级。希望在今后预定出版的有关书籍中能够提供更加便于使用的升级版。

语料库和加注标签的软件是两个非常有效的研究工具,这两个工具的配套使用可以大大地提高研究的速度,增加研究的深度。不过,手上有了砍柴的工具,不等于已经砍到了柴。接下来就需要读者自己上山去寻找自家炉灶需要的柴火,并把它们带回家了。如果嫌麻烦,连山都不愿意上,只等着别人给你送柴火来,那也只能望“山”兴叹了。

本书的第4章是国家社科基金项目《翻译教学理论、教学体系和家学模式的研究与翻译语料库的建设》(批准号:11BYY013,课题负责人:邱鸣)的研究成果之一,在此特致谢意。

于 康
2014年1月

目 录

卷首语	1
前言	1
第 1 章 语料库与标签	1
1.1 为什么要给例句加注标签	1
1.2 标签的种类	2
1.3 设计和加注标签的准则	4
1.4 加注标签的方式与标签的内容	4
1.5 加注标签软件的类型	5
1.6 加注标签时所需要的电脑配置和所需软件	6
1.6.1 电脑的配置	6
1.6.2 所需软件	7
1.7 小结	7
第 2 章 日语语料库专用的加注标签软件 TNR_JapaneseCorpus ...	9
2.1 加注标签的主要方法	9
2.2 加注标签的对象和加注标签的主要方式	10
2.3 TNR_JapaneseCorpus 的主要构成与拷贝	12
2.4 如何使用 TNR_JapaneseCorpus 给例句加注标签	13
2.4.1 收集例句	14
2.4.2 清理例句	17
2.4.3 给清理后的例句加注标签	24
2.4.4 制作带标签的语料库	57
2.5 小结	67

第 3 章 偏误语料库专用的加注标签软件 TNR_ErrorCorpus 68

3.1	TNR_ErrorCorpus 的主要构成与拷贝	68
3.2	加注标签的主要程序与各文件夹和软件的主要功能	69
3.3	如何使用 TNR_WritingCorrection 批改日语作文	70
3.3.1	标签的形式	70
3.3.2	建立文件夹与移动文件	71
3.3.3	批改日语作文的主要步骤	74
3.3.4	键入批改人的意见	85
3.3.5	删除批改的信息	87
3.3.6	保存批改后的作文	88
3.3.7	在保存的作文中自动生成正误标签	89
3.4	如何使用 TNR_WritingCorrection 转换用 Word 批改的作文	91
3.4.1	如何反映 Word 第 1 种批改方法批改的信息	94
3.4.2	如何反映 Word 第 2 种批改方法批改的信息	97
3.4.3	删除批改的信息	101
3.5	如何使用 TNR_ErrorCorpusTagger 给偏误作文加注标签	102
3.5.1	给偏误句子加注标签	103
3.5.2	修改标签	107
3.5.3	保存加注标签后的作文	111
3.6	制作带标签的语料库	112
3.6.1	使用 edamame_v21 转换文件的格式	113
3.6.2	使用 Himawari_1_3 制作带标签的语料库	114
3.7	小结	116

第 4 章 翻译语料库专用的加注标签软件 TNR_TranslationCorpus 118

4.1	加注标签的目的	120
4.2	TNR_TranslationCorpus 的主要构成与拷贝	121
4.3	加注标签的主要程序	122
4.4	如何使用 TNR_TranslationCorpus 给原文和译文加注标签	123
4.4.1	收集原文和译文	128
4.4.2	建立新的文件夹和向新文件夹内放入原文和译文	129
4.4.3	启动加注标签的软件	132
4.4.4	给日语原文加注标签	133
4.4.5	给汉语译文加注标签	160

4.4.6 给日语原文和汉语译文加注对应标签	176
4.5 小结	187
第5章 加注标签软件在日语教学和日语研究中的应用	190
5.1 加注标签软件与日语研究	191
5.1.1 「X 離れ」的研究程序	191
5.1.2 收集「X 離れ」的例句和清理例句	192
5.1.3 给「X 離れ」的例句加注标签	192
5.1.4 制作加注标签的语料库	196
5.1.5 检索加注标签的例句并进行统计和分析	203
5.2 加注标签软件与日语作文教学和偏误研究	210
5.2.1 批改作文软件 TNR_WritingCorrection 与日语作文教学	210
5.2.2 TNR_ErrorCorpusTagger 与偏误研究	223
5.3 小结	229
后记	231

第1章 语料库与标签

1.1 为什么要给例句加注标签^①

在《语料库的制作与日语研究》中我们详细介绍了如何利用免费软件来制作日语语料库和如何使用语料库。但如果有了自己的语料库就可以不受时间、地点和可否上网等条件的限制,随时随地根据学习和研究的需要检索例句。

的确,与手工收集例句相比,使用语料库检索例句不仅可以提高收集例句的速度,而且还可以解决文献体裁涵盖面小的问题。由于基本上可以根据自己的研究需要进行各种各样的检索并收集各种各样的例句,这就从根本上解决了例句不足的问题。

但是,学会了制作语料库并非大功告成,收集到了数以千计或数以万计的例句也并非意味着研究已见光明。还有一项非常关键和非常重要的、同时也是非常令人头痛的工作在等待着我们,那就是如何分析收集来的例句和从收集来的例句中找到研究的线索。

以前我们说过^②,无论从事什么样的研究,基本上都有两个方法。一个是归纳法,一个是演绎法。使用语料库进行研究,基本是以归纳法为主,即从大量的例句中去归纳和发现规则或规律。而问题是应该如何去归纳和发现规则或规律。

比如,在20个人中,有5个人是兄弟姐妹的关系,即一家人。其他皆为与这一家人、同时也与其他人无关的人。这时,大家都会用自己的大脑对20个人进行分类,然后在脑内进行各种各样的排列组合,最后辨别出这5个人来。

① 本书在论述标签问题时,有的时候使用“标注标签”,有的时候使用“加注标签”。两者在本书中视为同义词,只是根据文章的行文需要有所选择而已。另外,在先贤的研究中,会出现各种不同的说法,比如“加注标签”“标注标签”“赋予标记”“加标签”“赋标”等。这些术语指的都是同一种现象,所以,本书不作严格区分。

② 参见于康:《日语论文写作——方法与实践》,高等教育出版社2008年版;于康:《现代日语语言学丛书语法学》,高等教育出版社2012年版。

对 20 个人进行分类,依据的是区别特征。这些区别特征实际上就是一种标签。给每个人标注上各种各样的标签,然后再对这些标签反反复复地进行各种各样的排列组合,最后找出具有相同特征的标签,以此来认定具有兄弟姐妹关系的 5 个人。

由此可见,归纳例句和从例句中发现规则或规律比较有效的方法是给研究对象加注各类标签,通过对标签的不断分类和排列组合,剔除无用的信息,最后归纳出具有上位概念意义的特征来。这个特征就是具有能产性和普遍解释意义的规则或规律。

对研究的对象进行分类实际上就是一种合并同类项的工作。类型相同者一定会具有相同的特征,这个特征也就是用来证明自己不同于其他、并具有区别意义的条件。

合并同类项需要分类依据,只有同类的项目才能合并。在不断地合并同类项的过程中,区别特征会越来越清晰和显著。分类需要依据标记,只有相同的标记才能证明其为同类。这个标记就是我们所说的标签。

当研究对象的数量不太多时,可以依靠大脑来给对象加标记,并用大脑来对标记进行筛选和分类,最后实现合并同类项。但是,当对象的数量庞大时,由于大脑的记忆力有限,就无法快速准确地对研究对象进行分析和分类了。

比如,如果需要在 1 万个人中找出具有父子关系的人,用大脑来进行分类和归纳就很困难了。此时,需要给 1 万个人加注各类标签,然后使用电脑对这些标签进行分类和统计,最后归纳出所有具有父子关系的人来。

要想使检索出来的例句活起来,即从检索出来的例句中顺利地归纳出规则或规律来,给例句加注标签是一个必不可少的研究步骤。通过对标签的归纳和统计,可以发现通常依靠目视无法发现的问题和意想不到的规则或规律。这样就可以大大地提高研究的速度和深度,使枯燥无味的研究变成一种乐趣。

1.2 标签的种类

与日语研究和日语教学相关的语料库基本可以分为 3 大类。一类是日语语料库,一类是偏误语料库,一类是翻译语料库。日语语料库和偏误语料库属于单语语料库,翻译语料库属于双语或多语语料库。无论哪类语料库都需要加注标签,以此来提高分析问题和解决问题的速度和深度。

标签的种类会因研究领域以及研究目的的不同而有所不同。研究语言用的标签通常可以分为两个大类:

① 注明例句出处的标签

② 注明例句成分的标签

例句出处的标签指的是注明例句来自何处的标签。比如作者姓名、性别、年龄、身份,刊登的报刊、杂志、书籍以及体裁等信息。例句成分的标签指的是注明例句中各类句子成分性质的标签,比如词汇、语法、句法、语义、语用和篇章、话语等信息。

例句出处的标签中,日语语料库需要标注作者的姓名、性别、作品名、刊登的刊物、出版单位、出版时间、文章的体裁等信息。

偏误语料库由于以学习者的作文为主,通常需要标注作者的姓名、性别、年级、学习日语的时间、留学的经历和时间、文章的体裁等信息。

翻译语料库由两个小类构成:一个是标准翻译语料库,一个是翻译习作语料库。标准翻译语料库指的是原文与译文都是由正式刊行或出版的文章或书籍构成的语料库。翻译习作语料库指的是原文为正式出版物、而译文为翻译作业的语料库^①。在标准翻译语料库中,需要标注译者的姓名、性别、作品名、刊登的刊物、出版单位、出版时间、文章的体裁等信息。在翻译习作语料库中,需要标注译者姓名、性别、年级、学习日语的时间、留学的经历和时间、作品名、文章的体裁等信息。

在例句成分的标签中,日语语料库、偏误语料库和翻译语料库都可以标注词汇、语法、句法、语义、语用和篇章、话语等信息。加注标签时有两个方法:一个是穷尽加注法;一个是部分加注法。穷尽加注法指的是不仅给例句加注词汇、语法、句法、语义等标签,同时还加注语用、篇章和话语等标签。部分加注法指的是根据研究的需要在词汇、语法、句法、语义、语用和篇章、话语等标签中,选择最需要的标签给例句加注。

无论是采用穷尽加注法,还是部分加注法,都取决于读者的研究目的和加注标签的时间。穷尽加注法和部分加注法各有利弊。穷尽加注法可以相对地一劳永逸,但需要漫长的时间。部分加注法有的放矢,对症下药,可以大大缩短加注的时间,但由于只能“有的放矢和对症下药”,所以,当标签的种类不能满足研究的需要时,就需要再加注二次性或三次性的标签。

实际上,穷尽加注法并非能够真正达到穷尽,还会有很多注意不到的变数。在加注标签的过程中,二次加注标签和三次加注标签往往是不可避免的。

① 标准翻译语料库和翻译习作语料库的名称并不精准,属于暂时的叫法。因为,标准翻译语料库中的译文未必都是正确的译文,也会包括误译,而翻译习作语料库中的译文未必都是错误的译文,也会包括正确的翻译。只不过标准翻译语料库中的文章属于正式出版物,翻译习作语料库中的文章属于非正式出版物,而且,翻译习作语料库中误译率会远远大于标准翻译语料库,所以姑且暂用此名。

“二次加注标签”指的是给通过对加注标签的语料库进行检索得到的例句再次加注标签。三次加注标签的原理与二次加注标签相同。因此,从合理分配时间和精力这个角度来考虑的话,一开始不必求全,可以先从加注最需要的标签着手,然后结合二次和三次加注标签来不断深化自己的研究。

1.3 设计和加注标签的准则

给语料加注标签有两个必须遵守的准则:

- ①对标签内容的解释和理解不能过度地受到人为因素的左右
- ②加注标签者应该能够比较自如地给语料加注标签

人为因素有两类:一类是对词类的划分、句式的认定、构句成分的句法和语义功能的认定等是否会因学者的不同而出现不同的解释,这些解释是否会影响加注标签者对加注标签的对象的分类和认定的判断;另一类是对词类的划分、句式的认定、构句成分的句法和语义功能的认定等是否会因加注标签者的个人理解的不同而出现不同的解释。

如果分类和划分标准多种多样,对需要加注标签的成分的理解就会因人而异。如果加注标签者不熟悉各类学说的标准,即便是一个相同的成分,所加注出来的标签也会出现五花八门的情况,缺乏统一性。

比较自如地给例句加注标签指的是加注标签者应具备一定的语法和语言学的常识,能够在面对有多种划分标准的标签类型时,不过多地受其干扰,并依据一个比较容易遵循的客观标准来给例句加注标签。也就是说,在设计和加注标签时,应该尽量回避和减少使用易受人为因素干扰或难以判断的标签。在日语研究领域里,词类的划分、句式的认定、构句成分的句法和语义功能的认定等都相对地有据可循,容易操作。而语用功能的分辨则会明显地受到语境和说话人表达意图的左右,基本上无据可循,不太容易操作。因此,在设计标签时,需要暂时排除语用类的标签。

1.4 加注标签的方式与标签的内容

加注标签有两种方式:一个是手动加注标签,一个是使用加注标签软件加注标签。手动加注标签指的是采用目视的方法对需要加注标签的对象逐一加注各

类标签。使用加注标签软件加注标签指的是运用装备标签一览表的软件,采用点击一览表中标签的方式来给需要加注标签的对象加注各类标签。

目前带标签的日语语料库几乎看不到。属于个人或团队开发的尚未公开的带标签的语料库也都是用手动来加注标签的。采用手动加注标签时,必须十分熟悉标签的标准和分类,并做到信手拈来,不能凭借临时需要而随意添加标签。因为,加注的标签必须具有统一性和一贯性,即所有的例句都必须按照同一个标准来加注标签。只有这样,检索和统计出来的数据才会具有客观性。

但是,人脑的记忆力是有限的,参与加注标签的如果是一个团队的话,就需要有一个可供所有参与者参照的标签一览表。过去,我们在开发加注标签软件之前,通常都是把标签一览表放在手边,一边查阅和确认,一边加注标签。

标签一览表中的标签必须具有体系性。所谓标签的体系性指的是依据同一类学说设计具有一贯性的标签。标签的设定可以依据各类不同的体系分成不同的标签区域,但使用标签时,不能将各种不同体系的分类不加区别地混同在一起使用,否则会影响结论的可靠性和可验证性,这是需要特别注意的。

但是,即便是大多数人认可的体系实际上也是一种暂时的体系,里面会包含很多不完善的部分,有的地方甚至自相矛盾。个性强的体系虽然只是得到某一个流派或一部分人的认可,尚未普及,但其中也不乏有理有据之处。由此可见,所谓公认的体系和个性化的体系都各有所长和各有所短。

因此,在设计标签的时候,标签的种类应该包含两大类:一类是为基于求大同存小异并基本得到公认的体系所设计的标签,一类是为基于不求大同且个性化强的体系所设计的标签。也就是说,标签一览表中,应该有两大类内容:一类是反映公认体系的标签,一类是反映个性化体系的标签。其中,个性化体系的标签可以有多种。读者可以根据自己需要对使用何类标签进行选择。既可以只选择基于公认体系的标签,也可以只选择基于个性化体系的标签,还可以在选择基于公认体系的标签的同时,选择某一个个性化体系的标签。

要做到这一点,手动加注标签的方式显然比较费力和费时,而且还容易出现遗漏和混同的情况。为了解决这个问题,设计和开发加注标签软件就成了当务之急。

1.5 加注标签软件的类型

上面提到与日语研究和日语教学相关的语料库基本上有3类,即日语语料库、偏误语料库和翻译语料库。这3类语料库的内容和研究目的不尽相同,因此,我们分别设计和开发了日语语料库专用的加注标签软件、偏误语料库专用的加注

标签软件和翻译语料库专用的加注标签软件。

日语语料库专用的加注标签软件指的是给从日语语料库中检索出来的例句加注标签的软件。研究的对象是日语本体,加注标签是为了便于观察和分析日语的语法和语义等问题。

偏误语料库专用的加注标签软件指的是给日语学习者的日语作文加注标签的软件。研究的对象是日语学习者的日语使用问题,加注标签是为了便于观察和分析包括偏误在内的日语使用的规则。

翻译语料库专用的加注标签软件指的是给原文和译文同时加注标签的软件。研究的对象是译文,加注标签是为了便于从信息是否等量这个角度来观察和分析原文表达形式的内含信息与译文表达形式的内含信息的对应情况,由此分析出信息等量、信息不等量、信息部分等量时译文的表达形式,并从中抽取一般化的规则。

1.6 加注标签时所需要的电脑配置和所需软件

加注标签需要使用电脑和一些软件。软件大致可以分为4类:第1类是电脑自带的软件,第2类是根据自己的需要从网上下载的免费软件,第3类是购买的软件,第4类是本书所提供的加注标签的软件。

本书所提供的加注标签的软件是以日文版的 Windows 为基础设计的。只要是安装日文版 Windows 的电脑,并且没有下载国内免费提供的杀毒软件,通常都可以无障碍运行,不会出现死机、无法显示以及乱码等现象。为了便于顺利地加注和分析标签,下面就对电脑以及所需软件的一些最基本的要求做一个说明,以便建立一个良好的平台。

1.6.1 电脑的配置

加注标签的软件所处理的是文字材料,不包括图像和动画,所以无需高配置的电脑。通常,只要满足以下条件就足够了。

- ①操作系统为 Windows® XP 以上(本文以 Windows® 8 为例)
- ②内存为 1GB 以上

现在大多数电脑的内存都已经超过 4GB。内存的大小影响电脑的速度,如果有条件的,建议使用内存在 4GB 以上的电脑。

使用中文版的 Windows 时,有的时候会出现死机、无法显示以及乱码等现

象。这时有两个问题需要确认:一个是中文版的 Windows 是否为正版,不能是试用版或盗版;一个是电脑中是否下载和安装了网上免费提供的杀毒软件。如果 Windows 不是正版,或 Windows 是正版,但下载和安装了免费的杀毒软件,有时候也会因此严重影响加注标签软件的正常使用。这一点请读者务必注意。

1.6.2 所需软件

使用加注标签软件加注标签时,需要以下软件:

- ①Java
- ②秀丸(ひでまる)^①
- ③えだまめ
- ④ひまわり
- ⑤日语语料库专用的加注标签软件
- ⑥偏误语料库专用的加注标签软件
- ⑦翻译语料库专用的加注标签软件

Java、「秀丸」、「えだまめ」、「ひまわり」都是免费下载的软件。这4个软件都是由正式的网站发布的,其中,后两个是日本国立国语研究所研究开发的软件,具有很高的安全性,可以放心下载和使用。有关这4个软件的详细介绍、下载和安装的方法请参照浙江工商大学出版社2013年出版的《语料库的制作与日语研究》(于康著:10—21),这里就不再赘述。

“日语语料库专用的加注标签软件”“偏误语料库专用的加注标签软件”和“翻译语料库专用的加注标签软件”的安装和使用是本书的主要话题,我们将在下面的各个章节里专门进行详细的介绍。

1.7 小结

凡是使用归纳法从事日语研究的,至少需要经过两个步骤,一个是收集例句,一个是分析例句。

收集例句有两个方法:一个是手工收集例句,一个是使用语料库收集例句。现在语料库的研发不断深化,出现了很多语料库。但有些语料库受版权限制,不能公开。于是,学会制作个性化语料库后,例句的收集将会变得更加方便和快捷。

① 正式的名称是「秀丸エディタ」,这里简称为「秀丸」。