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

于康，笔名云丹加措。日本关西学院大学教授、博士生导师。主要致力于日语语言学、汉日语法对比、第二语言习得和日语学习中的误用等领域的研究。

田中良，关西学院大学研究生院博士研究生。主要致力于语料库语言学、社会语言学和机器处理的语言规则及话语规则等研究。

高山弘子，关西学院大学研究生院博士研究生。主要致力于日语语法研究和偏误研究。

■ 方法工具与日语教学研究丛书 I 《语料库的制作与日语研究》

■ 方法工具与日语教学研究丛书 II 《加注标签软件与日语研究》

■ 方法工具与日语教学研究丛书 III 《日语偏误研究的方法与实践》

■ 方法工具与日语教学研究丛书 IV 《日语统计分析软件与日语教材研究》

内容简介

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。

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，2018.7

(方法工具与日语教学研究丛书 / 张威主编)

ISBN 978-7-5178-2860-0

Ⅰ. ①加… Ⅱ. ①于… ②田… ③高… Ⅲ. ①日语－语料库－应用软件－研究 Ⅳ. ①H36－39

中国版本图书馆 CIP 数据核字(2018)第 154690 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　姚　媛
封面设计　林朦朦
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　杭州恒力通印务有限公司
开　　本　710mm×1000mm　1/16
印　　张　58
字　　数　1126 千
版 印 次　2018 年 7 月第 1 版　2018 年 7 月第 1 次印刷
书　　号　ISBN 978-7-5178-2860-0
定　　价　168.00 元(全 4 册)

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

重要声明

（此声明适用于本套丛书所有书籍）

软件下载地址：

http://www.zjgsupress.com/

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

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。