

日语统计分析软件与日语教材研究

本书是『方法工具与日语教学研究丛书』之一。主要介绍和讲解了日语词语和句子长度统计分析工作软件的功能与操作方法，同时还从不同的角度列举和展示了若干运用本书的工具与方法进行学术研究的实际案例。

于康 吕芳 朴丽华 [日]高山弘子 [日]田中良 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

于康，笔名云丹加措。博士。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。

吕芳，博士。关西学院大学国际学部讲师。主要从事日语语言学、汉日语法对比和日语学习中的误用领域的研究。

朴丽华，博士。关西学院大学国际学部非常勤讲师。主要致力于日语语言学和日语学习中的误用领域的研究。

高山弘子，关西学院大学研究生院博士研究生。主要致力于日语语言学和日语学习中的误用领域的研究。

田中良，博士。立命馆大学非常勤讲师。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。

日語

- 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。全书系统地介绍和讲解了日语词语和句子长度统计分析工作软件的功能与操作方法，同时还从不同的角度列举和展示了若干运用本书的工具与方法进行学术研究的实际案例，从而使广大读者能够感同身受地理解学习和掌握这个工具方法的意义和实际应用价值。

方法工具与日语教学研究丛书

丛书主编 张威

日语统计分析软件与日语教材研究

本书是「方法工具与日语教学研究丛书」之一。主要介绍和讲解了日语词语和句子长度统计分析工作软件的功能与操作方法，同时还从不同的角度列举和展示了若干运用本书的工具与方法进行学术研究的实际案例。

于康 吕芳 朴丽华 [日]高山弘子 [日]田中良 著



浙江工商大学出版社

图书在版编目(CIP)数据

日语统计分析软件与日语教材研究 / 于康等著. —
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日… II. ①于… III. ①日语—统计分析—应用
软件—研究 IV. ①H36—39

中国版本图书馆 CIP 数据核字(2018)第 154692 号

日语统计分析软件与日语教材研究

于 康 吕 芳 朴丽华 [日]高山弘子 [日]田中良 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)

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 0571-88904970

卷 首 语

在全球一体化的背景下,日本语教育在世界范围内日趋繁荣,日语语言学与日语教学的研究也在朝着多层次和专业化的方向不断走向深入。为了适应当前社会发展的实际需求,近年来计算机软件研发与数据库统计处理技术越来越多地在教学领域和学术研究中得到应用与推广,并在不断改善和提高学术研究与教学工作的质量与效率方面发挥着不可取代的催化剂作用。

“方法工具与日语教学研究丛书”就是在这样一个变革发展的历史时期,被精心设计打造的一套具有实际应用价值和普及意义的学术系列著作。本书的宗旨是将计算机应用软件领域的最新技术与日语教学研究的实际需求有机地结合起来,强调研究工具与方法的更新和普及。并通过将与时俱进的研究理念及相关的理论知识和工具方法结合起来,以通俗易懂的方式传授给广大读者,进一步拓展日语教学研究的方法与思路,为推动我国日语研究与教学事业做出较大贡献。

自2013年3月本丛书的首卷作品《语料库的制作与日语研究》问世以后,《加注标签软件与日语研究》和《日语偏误研究的方法与实践》也相继出版,并在国内外学界引起广泛的重视和反响,基本上达到了我们预期的效果。今天,本丛书的第四卷《日语统计分析软件与日语教材研究》又要与读者见面了,我们期待这部新书也同样可以受到广大读者的青睐。

《日语统计分析软件与日语教材研究》一书的编写目的是手把手地教会大家如何使用计量统计软件来对材料进行各类统计和解析。本书不仅在系统介绍和讲解日语词语和句子长度统计分析工作软件的功能与操作方法方面下了很大功夫,同时还从不同的角度列举和展示了若干运用本书的工具与方法进行学术研究的实际案例,从而使广大读者能够感同身受地理解学习和掌握这个工具方法的意义和实际应用价值。

学会和掌握统计软件的操作,可以在很短的时间内对大量的数据进行统计,这是手工统计无法比拟的。由于统计软件还可以同时处理不同的文件,这就使得

比较型的统计可以在瞬时完成。此外,统计内容可以依据研究者的思路自由设定,这就意味着所有统计出来的结果都具有原创性和独创性。由此可见,本书对于广大日语教师和学生具有不可忽视的参考意义和实用价值。尤其对于那些因为找不到合适的毕业论文题目或研究课题而纠结和苦恼的学生读者,本书或许可以为你提供“柳暗花明又一村”的灵感,让你从中找到摆脱困境的方向。

我们衷心地期待本书能够成为日语界广大读者的良师益友,为进一步推动我国的日语教学和日语研究起到积极的促进作用。

张 威

2016 年 3 月

重要声明

(此声明适用于本套丛书所有书籍)

本书所含全部软件系统版权均归田中良和于康所有。未经作者许可不得复制、转让、销售、改编或挪作他用。凡因此发生的法律问题由使用者承担,凡由擅自改编程序引发的任何问题作者一概不负责任。

软件下载地址:

<http://www.zjgsupress.com/>

前 言

2013年3月《语料库的制作与日语研究》(于康著,浙江工商大学出版社)的出版拉开了“方法工具与日语教学研究丛书”出版的序幕,此后,2014年3月出版了《加注标签软件与日语研究》(于康、田中良、高山弘子著,浙江工商大学出版社),2014年11月出版了《日语偏误研究的方法与实践》(于康著,浙江工商大学出版社),此次出版的《日语统计分析软件与日语教材研究》(于康、吕芳、朴丽华、高山弘子、田中良著,浙江工商大学出版社)算是我们团队所承担的此系列丛书中的最后一本。

《语料库的制作与日语研究》介绍如何制作日语语料库和如何使用语料库进行日语研究;《加注标签软件与日语研究》推出并介绍了我们团队自己研发的加注日语标签软件、带自动标注正误标签功能的日语作文批改软件、带自动识别语法标签功能的语法标签加注软件、汉日日汉翻译软件,并着重介绍了如何使用这些软件进行日语研究、日语偏误研究和翻译教学研究;《日语偏误研究的方法与实践》推出了最新版的带自动标注正误标签功能的日语作文批改软件、带自动识别语法标签功能的语法标签加注软件和带自动统计功能的偏误语料库软件,并就日语偏误研究的方法和实践进行了详细说明。如果说上述三本书主要用于日语研究、日语偏误研究和汉日日汉翻译研究,那么本书《日语统计分析软件与日语教材研究》所介绍的日语词语和句子长度统计分析软件,不仅可以用于日语语言研究,还可以用于日语教材研究以及社会语言学、文学、社会学、政治学等领域的研究。本书写作的具体分工为:于康负责软件构思,田中良负责软件制作。第1章、第2章、第3章、第4章和第5章的作者为于康,第6章的作者为朴丽华,第7章的作者为吕芳,第8章的作者为高山弘子。

我们研发的软件没有去申请专利,而是选择了免费赠送给每位购书的读者,目的是为了积极推动国内的日语教学研究。希望我们这套书能够对各位读者有所帮助,也非常感谢各位读者所给予的鼓励和支持。

于 康
2016年3月

目 录

第 I 部 工具与方法

第 1 章 日语词语和句子长度的统计分析	3
1.1 日语词语和句子长度的统计分析所需的工具	3
1.2 2 个软件的主要功能	4
1.2.1 TNR_Support_0.8.0	4
1.2.2 GetSentLeng_PosFreq_GUI_Ver.4	5
1.3 日语词语的统计分析	9
1.3.1 主要用途	9
1.3.2 7 种用途的具体举例说明	10
1.4 日语句子长度的统计分析	34
1.4.1 主要用途	34
1.4.2 7 种用途的具体举例说明	34
1.5 小结	42
第 2 章 电脑配置与所需软件	44
2.1 日语词语和句子长度的统计分析的主要程序	44
2.2 电脑配置	44
2.3 所需软件	45
2.4 所需软件的下载与安装	46
2.4.1 TNR_Support_0.8.0 和 GetSentLeng_PosFreq_GUI_Ver.4 的 安装	46
2.4.2 「秀丸エディタ」的下载与安装	47
2.5 软件的卸载	53
2.6 小结	53

第 3 章 TNR_Support_0.8.0 和「秀丸エディタ」与无关信息处理	55
3.1 语料的收集和各種語料的特点	55
3.2 『青空文庫』和『国会會議録検索システム』的語料下載	56
3.2.1 『青空文庫』的語料下載	56
3.2.2 『国会會議録検索システム』的語料下載	56
3.3 需要處理的語言信息	57
3.3.1 『青空文庫』中需要處理的語言信息	57
3.3.2 『国会會議録検索システム』中需要處理的語言信息	58
3.3.3 合併文件與一句一行	60
3.4 TNR_Support_0.8.0 与信息處理	61
3.4.1 刪除『青空文庫』中的无关信息	61
3.4.2 將文章的内容改為一句一行	64
3.4.3 合併文章	67
3.5 「秀丸エディタ」与信息處理	70
3.5.1 刪除會議的標題、人員名單和有關會議議程的說明	70
3.5.2 刪除「〔「異議なし」と呼ぶ者あり〕」等有關會場情況的說明	71
3.5.3 刪除文中的「○○○○休憩」「午後○○時開議」	73
3.5.4 行與行之間的裝飾性符號	76
3.5.5 文字前附加的各種符號	81
3.5.6 文章中「○○委員長」「○○委員」等文字的處理	84
3.6 小結	85
第 4 章 日語詞語和句子長度統計分析軟件的使用方法	86
4.1 軟件拷貝與主畫面中各類按鈕的主要功能	86
4.2 建立新的文件夾與保存統計分析用的文章	87
4.3 選擇統計分析用的文章	89
4.4 選擇文章的上限數	91
4.5 詞語使用的統計分析與主要步驟	92
4.5.1 單個作者單篇文章中詞語使用情況的統計分析方法	92
4.5.2 其他類型的詞語使用情況的統計分析方法	101
4.6 句子長度的統計分析與主要步驟	102
4.6.1 單個作者單篇文章中句子長度使用情況的統計分析方法	102
4.6.2 其他類型的句子長度使用情況的統計分析方法	106

4.7 小结	106
--------------	-----

第Ⅱ部 实证研究

第5章 中日韩三国四方大学生“人生设计”分析	109
------------------------------	-----

5.1 各种词类的选择使用情况与使用率比较分析	110
5.2 格助词的使用特征	113
5.2.1 中国大学生与日本大学生的格助词选择比较	114
5.2.2 三国四方大学生与日语教材的格助词使用情况比较	115
5.3 「VPと思う」的选择使用分布	117
5.4 「結婚」类词语的使用情况与未来人生设计	119
5.4.1 「結婚」类词语的使用情况分布	120
5.4.2 「家庭」的例句与分析	121
5.4.3 「結婚」的例句与分析	124
5.4.4 「子供」的例句与分析	129
5.5 小结	134

第6章 日语教材分析	137
------------------	-----

6.1 收集语料	137
6.2 统计分析日语教材中的词语	141
6.2.1 日语教材中词类的整体使用情况	141
6.2.2 关于教材中人称代词「私/わたし」的使用	145
6.3 统计分析日语教材的句子长度	150
6.3.1 句子的平均长度	150
6.3.2 各册教材中句子长度的分布情况	153
6.4 小结	156

第7章 作家作品分析	157
------------------	-----

7.1 词语的统计与分析	157
7.1.1 相同时期不同作家作品中的词语统计与分析	157
7.1.2 同一题材不同时期作品中的词语统计与分析	164
7.2 句子长度的统计与分析	172
7.2.1 相同时期不同作家作品中句子长度的统计与分析	172
7.2.2 同一题材不同时期作家作品中的句子长度的统计与分析	174
7.3 小结	176

第 8 章 历代首相施政讲演分析	178
8.1 時代による表現と演説スタイルの変化	179
8.1.1 文の長さから見る演説の傾向性	180
8.1.2 文末の助動詞、動詞から見る演説の傾向性	183
8.1.3 演説スタイルの変化	193
8.2 安倍首相に見られる表現と演説スタイルの変化	193
8.2.1 文末表現から見る変化	195
8.2.2 名詞、接尾辞から見る変化	199
8.2.3 演説スタイルの変化	200
8.3 まとめ	201
参考文献	203
后 记	205

第 I 部 工具与方法

第1章 日语词语和句子长度的统计分析

重要提示：

1. 日语词语的统计分析指的是使用「TNR_語彙形態素と文の長さを分析する日本語コンコーダンサー 2015」Ver. 3 来统计分析文章中所有词类的使用情况,以此计算出百分比,并统计分析不同词类中词语的使用次数和排位顺序。
2. 日语句子长度的统计分析指的是使用「TNR_語彙形態素と文の長さを分析する日本語コンコーダンサー 2015」Ver. 3 来统计分析文章中句子长度的使用情况。
3. 统计分析日语词语与句子长度时,需要使用以下两个软件。TNR_Support_0.8.0 是一个辅助软件,专门用来处理文章中的无关信息。GetSentLeng_PosFreq_GUI_Ver. 4 是专门用来统计分析日语词语与日语句子长度使用情况的软件。

日语词语与句子长度的统计分析指的是使用「TNR_語彙形態素と文の長さを分析する日本語コンコーダンサー 2015」Ver. 3^① 来统计分析文章中所有词类的使用情况和句子长度的使用情况。

1.1 日语词语和句子长度的统计分析所需的工具

日语词语与句子长度的统计分析共需 2 个工具,即 2 个软件。软件的名称如下:

①TNR_Support_0.8.0

②GetSentLeng_PosFreq_GUI_Ver. 4

TNR_Support_0.8.0 是关西学院大学研究生院于康研究室的田中良设计和制作的软件,GetSentLeng_PosFreq_GUI_Ver. 4 是于康策划和设计,由田中良制

① 中文名为“日语词语和句子长度统计分析软件 2015”。本软件中所使用的词素分析词典 ChaSen(茶筌)是由奈良先端科学技术大学松本研究室研制开发的。词类的名称和词素的切分都将遵循 ChaSen 的规定,所以,有些术语和词素切分可能会使读者产生疑问。这个问题不是本书作者能够控制的,而且,现在可供使用的任何一种词素分析词典都无法达到 100% 的精确度,误差是在所难免的,读者在使用分析结果时需要对分析结果进行核实和一些技术上的处理。