

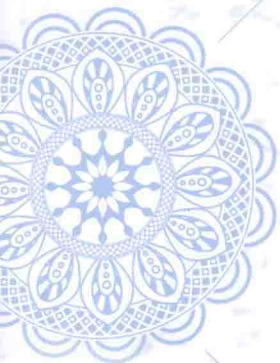
[瑞典] Aarne Ranta ○ 著
田 艳 ○ 译

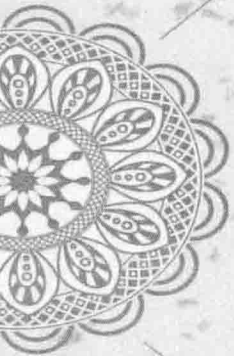
Grammatical Framework
Programming with Multilingual Grammars

语法框架 为多种自然语言语法编程



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS





[瑞典] Aarne Ranta 著

田 艳 译

乔海燕 审校

Grammatical Framework
Programming with Multilingual Grammars

语法框架★

为多种自然语言语法编程



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS



内容提要

语法框架是一种计算机编程语言,专门为编写自然语言的语法而设计,它有能力并行处理多种自然语言。书中全面介绍了如何利用语法框架为自然语言编写语法,以及如何在旅游手册、口语对话系统和自然语言处理系统等实用系统中加以应用。书中的例子和练习涉及多种自然语言,读者可以从中学习如何从计算语言学的视角看待自己的母语。

阅读本书不需要语言学基础知识,因此,特别适合计算机科学家和程序员使用。此外,本书从程序语言理论的视角展示了处理多种自然语言语法的新途径,因此,语言学家也会对本书产生兴趣。

© by CSLI Publications, Leland Stanford Junior University.
Translated and published by arrangement with CSLI Publications.
All rights reserved.

图书在版编目(CIP)数据

语法框架 / (瑞黄)兰塔(Ranta, A.)著;田艳译. — 上海:上海交通大学出版社,2014
ISBN 978-7-313-12214-8

I. 语... II. ①兰... ②田... III. 自然语言处理 IV. TP391

中国版本图书馆CIP数据核字(2014)第239268号

语法框架

——为多种自然语言语法编程

著者:[瑞典]Aame Ranta	译者:田艳
出版发行:上海交通大学出版社	地址:上海市番禺路951号
邮政编码:200030	电话:021-64071208
出版人:韩建民	
印制:常熟市梅李印刷有限公司	经销:全国新华书店
开本:787mm×960mm 1/16	印张:15
字数:280千字	
版次:2014年10月第1版	印次:2014年10月第1次印刷
书号:ISBN 978-7-313-12214-8/TP	ISBN 978-7-89424-076-6
定价(含光盘):48.00元	

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:0512-52661481

中文版序言

One of the most pleasant surprises in my life was in December 2011, when I received a mail from Professor Yan Tian of Shanghai Jiao Tong University, proposing to translate the GF book to Chinese. Wow, I thought, this would almost double the potential audience of my book! Less than a year later, this collaboration spawned another pleasure, which was to visit Professor Tian and her colleagues in Shanghai. She then showed me a complete manuscript of the translation, of which I unfortunately understood next to nothing. But we had long sessions discussing the translation of the terminology, which convinced me that the book is being produced in a serious way. We also had a tutorial course and technical discussions, testing and assessing the newly released Chinese resource grammar, and looking at future project opportunities. Professor Tian suggested that, in order to make the book genuinely interesting to the Chinese audience, we should add an Appendix that shows what GF means for Chinese. Therefore, I wrote the appendix specifically for this Chinese version. It gives Chinese examples and code as a commentary to the main book, thus following its order of presentation. It is probably hard to read without access to the book.

I am grateful to Professors Yan Tian and Yinglin Wang, as well as Li Peng, for their hospitality in Shanghai, to Jyrki Nummenmaa from Tampere for starting this collaboration and guiding me in Shanghai, to Dr. Chen Peng from Beijing for also joining the discussions in Shanghai and contributing to the Chinese resource grammar, and to Dr. Qiao Haiyan from Guangzhou, who wrote the first Chinese GF code—still in use in the resource grammar—back in 1999. The most demanding part of work with the Chinese resource grammar was made by Jolene Zhuo Lin Qiqige, as a part of a Masters course at the University of Gothenburg.

Gothenburg, November 2012

Aarne Ranta

中文版序言(译文)

我生命中最令人愉快的惊喜之一是 2011 年 12 月收到了来自上海交通大学田艳教授的邮件,提议要把《语法框架》翻译成汉语。哇,我想,这几乎会使本书的潜在读者增加一倍!不到一年时间,这一合作又带给我另外一段愉快经历:我在上海造访了田艳教授和她的同事。她给我看了翻译完成的书稿,遗憾的是,我几乎什么也看不懂。但是我们长时间讨论了一些术语的翻译问题,这使我相信本书的翻译极为认真。我们也开了辅导课,讨论了一些技术问题,测试并评估了新发布的中文资源语法,并展望了未来项目合作的可能性。田教授建议:为了使中国读者对本书更感兴趣,我们应该增加一个附录,展示语法框架对汉语的意义。因此,我专门为中文版写了附录,给出了一些汉语例子和代码,作为对原书的补充,附在书后。如果没有事先阅读全书,可能很难读懂专为中文版所写的这个附录。

我对在上海受到的热情款待向田艳教授和王英林教授以及李鹏博士致以衷心地感谢!对发起此次合作并在上海陪同我的芬兰坦布尔大学的杰瑞基·那门马表示感谢!对从北京到上海参加此次讨论并贡献了中文资源语法的陈鹏博士表示感谢!也对在 1999 年就写出了第一个且始终在资源语法中使用的中文语法框架代码的广州的乔海燕博士表示衷心感谢!中文资源语法中最需要付出艰辛劳动的部分是由卓琳其其格完成的,那是她在哥本哈根大学硕士学位论文的一部分。

阿尔内·兰塔

2012 年 11 月于哥德堡

译者序

第 23 届国际计算语言学大会 (The 23rd International Conference on Computational Linguistics) 的主题是“规则与统计共舞, 语言与计算齐飞” (Computational Linguistics—Rules dance with numbers, language soars with information)。在自然语言处理界, 统计方法占据主导地位的今天, 规则方法是否仍然可以发挥其应有的作用?

众所周知, 自然语言的语法对于语言学习者, 尤其是成年语言学习者的重要性。因为语法是语言的一般规则, 掌握了语法, 就可以更快、更好地学习一门语言, 也可以减少大量的具体操练。掌握了语法, 还可以提高语言学习者语言产出的质量, 与一个仅仅通过直接接触语言学习语言的人相比, 其在写作与口语中所犯的错误也会更少。

语法为语言学习者提供了一条捷径。机器处理自然语言同样如此, 在处理比英语结构更为复杂的自然语言时更是如此。译者非常赞同作者的观点: “知道的, 就不要去猜。”

Aarne Ranta 教授创建的语法框架 (Grammatical Framework, 简称 GF) 是为多种自然语言的语法编写规则的一种编程语言。自 1998 年问世以来, 语法框架系统不断发展, 并且逐年增加新的功能。语法框架编程语言的通用语法始于 2000 年, 通过手写或从其他语言资源中自动生成方式, 其资源语法库及其所有的扩展已经超过 50 万行语法框架代码, 其应用已包括了上百万行代码。资源语法库已增加到包括汉语在内的 26 种自然语言。

《语法框架》一书以教材的形式编写, 涵盖了语法框架的基本理论和实际应用。正文中有全面的语法框架辅导章节, 附录中有详细的语法框架参考手册。读者不需要有很深的背景知识, 也不一定按照专著的顺序, 而只需按自己的需要研读。《语法框架》还有专门的网站, 为实际操作提供更多的材料和支持。

衷心希望《语法框架》中译本的出版能促进我国自然语言处理, 尤其是中文信息处理研究的发展, 使规则方法能够有机地与统计方法相结合, 提高自然语言处理的质量。

Aarne Ranta 教授在《语法框架》翻译的过程中给予译者热情的鼓励和大力帮助。他专门为中译本撰写了序言和针对中文处理的附录。中文处理的附录中所举的例子可以帮助中文读者更好地了解《语法框架》在汉语语法编写中的运用。中山大学计算机系语法框架方面的专家乔海燕博士在百忙之中审校了全部

译稿,并进行了非常专业的修改。瑞典哥德堡大学语言技术中心(The Centre for Language Technology of the University of Gothenburg)在出版资金上鼎力支持。国家自然科学基金项目:“用户异构需求间的半自动化映射及需求缺失识别方法研究”(编号:61375053)也提供了帮助。

虽然译者多年从事计算语言学研究,然而,本书的翻译仍然是一个很大的挑战。书中存在的不妥之处,敬请读者不吝指正。

译者
2014年10月

前 言

语法框架是为编写自然语言的语法而设计的一种专用编程语言,它可以处理各种复杂的自然语言,也可以为需要许多程序员共同参与的大型工程项目提供编程工具。此外,它还支持仅在单一自然语言内部以及跨多种自然语言的抽象和代码共享。在语法框架走过的 12 年间,它已被广泛应用于多语言翻译系统、基于语言的人机交互系统以及创建计算语言学资源等方面。

语法框架是基于类型化函数式编程语言 Haskell 和 ML(元语言)的现代设计,旨在供不同背景的使用者,尤其是没有经过语言学训练的使用者使用。由于语法框架支持强大的概括和多语种应用,它也会让理论语言学家产生兴趣。因此,语法框架可供计算机科学家、语言学家、工程师、数学家、学生以及资深专业人士使用。

本书旨在汇集语法框架的全部内容以便于高效使用,其中包括了从零起点开始并涵盖了语法框架所有特征的辅导部分、一系列有关语法框架实际应用和高级语法编写的章节,以及勾画了语法框架全貌的参考手册。书中的有些内容曾经在 50 余种出版物以及语法框架网站上发表,但对于语法框架使用者来说,所遇到的困扰一直是无法在一个地方找到关于语法框架的全部内容。因此,我们希望本书的出版不但对于语法框架的初学者,而且对于经验丰富的语法框架程序员也有所帮助。

作 者
2011 年 2 月

致 谢

早在 2007 年,我使用了本书的早期版本给瑞典语言技术研究生院的 15 名研究生上了语法框架的课程。2009 年修订后,发给了 30 名参加“2009 年语法框架暑期班”的学生和老师们,他们随即开始使用语法框架对 16 种新的自然语言进行编程。德国慕尼黑大学的 Hans Leiß 和西班牙巴塞罗那技术大学的 Sebastià Xambó 也使用过暑期班的版本。我对使用过语法框架并给予我反馈的所有老师和学生以及同事们表示诚挚的谢意!尤其需要感谢的是 Krasimir Angelov 和 Björn Bringert,因为他们实际上撰写了书中的部分内容。此外,Håkan Burden、Olga Caprotti、Ramona Enache、Julia Hammar、Arto Mustajoki、Bengt Nordström、Jordi Saludes 以及 Annie Zaenen 曾对我手稿的多个版本提出了非常有价值的意见。我还要特别感谢 Ann Copestake,她作为斯坦福大学“语言与信息研究中心”(CSLI)系列丛书的编辑和语法形式化方面的著名同行专家给予了我富有远见卓识的帮助。Emma Pease 为本书撰写了非常专业的结尾。

语法框架软件是一个开源的合作项目,代码由(排序按时间先后)Petri Mäenpää、Thomas Hallgren、Janna Khagai、Markus Forsberg、Kristofer Johannisson、Peter Ljunglöf、Håkan Burden、Hans-Joachim Daniels、Björn Bringert、Krasimir Angelov、Moisés Salvador Meza Moreno、Kevin Kofler、Jordi Saludes、John Camilleri、Ramona Enache 和 Grégoire Détrez 提供。

语法框架资源语法库是语法框架的主要精华之一,对此做出重要贡献的有:Janna Khagai、Markus Forsberg、Inger Andersson、Thérèse Söderberg、Anni Laine、Jordi Saludes、Harald Hammarström、Ali El Dada、Jean-Philippe Bernardy、Andreas Priesnitz、Krasimir Angelov、Ilona Nowak、Adam Slaski、Ramona Enache、Muhammad Humayoun、Shafqat Virk、Server Qimen 和 Markos Kassa Gobena。

除了上述对语法框架做出直接贡献的人士外,语法框架的语言和理论也受益于类型论、函数编程和语言学等已知理论和方法。此外,在我与许多同事的个人交流中,我也获得了很多重要的设计灵感。这些同事是(排序仍按时间先后):Per Martin-Löf、Thierry Coquand、Bengt Nordström、Lena Magnusson(后来改为 Pareto)、Gilles Kahn(†2006)、Rod Burstall、Marc Dymetman、Sylvain Pogodalla、Veronika Lux、Lauri Karttunen、Patrik Jansson、Koen Claessen、Robin Cooper、Gérard Huet、Philippe de Groote、Sibylle Schupp、Lauri Carlson、Hans Leiß 和 Elisabet Engdahl。

最后,我想感谢我的亲人们:Pihla、Eemu、Uula、Aamos 和 Luukas,感谢他们陪伴我度过的美好时光,并常常使我感到更加幸福和睿智。

2011 年 2 月于哥德堡

阿尔内·兰塔

aarne@chalmers.se

网上资源和语法框架系统

辅助资料可以在本书的网站上找到：

<http://www.grammaticalframework.org/gf-book>

资料包括：

- 书中的大部分代码实例
- 书中参考手册中给出的链接
- 供课堂上使用的幻灯片
- 一些练习的答案

本书的网站托管于语法框架的主网站：<http://www.grammaticalframework.org>。除其他内容外，该主网站还含有：

- 带有链接的更新后的参考手册
- 完整的资源语法库应用程序界面
- 语法框架系统和资源语法库源代码
- 语法框架系统可执行的二进制文件

鼓励读者下载和安装语法框架系统并尝试代码示例和习题。本系统适合包括 Linux、Mac OS 和 Windows 的所有主要计算机平台。获得本系统最容易的途径是通过如下的链接：<http://www.grammaticalframework.org/download>。

从此链接可以获得语法框架的最新版本，只需下载和解压，就可得到现成的可执行文件。语法框架 3.2 版本经测试，证明与本书完全匹配。今后的版本也力图与之前的版本兼容，只增加新的功能而并不清除原有功能。

3.2 版本可通过本书的网站获得，包括资源、在线文件和适合 Linux、Mac OS 和 Windows 操作系统的二进制文件。如果要使用其中的一种二进制文件，只需下载和解压，并放在你的可执行程序的路径上即可。

目 录

第 1 章 导论	(1)
1.1 本书内容	(1)
1.2 如何使用本书	(2)
1.3 语法在语言处理中的作用	(3)
1.4 语法成本	(6)
1.5 多语性	(8)
1.6 语义动作和互操作性	(12)
1.7 应用语法和资源语法	(14)
1.8 GF 的历史及其应用	(18)
1.9 相关工作	(19)

第 I 部分 语法框架辅导

第 2 章 多语种语法的基本概念	(22)
2.1 BNF 语法格式	(22)
2.2 GF 系统的使用	(24)
2.3 在 GF 系统中测试一个语法	(25)
2.4 用于意大利语的 BNF 语法	(28)
2.5 BNF 语法和翻译	(28)
2.6 抽象句法与具体句法	(29)
2.7 在 GF 里翻译	(32)
2.8 语法模块结构	(34)
2.9 BNF 语法的局限性	(35)
2.10 阻止和元变量	(36)
2.11 自由变异	(37)
2.12 歧义	(38)
2.13 剩余问题	(39)
2.14 图示形象化和命令行解释器转义	(40)
2.15 词法分析和反词法分析	(42)
2.16 字符编码	(43)

第 3 章	参数、表格和记录	(45)
3.1	形态变异问题	(45)
3.2	参数和表格	(46)
3.3	变量特征与固有特征	(48)
3.4	记录与记录类型	(49)
3.5	线性化类型与一致	(49)
3.6	GF 中的函数编程:运算定义	(51)
3.7	Food 语法的再探讨	(53)
3.8	在 GF 里测试词形变化和运算	(55)
3.9	部分调用	(55)
3.10	非连续成分	(56)
3.11	非衔接词法	(58)
第 4 章	可模块化与可扩展的语法编写	(60)
4.1	可复用的资源模块	(60)
4.2	数据抽象	(63)
4.3	分支表达式和字符串匹配	(65)
4.4	智能词形生成函数	(66)
4.5	阿拉伯语词法回顾	(67)
4.6	区分运算类型和定义	(67)
4.7	运算重载	(68)
4.8	模块扩展和继承	(69)
4.9	继承和打开	(72)
4.10	依存图	(73)
4.11	参数的代数数据类型	(73)
4.12	记录扩展和子类型	(75)
4.13	元组与乘积类型	(76)
4.14	前缀依存选择和模式宏	(76)
4.15	编译时字符串和运行时字符串	(77)
第 5 章	使用资源语法库	(79)
5.1	库的目的和覆盖范围	(79)
5.2	词典和短语规则	(80)
5.3	词汇范畴和规则	(81)
5.4	短语范畴和规则	(82)
5.5	资源 API	(83)
5.6	库的路径	(84)

5.7	示例:英语	(87)
5.8	多语言语法的函子实现	(87)
5.9	界面和实例	(89)
5.10	多语言语法的设计模式	(90)
5.11	分工再探讨	(91)
5.12	重写函子	(94)
5.13	编译时转移	(95)
5.14	作为语言本体的资源语法	(96)
5.15	资源 API 概览	(97)
5.16	结构扁平化	(99)
5.17	时态和极性	(100)
5.18	浏览资源语法库	(101)
第 6 章	抽象句法里的语义动作和条件	(104)
6.1	作为逻辑框架的 GF	(104)
6.2	依存类型	(105)
6.3	选择限制	(106)
6.4	多态性	(109)
6.5	具体句法中的依存类型	(109)
6.6	证明对象	(110)
6.7	证明携带文件	(111)
6.8	受限多态性	(112)
6.9	约束变量与高阶抽象句法	(114)
6.10	照应表达式	(116)
6.11	语义定义	(118)
6.12	内涵等式和外延等式	(120)
6.13	语义动作和运行时转换	(121)
6.14	预定义类	(123)
6.15	概率 GF 语法	(124)

第 II 部分 更大规模语法及其应用

第 7 章	嵌入语法与代码生成	(126)
7.1	可移植语法格式	(126)
7.2	嵌入解释器及其 API	(127)
7.3	Haskell 中的嵌入 GF 应用	(127)
7.4	PGF 模块	(128)

7.5	独立翻译器	(128)
7.6	翻译器循环	(130)
7.7	答疑系统	(130)
7.8	GF 数据类型的输出	(132)
7.9	将其全部放在一起	(134)
7.10	网络服务器应用程序	(135)
7.11	其他宿主语言的嵌入语法	(136)
7.12	多语言句法编辑	(137)
7.13	语音识别的语言模型	(139)
7.14	统计语言模型	(141)
7.15	多模态对话系统	(141)
第 8 章	形式语言与自然语言接口	(144)
8.1	算术表达式	(144)
8.2	作为线性化的代码生成	(147)
8.3	带变量的程序	(148)
8.4	赋值的具体句法	(149)
8.5	变量的自由句法	(150)
8.6	GF 对于定义形式语言是有用的吗	(150)
8.7	由逻辑生成自然语言	(151)
8.8	自然语言的逻辑语义学	(156)
8.9	分形语法	(160)
第 9 章	开始资源语法编程	(161)
9.1	概览	(161)
9.2	微型资源语法	(162)
9.3	特征设计	(163)
9.4	谓项	(163)
9.5	互补性	(165)
9.6	限定	(166)
9.7	修饰	(167)
9.8	词汇插入	(168)
9.9	意大利语中的微型资源	(169)
9.10	实现词法	(170)
9.11	实现修饰和限定	(171)
9.12	实现动词短语和互补	(174)
9.13	实现谓项	(174)

9.14	实现剩余的部分	(176)
9.15	并列和抽取	(177)
第 10 章	扩展资源语法库	(179)
10.1	资源语法的模块结构	(179)
10.2	所需工作量统计	(181)
10.3	一种新语言的工作流程	(182)
10.4	重用微型资源的代码	(183)
10.5	研发 - 测试周期	(184)
10.6	非 ASCII 字母和转写	(185)
10.7	编码原则	(186)
10.8	资源语法中的函子	(187)
10.9	扩大句法分析文本的覆盖范围	(190)
10.10	引导一个资源词典	(191)
附录 A	小型资源语法^①	(附 1)
A.1	抽象句法	(附 1)
A.2	意大利语助动词资源模块	(附 2)
A.3	意大利语具体句法	(附 6)
A.4	意大利语词法词形变化表 API	(附 8)
A.5	测试词典	(附 9)
A.6	语法 API	(附 10)
附录 B	语言学术语词汇表	(附 12)
第 III 部分 语法框架参考手册		
附录 C	语法框架编程语言	(附 15)
C.1	GF 概览	(附 15)
C.2	模块系统	(附 16)
C.3	判断	(附 23)
C.4	类型与表达式	(附 29)
C.5	标志与编译指示	(附 40)
C.6	GF 语法	(附 41)
附录 D	语法框架资源语法库	(附 47)
D.1	范畴系统	(附 47)
D.2	语法规则	(附 50)

① 附录 A ~ 附录 F 的内容在随书的光盘中。

D.3 词汇范例	(附 54)
D.4 其他库模块	(附 55)
附录 E 语法框架软件系统	(附 57)
E.1 GF 命令解释程序	(附 57)
E.2 GF 批处理编译器	(附 59)
附录 F 中文版附录	(附 61)
参考文献	(194)
索引	(206)