计算机辅助翻译

Computer Assisted Translation

杨 蔚◎主编

同济大学出版社
TONGJI UNIVERSITY PRESS

计算机辅助翻译

杨 蔚◎主编

同济大学出版社
TONGJI UNIVERSITY PRESS

内 容 提 要

本书介绍计算机辅助翻译的基本知识与核心概念，根据目前翻译行业的业务特征，详细讲述网站、软件、联机帮助等翻译文本类型的流程与特点，介绍专业及通用辅助翻译工具的使用，目的在于帮助学生适应信息化社会及翻译行业对他们专业知识和技能提出的挑战。

图书在版编目（CIP）数据

计算机辅助翻译 / 杨蔚主编. —上海：同济大学出版社，2018.1

ISBN 978-7-5608-7673-3

Ⅰ. ①计… Ⅱ. ①杨… Ⅲ. ①自动翻译系统—研究 Ⅳ. ①TP391.2

中国版本图书馆CIP数据核字（2017）第330254号

计算机辅助翻译

杨 蔚 **主编**

责任编辑 刘 睿 魏国旺 **责任校对** 徐春莲 **封面设计** 文 一

出版发行 同济大学出版社 www.tongjipress.com.cn

（地址：上海市四平路1239号 邮编：200092 电话：021-65985622）

经 销 全国各地新华书店

印 刷 虎彩印艺股份有限公司

开 本 787mm×1092mm 1/16

印 张 13.25

字 数 260千字

版 次 2018年1月第1版 2018年1月第1次印刷

书 号 ISBN 978-7-5608-7673-3

定 价 39.00元

本书若有印装质量问题，请向本社发行部调换 版权所有 侵权必究

前　言

翻译是人类社会生活沟通与交融的必要工具，技术融入翻译活动是社会发展的必然。虽然实现全自动、高质量且无需人工介入的机器翻译目前尚未达到理想效果，由计算机软件作为辅助工具，实现人译机助却在近年来得到了长足的发展，并且带动了翻译市场的繁荣。计算机辅助翻译能力如今已成为译员的必备技能之一。

本书介绍了计算机辅助翻译的基本知识与核心概念，根据目前翻译行业的业务特征，详细讲述了网站、软件、联机帮助等翻译文本类型的流程与特点，介绍专业及通用辅助翻译工具的使用，目的在于帮助学生适应信息化社会及翻译行业对他们专业知识和技能的挑战。学习本书后，学生应能做到以下几点：

熟悉语料库语言学基本原理，了解计算机辅助翻译的核心概念。

学会创建翻译记忆库与术语库，学会借助各类CAT工具从事翻译工作。

熟悉机器翻译和机器辅助翻译的工作机制，对主要的机译和机助翻译系统有专业认识。

学会使用工具软件进行技术文档、网站、软件翻译、联机帮助等翻译工程。

本书共有十个章节，可供一学期教学使用，适用于英语翻译专业本科及硕士生（MTI）教学及本科翻译专业教学。学生需具备较好的计算机基础及入门水平以上的计算机使用能力。每一章的最后附有思考与练习，供教师或学习者参考使用。

作者衷心感谢南京理工大学外国语学院赵雪琴院长，2012年4月支持我参加中国译协主办的“全国高等院校翻译专业师资翻译与本地化技术培训”，是我第一次接触计算机辅助翻译技术，并在同年5月鼓励我开设MTI专业课程“计算机辅助翻译”，建设计算机辅助翻译实验室，从此与翻译技术结下了不解之缘。还要感谢2012年4月南京技术培训班的各位老师，崔启亮博士、王华树博士、曾立人博士和闫栗丽老师，各位老师的无私传授让我能够站在一个很高的起点上跨入计算机辅助翻译与本地化的大门。感谢南京理工大学研究生院，本教材的编写与出版离不开研究生创新工程项目的大力支持。

由于作者水平有限，加之翻译技术的发展与行业的发展日新月异，疏漏与不足在所难免，敬请读者指正。

杨　蔚

二〇一七年七月

南京理工大学竹园

目　录

试读结束：需要全本请在线购买：　www.ertongbook.com

第一章　绪　论

第一节　翻译与技术

一、概述

翻译是人类社会生活沟通与文化交流的必要工具，翻译活动与人类社会共同发展。一般认为西方最早的译作出现在公元前3—公元前2世纪之间，包括72名犹太学者在埃及亚历山大城翻译的《圣经·旧约》，即《七十子希腊文本》，以及公元前3世纪中叶安德罗尼柯在罗马用拉丁语翻译的希腊《荷马史诗》《奥德赛》。而中国的翻译始于西汉哀帝时期的佛经翻译。无论中西，人类翻译活动至少已有2000多年的历史。在这漫长的历史进展中，翻译工具亦随着社会的发展不断变化更新。从古代的笔墨纸砚到近代的纸笔，再到如今的计算机应用，翻译活动的发展与科技的发展密不可分。

几乎从计算机诞生之日起，人们就试图利用计算机来进行自然语言的翻译工作。最初的目标是不需人工介入的、全自动的、高质量的机器翻译（FAHQT），然而至今未能达到理想的效果。在曲折的机器自动翻译发展过程中，出现了机译人助翻译（HAMT）、人译机助翻译（MAHT）和计算机辅助翻译（CAT）等形式（图1-1）。

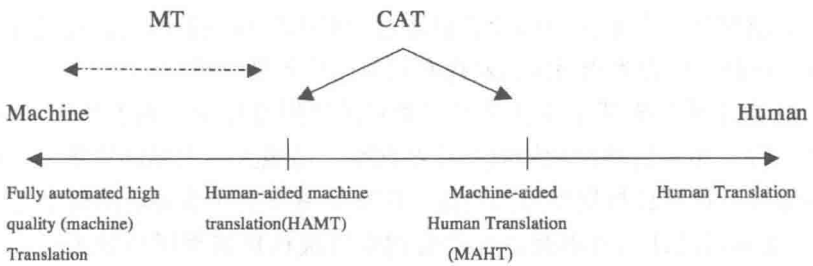


图1-1　翻译形式分类（Hutchins & Somers, 1992）

二、机器翻译（Machine translation）

使用机器完成翻译的最初目标是不需人工介入的、全自动的、高质量的机器翻译（FAHQT），但是在1960年，Bar-Hillel提出，这样的目标不切实际，因为全自动翻译与高质量翻译之间存在冲突，要么牺牲翻译质量，要么降低机器翻译中全自动的比重。如果保证译文的高质量是唯一目的，则机器翻译产出的译文必须经过译后编辑（post-editing）。在这样的

情况下，机器翻译实际上成为了机器协助下的翻译（Machine aids to translation）。

Bar-Hillel的观点得到了研究者的认同。机器翻译研究虽未因此而停止，但目的有了些许调整，即“自动生成符合翻译目的而非高质量的翻译”（Quah, 2006），也就是现在通常认为的机器翻译（Machine Translation），“利用计算机程序完成一种自然语言到另一种自然语言的文本翻译任务[①]”。

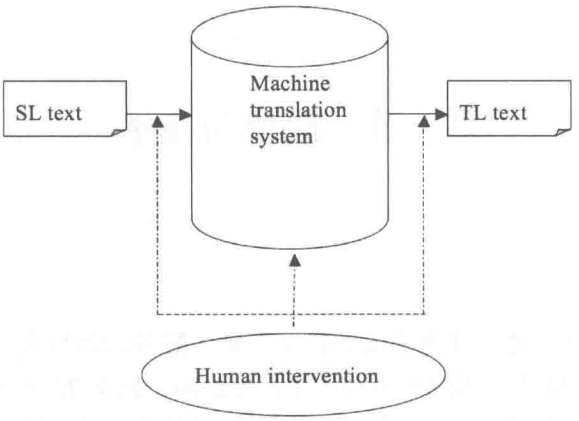


图 1-2　机器翻译模式（Quah, 2006）

第一代机器翻译系统设计原理为基于规则的机器翻译（Rule-based MT）。1954 年美国乔治敦大学与IBM公司共建的机器翻译系统在六条语法规则的基础上将 250 个单词由俄语翻译为英语，引起世界范围内研究机器翻译的热潮。但是，基于语言规则的机器翻译研究很快陷入困境，因为无法完全依靠规则和词汇表来分析真实语境中千变万化的语句，甚至规则本身也在变化之中。只有在一些语法规则高度统一的专业领域，基于规则的机器翻译有较成功的例子。

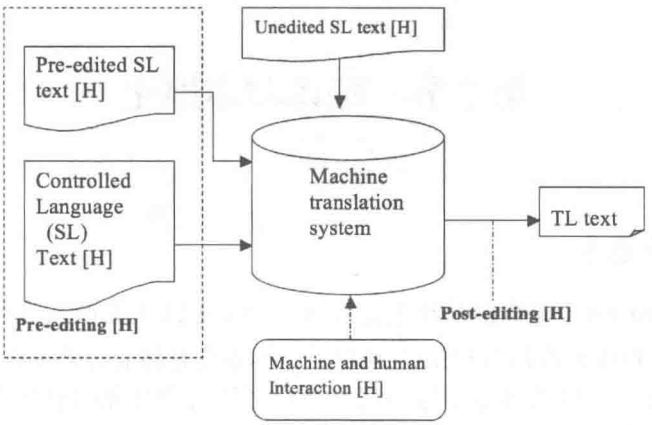
第二代机器翻译系统的设计理念是基于统计的（Statistical MT），通过收集大量真实的双语文本建成平行语料库，在翻译时由计算机通过语料库进行词汇匹配，靠统计结果进行歧义消解处理和译文选择。创建大规模的语料库在信息时代不是难题。

新一代的机器翻译系统基于深度学习技术和神经网络技术。深度学习（Deep Learning）是机器学习的一种，用多层神经元构成的神经网络，仿照人类大脑的思维方式以及神经网络的接收和反馈方式，达到让机器学习的功能。深度学习又名深度神经网络学习（Deep Neural Networks）。深度学习设计与机器翻译在结合初期以统计机器翻译系统为框架，利用神经网络来改进其中的关键模块。2016 年 9 月，谷歌发布新的神经机器翻译系统GNMT（Google Neural Machine Translation），直接用神经网络将源语言映射到目标语言，形成端到端（ end to end）的神经网络机器翻译。根据谷歌机器翻译研究团队的论文，较之基于短语的统计机器翻译（PBMT），能够降低 60% 的错误率。微软也已经采用深度神经网络统计翻译。

① 欧洲机器翻译委员会的定义，http://www.eamt.org/mt.html。

三、机译人助翻译（Human-aided machine translation）

机译人助翻译指“由计算机程序完成翻译过程，但在各个阶段均存在人工介入（Slocum,1988）”的翻译体系。翻译工作由机器翻译系统完成，但需要人工进行译前编辑（pre-editing），并且在译文生成后，由人工进行译后编辑（post-editing）。译前编辑的目的在于挑出源文本中可能在机器翻译过程中导致错误的词汇、短语及印刷错误等，并进行修正。译后编辑需人工以一定的标准评判译文的适用性和语言风格。在机器翻译程序运行过程中也有可能进入人工介入，如系统出现错误提示后由人工修订出错词汇或短语。

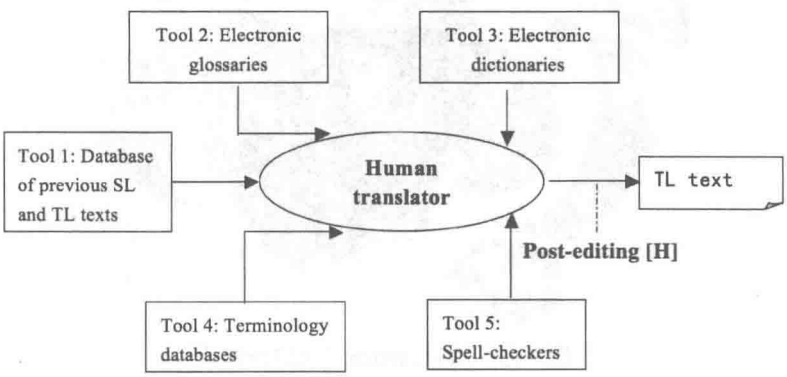


SL=source language; TL=target language； H=human

图 1-3 机译人助翻译模式（Quah, 2006）

四、人译机助翻译（Machine-aided human translation）

机译人助模式的核心是机器，而人译机助模式的核心则是人，即由译者使用多种计算机翻译工具完成翻译的过程。综合性的人译机助系统由于集合了多种翻译工具，有时也称为“工作站”。这些工具包括（图 1-4）平行语料库、词汇表、词典、术语库及拼写检查工作等。



SL=source language; TL=target language； H=human

图 1-4 人译机助翻译模式（Quah, 2006）

五、计算机辅助翻译（Computer-aided translation）

如本章图 1-1 所示，机译人助翻译与人译机助翻译均可定义为计算机辅助翻译（CAT）。二者的区别在于翻译工作的主导因素是人还是机器。也有学者据此将机译人助归于机器翻译的范畴，因为在机译人助模式下，机器为翻译的主导因素。广义地来看，在翻译过程中只要有计算机的介入即可称之为计算机辅助翻译，包括使用WORD软件、使用网络查询、使用在线词典等。狭义地来看，计算机辅助翻译等同于以人为主导因素的人译机助翻译，也是本书采用的定义。

第二节　翻译与本地化

一、本地化的概念

本地化（Localization）通常指外资企业在进入国（地区）生产或采购零部件和原材料的活动[①]。根据本地化行业标准协会LISA[②]的定义，本地化是将企业产品按特定国家/地区或语言市场的需要进行加工，使之满足特定市场上用户对语言和文化的特殊要求的生产活动。业界人士常将本地化写为“L10N”，“10”表示在首字母“L”和尾字母“N”之间的 10 个字母。

从LISA的定义中明确看到了翻译的身影。“按特定国家/地区或语言市场的需要进行加工”就是将外国产品上的文字信息转化为进入国的文字信息；然而本地化不能等同于翻译，因为本地化的产品必须“满足特定市场用户对语言和文化的特殊要求”。举例来看，图 1-5 是苹果公司英文网页上对apple watch的宣传图片，包含了一条文字信息“2PM Meet Erika at Gym”。

图 1-5　Apple watch英文宣传图片

① 由全国科学技术名词审定委员会审定公布。

② LISA全称为Localization Industry Standards Association，1990 年由九家本地化行业公司创办，LISA制定了包括TMX、TBX、SRX等在内的本地化行业交换标准，于 2011 年宣布关闭。

如果这条信息出现在英汉翻译练习中，将之译为：下午两点和埃里卡在健身房见面。当然没问题。但是，如果该产品进入了中国市场，需要将这段信息进行本地化，就不可以采取上述翻译。因为“埃里卡”不是中国人的姓名，不适用于中国市场。请看苹果公司中文网站上的相同图片。

图 1-6 Apple watch 中文宣传图片

图片上的信息为“下午 2:00 和玲玲一起健身”，“玲玲”就是典型的中文名字。因此，本地化的一个重要标准是要让产品看上去是在本地文化氛围中设计开发的。

二、本地化的主要内容

目前本地化翻译的主要内容包括软件本地化、网站本地化、技术文档本地化及媒体文档本地化等。

1. 软件本地化（Software Localization）

随着全球消费软件和商业软件市场的不断扩大，软件本地化的需求增长很快。软件本地化是指改编软件产品使之适合目标市场的语言、文化和技术要求。软件本地化既可以针对已经投放市场的软件产品，也可以针对即将发布的软件产品。对产品国际化需求较强的大公司来说，后一种情况往往更多。因此，软件本地化常与软件产品的开发同步进行，以保证将多语种版本同时投放市场。

软件本地化的对象包括软件产品本身及其所有相关的文本，文字信息在不同语言之间的转换仅仅是软件本地化过程中的一部分。

软件本地化的步骤可包括：

（1）分析材料，评估本地化所需工具与资源。

（2）分析目标市场的文化因素、技术因素和语言因素。

（3）准备术语，创建术语库或词表。

（4）进行翻译。

（5）根据需要改编用户界面，包括图表及对话框大小。

（6）图表、脚本及其他含有文字信息、符号信息的媒体本地化。

（7）编译本地化版本并测试。

（8）审校语言并保证功能完善。

试读结束：需要全本请在线购买： www.ertongbook.com

（9）交付产品。

2. 网站本地化（Website Localization）

随着电子商务的迅速发展，对网站进行本地化意味着可以与不同国家的潜在客户进行更方便、更有效的交流和沟通。大型的网站几乎每天都在变化——新的功能、内容、产品发布、市场活动等。而且，网站也需要以多种语言吸引网络用户。

一个网站由某一域名下的所有可访问的网页构成，比如www.njust.edu.cn。网站的内容包括一系列的文档、图像、视频、程序等，每一项内容由一个通用资源标识符（Uniform Resource Identifier, 简称URI）进行定位。网站本地化即改编所有网站内容，使之适应特定目标市场的语言、文化和技术要求。

同样，网站本地化不仅仅意味着翻译，还涉及用户界面布局调整、本地特性开发、联机文档的制作以及保证本地化版本能正常工作的软件质量保证活动等。网站本地化不仅需要高超的翻译技巧，而且需要精通HTML、脚本语言、图像本地化以及功能测试；还需要掌握多语种和方言的解决方案，为目标客户的理解搭建起一座信息沟通的桥梁。

3. 各类文档本地化

随着各种计算机技术的发展，文档的格式种类繁多。产品开发的技术文档、用户手册、帮助文件及教学电子文档种类繁多，也是本地化的重要内容。其中，媒体文档本地化（Multimedia Localization），如游戏、电影的旁白音频文件、字幕文件等，尤为突出。

三、翻译与本地化的关系

翻译的对象可以分为文学文本与非文学文本，翻译实践也有文学翻译与非文学翻译之分。不难看出，本地化的主要对象属于非文学文本。

本地化的内容不仅包括将信息用不同语言重新表达，还涉及与本地化对象相关的技术，比如软件本地化中的用户界面调整，网站本地化中的页面布局调整等。因此，文字翻译仅仅是本地化工作中的一个环节。

在本地化过程中，涉及语言文字的翻译往往并非字面翻译，而是更加注重使用目标语言传递语言文字的内涵意义。

由于翻译在本地化中的重要地位，本地化有时也称为本地化翻译。同时，技术在本地化翻译上不可或缺，也可以说，本地化的需求带动了计算机辅助翻译的发展。

全球范围内的本地化需求也带动了翻译市场的繁荣，呈现出行业化、规范化、服务化、流程化的特点。

四、本地化行业的发展现状

下面引用崔启亮博士的文章《中国本地化行业二十年（1993—2012）》来介绍本地化在中国的发展。

20世纪90年代初，本地化服务行业在我国萌芽。随着国际大型软件公司加快软件全球化的步伐，软件本地化服务需求不断提高，本地化服务行业也在不断探索中逐渐积累了技术和经验。1995—2002年，我国本地化服务行业进入快速发展的“黄金时期”，当今知名的中国本地化服务公司几乎都是在这一时期成立的。2009年，中国翻译协会本地化服务委员会正式成

立，标志着国内本地化服务行业结束了无序发展的状态，确立了中国本地化服务的行业地位。本地化服务委员会成立后，与中国翻译协会、本地化公司以及多所大学展开了一系列工作，使得我国本地化服务行业的面貌焕然一新。委员会通过多种方式，积极促进本地化在国内外的传播，促进了行业会议与专题沙龙的规范化与多样化、本地化和翻译专业人才培养的规范化与专业化，制定了行业规范促进本地化行业的规范化发展，并建立了语言服务业调研与报告机制。

根据中国翻译协会的定义，新时代语言服务业的范畴不仅包括传统的翻译，还包括本地化服务、语言技术工具开发及应用、语言教学与培训、语言服务咨询、语言服务管理、语言服务业发展战略、语言服务智库等。根据《2016 中国语言服务行业发展报告》，2011—2016 年，全球语言服务内容比重，所有本地化相关服务总和已经超过现场口译。

表 1-1 2011—2016 年全球语言服务内容比重 （单位：%）

语言服务内容	2011 年	2012 年	2013 年	2014 年	2015 年	2016 年
笔译	45.68	45.70	45.56	33.13	56.01	65.31
现场口译	14.44	14.05	11.38	9.64	9.20	7.61
软件本地化	6.55	6.17	6.53	6.38	6.57	3.12
电话口译	3.40	2.40	2.22	5.47	3.94	1.05
项目管理	—	—	—	4.60	1.10	0.89
网站全球化	4.72	5.44	5.02	4.31	3.86	1.67
桌面排版	—	—	—	4.14	3.69	3.06
创译	1.90	2.71	2.77	3.89	2.75	3.46
游戏本地化	—	—	—	3.82	1.03	0.54
国际化服务	2.29	2.91	2.59	3.80	0.49	0.36
笔译技术	3.99	3.40	3.21	3.46	—	—
机器翻译译后编辑	2.33	2.47	2.42	3.33	1.40	3.94
多媒体本地化	3.27	3.43	3.79	3.15	2.71	1.10
测试和质量保障	2.35	2.51	2.69	3.15	1.39	0.56
旁白/配音/叙述/字幕	4.35	4.20	3.93	3.14	1.64	1.75
手机应用本地化	—	—	—	2.39	0.77	0.51
视频口译	0.89	1.29	1.18	1.37	1.41	0.31
口译技术	1.59	1.36	1.08	0.83	—	—

行业需求增大意味着更多的服务需求和人才需求。《2012 中国语言服务行业报告》显示，截至 2011 年，全国在营语言服务及相关企业为 37 197 家 。这个数据在《2014 版报告》[①]中为：到 2013 年底，语言服务及相关企业增加到了 55 975 家；在《2016 版报告》中为：到 2016 年底，注册的含有语言服务业务的企业为 72 495 家。语言服务的需求从数量、内 容、语种、地域等方面正处于从过去改革开放几十年来的“外译中”向未来几十年的“中译外”

① 《2014 中国语言服务行业报告》

的大转折与大变革时代。这是中国语言服务主动走向世界的历史契机，是中国文化融入世界的现代潮流，是中华文明影响世界的未来趋势。

大致来说企业对本地化人才的需求可以分为以下四个方面：一是需具备行业基本知识，即熟悉本地化的概念、本地化工程的流程等；二是需具备语言翻译能力，包括两种及以上语言之间的相互转换、熟练使用计算机辅助翻译工具等；三是需具备一定的技术能力，包括对网络、软件等进行测试和发布；四是团队协作和沟通能力。除了两种及以上语言之间转换能力及团队协作和沟通能力，以上内容在本教材中均有涉及。

思考与练习

1. 请说明首字母缩略CAT的含义。
2. 请说明机译人助与人译机助的区别是什么。
3. 请举例说明翻译与本地化的关系。
4. 如果有过翻译经验，请根据个人的翻译经验归纳总结翻译流程。
5. 在翻译的过程中，你使用过哪些工具？

第二章　计算机辅助翻译的理论基础

第一节　语料库与翻译

一、语料库的概念

语料库即语言资料库。语言资料存储在电脑上之后，就可以对其按条件进行提取、分析、归类、对比等，从而帮助语言学家探索更多的语言奥秘。

建立语言资料的集合并以各种手段对此集合中的资料加以提取和分析，作为一种研究方法，可以追溯到19世纪，甚至更为久远。现在一般以乔姆斯基转换生成语法的兴衰时间为参照，可以说，语料库作为一种语言研究方法的衰落及复兴恰恰与乔姆斯基转换生成语法的兴衰相反，这也就是经验论与唯理论之间此兴彼衰的历史。语言学家对语料库有很高的评价，比如，Halliday（1994）认为："语料研究在语言的理论探索中具有中心位置，对语料的开发途径很多。" Leech（1993）认为："从科学方法的角度，语料研究方法是一种更为强有力的方法，因为其结果是可以验证的。" 顾曰国（1998）认为："语料库和语料库语言学可以说是两阵对垒的天平上的一个举足轻重的砝码。"

二、语料库的分类

语料库的种类不少，且与其功能相关。参照潘永樑（2001）的分类方式。

1.以语料的媒体形式分类：可分为书面语语料库、经过转写的（transcribed）口语语料库、视频语料库及上述几种形式的混合语料库。

2.以语料库设计结构分类：可分为均衡结构语料库、无结构的开放式的监控性语料库（monitor corpus）。"监控语料库"的主张由COBUILD项目主持人Sinclair教授（1991）提出，即建立一个专用语料库，不断采集最新语料，对语言的发展变化实施监控。

3.以语料的来源分类：有单语种语料库与多语种平行语料库之分，源文语料库与翻译语料库之分，母语语料库与外语学习者语料库之分。

4.以语料的时效分类：有共时语料库与历时语料库之分。

5.以语料的处理方式分类：有未经赋码的文本语料库与经过赋码的文本语料库。在这一类中，前者是自然语料的文本，即未经任何赋码或标注处理的原始语料文本；后者是根据研究需要对文本经过赋码处理的语料文本。

一个语料库可以兼备上述特征中的几种，以英语国家语料库为例。1991年，牛津大学出

版社、朗文出版公司、钱伯斯－哈勒普出版公司与牛津大学计算机中心、兰开斯特大学、英国图书馆通力合作，在英国工贸部和工程与自然科学研究委员会的资助下，历时四年时间建立了“英国国家语料库”（British National Corpus，简称BNC），所收语料总量为1亿英语单词，是单语种语料库。90%的语料为书面语语料，10%的语料为口语语料，因此，BNC既是书面语语料库也是口语语料库。书面语语料的来源包括各类报刊、书籍和文本，甚至有未公开发表的信件、文件、备忘录等。口语语料有三个来源，一是私人谈话录音，二是工作录音，如政府工作会议、商务会谈，三是广播影视节目录音。这些语料均为20世纪末期的资源，涉及领域广泛，所占比例不同（图2-1）。1994年建成之后，没有再添加新的语料。因此，BNC也是共时的均衡型语料库。

Table 1. Composition of the BNC World Edition

Text type	Texts	Kbytes	W-units	S-units	percent
Spoken demographic	153	4206058	4.30	610563	10.08
Spoken context-governed	757	6135671	6.28	428558	7.07
All Spoken	910	10341729	10.58	1039121	17.78
Written books and periodicals	2688	78580018	80.49	4403803	72.75
Written-to-be-spoken	35	1324480	1.35	120153	1.98
Written miscellaneous	421	7373707	7.55	490016	8.09
All Written	3144	87278205	89.39	5013972	82.82

图2-1　BNC各类语料比例表

BNC同时也是经过赋码（annotation）的语料库，采取最常用的词性标注（Part-of-Speech / POS tagging）。文章的结构属性，如标题、段落、表格等也有标注。标注的存在使依条件搜索语料成为可能。希望了解 pay作为动词与名词的搭配用法。对检索关键词pay执行搜索，并限定后一个词的词性为名词（图2-2），就会准确地得到pay+名词的例句在语料库中的存在情况。

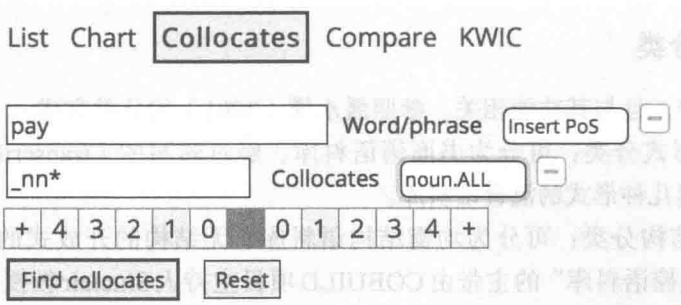


图2-2　在BNC语料库中对检索词后一个词进行词性选择

BNC首先以表格的形式展现Pay+名词的共现情况，如图2-3所示，可以通过点击任一词汇，查看具体的例句情况，比如点击共现率最高的名词attention，就能够得到图2-4，所有例句中的pay+attention高亮显示，清晰直观。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]　　COMPARE

		CONTEXT	FREQ	ALL	%	MI
1		ATTENTION	229	13190	1.74	5.29
2		TRIBUTE	186	1461	12.73	8.17
3		RISE	159	10263	1.55	5.13
4		TAX	136	16058	0.85	4.26
5		RISES	97	1773	5.47	6.95
6		INTEREST	91	26773	0.34	2.94
7		COMPENSATION	88	3066	2.87	6.02
8		CENT	82	37163	0.22	2.32
9		PEOPLE	70	119936	0.06	0.40
10		REDUNDANCY	66	1130	5.84	7.04
11		DIVIDENDS	57	974	5.85	7.05
12		TAXES	51	2759	1.85	5.38
13		INCREASE	49	16599	0.30	2.74
14		PACKET	47	1127	4.17	6.56
15		RATES	47	11286	0.42	3.23
16		INCOME	47	11862	0.40	3.16
17		INCREASES	45	4170	1.08	4.61
18		SECTOR	44	8617	0.51	3.53
19		REVIEW	44	9475	0.46	3.39
20		DAY	44	59298	0.07	0.74
21		MONEY	43	36031	0.12	1.43
22		MATERNITY	42	695	6.04	7.09

图 2-3　BNC中Pay+名词的总体情况

time (SP:KB2PSUNK) Two fifty a line and six pound a house (SP:KB2PSUGP) (unclear) (SP:KB2PSUNK) now **pay attention**, right, first number (SP:PS01V) Oh ye
It's really timid isn't it? (SP:PS0LK) but I don't think they **pay attention** to it across there (SP:PS0LM) Ah! (SP:PS0LK) it's not used to people
(SP:PS0L6) Okay? (SP:PS0L2) Right (unclear) (SP:PS0KY) (laugh) (SP:PS0L6) (shouting) Right if you didn't **pay attention** there, I don't see any others from next
just winding the teachers up and like generally taking the piss, I'd always **pay attention** and I'd always, always get good marks but somehow I think they
(unclear) Danny see what you can do. (SP:PSS3C) Erm (pause) (unclear) (SP:KPAPS000) Alright well **pay attention**. Matthew? (SP:KPAPSUNK) (unclear) (SP:KP/
It's running now. Thank you. (SP:PS1L8) Well we don't have to **pay attention** to it and er it's not necessary to record every word is it
earlier retirement (SP:J3WPSUNK) (laugh) (SP:PS3PD) alleged increased time and affluence, it beholds us to **pay attention** to this influential body of citizens,
fire risk we've probably got on the site. (SP:PS4UJ) Yeah. (sniff) (unclear) **pay attention** (pause) to get the Tats list (SP:PS4UK) Yeah y-- yeah yeah yeah (SP:PS4
one so that this ended up looking as clean as that. You have to **pay attention** to that part and that part too. Then you would (pause) take the
to you at the very beginning of September I'm the star, so you **pay attention** to me. That was the correct thing to do, very well though
, the parent's likely to notice and get cold themselves which means they'll **pay attention** to the baby, if the baby is far end of the house,
a day, turned round half way. (pause) Right can I ask you to **pay attention** to a couple of things when you do this diagram. Firstly, I
what was going wrong. (SP:PS46V) with you the mistakes. I want you to **pay attention**. Then you can get a chance to do it again in your book
look at how people recognise characters, it's not a-- the features that they **pay attention** to, but the overall shape. Quite often when you're looking at
in problem solving so I suppose it's there in a way. Do you **pay attention** to local features to build up global or d-- the other way round?
well but I can't remember what the question is, so you'd better **pay attention** for the whole of the rest of this lecture in case I (laugh) in
if your dog's called Rover, it's if you want get it to **pay attention**, it's the way you say Rover or B M W or (laugh)
the nineteen nineties. The thing has gone through a tidal change and we know **pay attention** to the views of women themselves. (SP:PS5GN) Both the Royal
sounds,' Gaver argues.' When we hear such sounds we do not **pay attention** to things like pitch and timbre. Instead we think about what made them
one of the things which makes Shared Ark work. You don't have to **pay attention**, but you are likely to notice a sudden hush, or any sounds

图 2-4　BNC中pay+attention搭配情况部分例句

三、平行语料库

语料库中包含两种或以上语言语料，则可称为多语种语料库。比如1988年启动的国际英语语料库项目（The International Corpus of English，简称ICE）。它的建设宗旨是为比较英语研究提供素材。全世界有15个研究小组参与其中，各自建立了提供本国英语素材的电子数据库，但都遵从统一的建库原则及语法附码方式。严格来说，ICE包含的语料是同一语言的不同变体。实际上，包括两种以上语言的语料库多为平行语料库（parallel corpus），即语料的构成为源文本及其不同语言的译文文本，语料在意义单位的不同层次，即字、词（组）、句、段

试读结束：需要全本请在线购买：　www.ertongbook.com