新编同义词词林

亢世勇 主编

XINBIAN TONGYICI CILIN

上海辞书出版社

新编同义词词林

安徽师范大学图书馆
藏书

孟琏勇 主编

XINBIAN TONGYICI CILIN

上海辞书出版社

图书在版编目(CIP)数据

新编同义词词林/亢世勇主编.—上海:上海辞书出版社,2015.12
ISBN 978-7-5326-4471-1

Ⅰ.①新… Ⅱ.①亢… Ⅲ.①汉语-同义词词典 Ⅳ.①H136.2-61

中国版本图书馆CIP数据核字(2015)第202706号

新编同义词词林

亢世勇 主编
责任编辑/徐 梅 贺晟威
装帧设计/姜 明
上海世纪出版股份有限公司
上海辞书出版社出版
(上海市陕西北路457号)
www.ewen.co www.cishu.com.cn
上海展强印刷有限公司印刷
开本850毫米×1168毫米 1/32 印张15.875 插页3 字数600 000
2015年12月第1版 2015年12月第1次印刷
ISBN 978-7-5326-4471-1/H·612
定价:48.00元

如发生印刷、装订质量问题,读者可向工厂调换
联系电话:021-66511611

顾问：冯志伟

主编：亢世勇

编者：王兴隆　姜　岚　徐艳华　刘海润　徐德宽　姜仁涛
李海鹰　李　艳　于屏方　冯海霞　石少伟　罗　琳
郭嘉伟　王　莉　赵　文　杨慧丽　张超男　阎　君
吕玲娜　刘　京　马永腾　武楠稀　韩淑红　张青文
王心玉　刘丽丽　王桂花　郑金勇　张业梅　张　焱

序

冯志伟

1982年7月，郭绍虞先生为《同义词词林》（上海辞书出版社，1983）作序，他从修辞和文法的角度，论述了学习词汇的重要性。他引用《文心雕龙》中的“句之清英，字不妄也”来说明，“古人学文在于记住字和词的用法，这才是一个真正难关”。他明确指出，“学中文的可以不必从文法入手，但是不能不从这些繁多的词汇入手”；他又指出，像《同义词词林》“这一类词书，看似不讲文法和修辞，但把汉语文法修辞两种学科，都包赅在内，经过这具体训练，比学语法修辞要好得多，因为就实用的意义讲，确实比空谈语法修辞之学者要实际”。我完全同意郭绍虞先生的这种看法。《同义词词林》是一个词汇的宝库，当我们写作时感到词穷而难以表达意思的时候，查一查《同义词词林》，我们就会豁然开朗，从中挑选到恰如其分的词语来表达我们的思想。《同义词词林》帮助我们排难解惑，常常使我们体会到“山穷水尽疑无路，柳暗花明又一村”的快乐。《同义词词林》出版三十多年以来，对于中文写作和外文翻译是非常有帮助的，它成了我们写作和翻译的好助手。

近年来，语言信息处理需要进行语义的形式分析，急需一套能够反映汉语单词语义特征的代码化的语义系统，而《同义词词林》中的每一个单词都有表示语义的代码，正好是一个代码化的语义系统，因此，语言信息处理学界的专家们把《同义词词林》当作一个宝贵的语言资源，并且把它改造成为计算机可读的电子文本，有力地推动了我国语言信息处理的研究。《同义词词林》在语言信息处理中的这种作用是郭绍虞先生在他的序中没有提到的，也是《同义词词林》的4位编者在编写时没有料到的。

然而《同义词词林》的初衷是为了写作和翻译而编写的，编者并

此为试读，需要完整PDF请访问：www.ertongbook.com

没有考虑到语言信息处理的特殊要求，因此，在语言信息处理中，《同义词词林》的语义代码往往会出现左支右绌、穷于应付的局面。

在这种情况下，我们深切地感到，需要从语言信息处理的需要出发，同时又要考虑到写作和翻译的需要，需要在《同义词词林》的基础上，重新编写一本同义词词林。鲁东大学多年来一直进行汉语语料库的研究，他们在词语的语义分类方面做了很多有价值的工作，成绩显著，因此，上海辞书出版社委托他们编写了这部《新编同义词词林》(以下简称《新词林》)。

在编写过程中，他们邀请我作为他们的顾问，我根据自己在机器翻译研究中设计的 ONTOL-MT 本体知识体系，参考《同义词词林》的语义代码，为《新词林》设计了一个新的代码系统。

我在设计这个新的代码系统时提出了如下 4 个原则：

第一，**普遍性原则**：对于任何两个意义相同的单词，不管这两个单词属于什么语言，它们在新的代码系统中的概念只有一个。

远在 1949 年，美国洛克菲勒基金会的副总裁韦弗(W.Weaver)在讨论机器翻译的时候就提出，当机器把语言 A 翻译为语言 B 的时候，可以从语言 A 出发，通过一种中间语言(Interlingua)，然后再转换为语言 B，这种中间语言是全人类共同的。我们的代码系统中的概念结点也应当是全人类共同的，它们应当适用于不同的语言，应当具有普遍性。

在普遍性原则的前提下，在编写不同语言的代码体系时，又应当考虑不同语言的特殊性。不过，特殊性是服从于普遍性的。新的代码系统表示的是语义，具有中间语言的性质，我们要首先考虑普遍性，其次才考虑特殊性。

目前这个代码系统只在《新词林》的编写工作中使用，只局限于汉语。但是，我们在设计代码体系时，是充分地考虑到它的普遍性的，它应当是多种语言共同的、通用的。

第二，**完备性原则**：新的代码系统中的概念代码应当具有完备性，它们应当尽量能够覆盖人类在自然语言中表达的所有通用的基本概念。

第三，**明晰性原则**：新的代码系统中的概念代码之间应当是泾渭分明的，它们应当具有明晰的界限，尽量避免交叉或重叠。在使用代码来标注词典的时候，应当尽量把不同的概念明晰地区分开来。

第四，**多角度原则**：事物从不同的角度观察，可以具有不同的特性，因此，同一个单词也可能具有不同的代码标记，这正说明了事物本身的多义性，应该是正常的。在新的代码系统中，同一个单词可以具有不同的属性，因而可以从不同的角度标注以不同的代码。

鲁东大学的师生们在这个新的代码系统的基础上，使用语料库技术对大量的词语进行了归类。经过数年时间艰苦的工作，今天这部《新词林》终于与读者见面了，这是值得庆幸的。

在这里，我愿意结合语言信息处理技术发展的需要，谈一谈《新词林》的作用，而且也像当年郭绍虞先生在他的《序》中所表示的那样，“放肆地谈一谈”。

目前，随着信息技术的进步和网络的发展，因特网(Internet)逐渐变成一个多语言的网络世界。在因特网上除了使用英语外，越来越多地使用汉语、西班牙语、德语、法语、日语、韩国语等英语之外的语言。据统计，从2000年到2005年，因特网上使用英语的人数仅仅增加了126.9%，而在此期间，因特网上使用俄语的人数增加了664.5%，使用葡萄牙语的人数增加了327.3%，使用中文的人数增加了309.6%，使用法语的人数增加了235.9%。因特网上使用英语之外的其他语言的人数增加得越来越多，英语在因特网上独霸天下的局面已经打破，因特网确实已经变成了多语言的网络世界。

语言是信息的最主要的负荷者，如何有效地使用现代化手段来突破人们之间的语言障碍，成为了全人类面临的共同问题。语言信息处理又叫作“自然语言处理”(Natural Language Processing，简称NLP)，这项技术包括机器翻译技术、跨语言信息检索技术、多语言问答式信息检索技术、多语言文本的自动分类技术、高效的搜索引擎技术，这些技术是解决语言障碍问题的有力手段之一。由于自然语言是人类历史长期发展的产物，带有浓厚的人文色彩，其结构极端复杂，其使用具有随机性和灵活性。在千百年数亿人的频繁使用中，由于历史长期积淀的差异和人们约定俗成方式的不同，自然语言在词汇层、句法层、语义层、语用层都充满了“歧义性”(ambiguity)，而且自然语言处理技术又往往涉及到多种语言，需要语言学、计算机科学、数学等多学科联合攻关，因此就更加复杂和困难。

可以毫不夸张地说，在进入21世纪之后，几乎每一个生活在信息网络

时代的现代人，都要直接或间接地与自然语言处理技术打交道。不论对于社会政治还是对于经济发展，自然语言处理技术都无疑是一个重要的研究领域。

在信息时代，科学技术的发展日新月异，新的信息、新的知识如雨后春笋般不断增加，出现了“信息爆炸”(information explosion)的局面。现在，世界上出版的科技刊物达 165 000 种，平均每天有大约 20 000 篇科技论文发表。专家估计，我们目前每天在因特网上传输的数据量之大，已经超过了整个 19 世纪的全部数据的总和；我们在新的 21 世纪所要处理的知识总量将要大大地超过我们在过去 2 500 年历史长河中所积累起来的全部知识总量。据 CNNIC 统计，2002 年底全球的网页总数已经达到 10^9 这样的天文数字。信息量的丰富大大地扩张了人们的视野，人们希望能够准确、迅速地搜索到自己需要的信息；而以自然语言为主要搜索对象的搜索引擎技术，将为解决海量信息的获取问题提供强有力的手段。

从 1954 年美国第一个俄语到英语的机器翻译实验获得初步成功开始，自然语言处理的研究已经有六十多年的历史了。在这六十多年的发展历程中，自然语言处理把语言学、计算机科学、数学、心理学、哲学、统计学、电子工程、生物学等学科融合起来，形成了一门独立的边缘性交叉学科。自然语言处理的范围涉及到众多的部门，如语音的自动识别与合成、机器翻译、自然语言理解、人机对话、信息检索、文本分类、自动文摘等等。在这六十多年中，自然语言处理逐渐形成了自己独特的理论和方法，在当代语言学和计算机科学中独树一帜。

目前，自然语言处理中的主流技术，是基于词法和句法分析的技术。尽管这些技术在某些受限的“子语言”(sub-language)中也曾经获得一定程度的成功，但是，这些技术难以有效地解决自然语言中普遍存在的“歧义性”问题，因而系统的质量不高，在实际应用中有很大的局限性。为了克服这样的局限性，自然语言处理需要在理论、方法和工具等方面实行重大的革新，其中一个重要的问题，就是在各种自然语言处理系统中，引入语义和概念的信息，以便进一步提高自然语言处理系统的智能。

例如，汉语中的“张三吃面包”“张三吃大碗”“张三吃食堂”三个句子，它们表面上的句法结构都是“主语—谓语—宾语”的形式：“张三”是主语，“吃”是谓语，“面包、大碗、食堂”类似宾语。在汉语中，它们在句法结构的形式上

是没有任何差别的。但是，这三个句子的英语译文却各不相同："张三吃面包"的英语译文是"Zhang San eats the bread"；"张三吃大碗"的英语译文是"Zhang San eats with a big bowl"；"张三吃食堂"的英语译文是"Zhang San eats in the restaurant"。在汉英机器翻译中，如果我们只使用词法和句法信息，按照"主语—谓语—宾语"的格式来翻译，这三个句子的英语译文都是相同的"Zhang San eats the bread""Zhang San eats the big bowl""Zhang San eats the restaurant"。显而易见，只有第一个句子的英语译文是正确的，而第二个和第三个句子的英语译文都是错误的，也是难以理解的。

但是，如果我们在机器翻译系统中引入关于单词的概念类别的语义信息。比如，"面包"的概念类别是"食物"，"大碗"的概念类别是"餐具"，"食堂"的概念类别是"建筑物"。根据这些概念类别来判定"面包"的语义功能是"受事者"，"大碗"的语义功能是"工具"，"食堂"的语义功能是"地点"，从而泾渭分明地把这三个在语义上不同的句子区别开来，在机器翻译时给它们分别赋以不同的语义功能，分别翻译为不同结构的英语句子："Zhang San eats the bread""Zhang San eats with a big bowl""Zhang San eats in the restaurant"。

可以看出，一旦在自然语言处理系统中引入概念语义信息，便可以进行自然语言的"歧义消解"(disambiguation)，使自然语言处理系统如虎添翼，把自然语言处理提高到一个新的水平。

近年来，国内外自然语言处理研究者已经逐渐认识到概念语义信息的重要性，开始在自然语言处理系统中引入一些概念语义信息。但是，这些概念语义信息大多数还是零零星星的、片段的、偶发性的，它们难以构成一个完整的系统。可以说，目前大多数自然语言处理系统还没有对于概念语义信息进行过全面的、科学的、系统的研究。

《新词林》的语义代码是建立在"知识本体"(ontology)的基础之上的，我们力图从知识本体的角度出发，对自然语言中的概念语义信息进行全面的、科学的、系统的研究，建立一个比较完善和全面的语义代码系统，从而为机器翻译、信息检索、搜索引擎提供强有力的概念语义信息支持，大大地提高这些系统的智能化水平，推动我国自然语言处理的发展。当然，建立在知识本体基础之上的《新词林》同时也是一部同义词词典，它同样能够丰富读者的词汇知识，帮助读者提高写作和翻译的水平。这样，《新词林》不但能够

为自然语言处理服务，也能为写作和翻译服务。

知识本体是语言概念知识系统的、科学的描述方法。知识本体这个学科来自古希腊的哲学。为了整理清楚“知识本体”这个科学的概念以及它和《新词林》的代码体系的关系，我们这里有必要对于知识本体研究的发展情况作简要的说明。

如果我们对于一个领域中的客体进行分析，找出这些客体之间的关系，获得了这个领域中不同客体的集合，这一个集合可以明确地、形式化地、可共享地描述这个领域中各个客体所代表的概念的体系，它实际上就是概念体系的规范，这样的概念体系规范就可以看成这个领域的“知识本体”。

人们很早就开始研究知识本体，因此，知识本体有很多不同的定义。这些定义有的是从哲学思辨出发的，有的是从知识的分类出发的，最近的一些定义则是从实用的计算机推理出发的。

牛津英语词典对于知识本体(ontology)的定义是“对于存在的研究或科学”(the science or study of being)。这个定义显然是非常广泛的，因为它试图研究存在的一切事物，为存在的一切事物建立科学。不过，这个定义确实是关于知识本体的经典定义，它来自哲学研究。

什么是事物(things)？什么是本质(essence)？当事物发生改变时，本质是否仍然存在于事物之中？概念(concept)是否存在于我们的心智(mind)之外？怎样对世界上的实体(entities)进行分类？这些都是知识本体要回答的问题，所以，知识本体是“对于存在(being)的研究或科学”。

远在古希腊时代，哲学家就试图研究当事物发生变化的时候，如何去发现事物的本质。例如，当植物的种子发育变成树的时候，种子不再是种子了，而树开始成为了树。那么，树还包含着种子的本质吗？巴门尼德(Parmenides)认为，事物的本质是独立于我们的感官的，种子在表面上虽然变成了树，但是，它的本质是没有改变的，所以，在实质上种子并没有转化为树，只不过是我们的感官原来感到它是种子，后来感到它是树。亚里士多德(Aristotle，公元前384—322)认为，种子只不过是还没有完全长成的树，在发育过程中，树的本质并没有改变，只是改变了它存在的形式，从没有完全长成的树(潜在的树)变成了完全长成的树(实在的树)。种子和树的本质都是一样的。知识本体就是要研究关于事物的本质的问题。亚里士多德还把“存在”区分为不同的模式，建立了一个范畴系统(system of categories)，包

含的范畴有十个：substance(实体)，quality(质量)，quantity(数量)，relation(关系)，place(空间)，time(时间)，attribute(属性)，state(状态)，action(行动)，passive action(承受)。这就是著名的十大范畴系统。这个范畴系统是最早的概念体系，实际上也就是最早的知识本体。亚里士多德以他卓越的学识和深刻的洞察力，抓住了人类认识中最关键的概念。我们在设计《新词林》的语义代码体系时，仔细地研究了亚里士多德的十大范畴系统，把它作为我们的重要参考。

在中世纪，学者们研究事物本身和事物的名称之间的关系，分为唯实论(realism)和唯名论(nominalism)两派。唯实论主张，事物的名称就是事物本身；而唯名论主张，事物的名称只不过是引用事物的词而已。在中世纪晚期，大多数学者都倾向于认为，事物的名称只是表示事物的符号(symbol)，例如，英语的 book 这个名称只不过是用来引用一切作为实体的“书”的一个符号。这是现代物理学的一个起点，在现代物理学中，采用不同符号来表示物理世界的各种特征(如速度的符号为 V，长度的符号为 L，能量的符号为 E，等等)。这些用符号表示的特征，实际上都是物理学中的概念或范畴。

德国哲学家康德(Immanuel Kant，1724—1804)认为，事物的本质不仅仅由事物本身决定，也受到人们对于事物的感知或理解的影响。康德提出这样的问题：“我们的心智究竟是采用什么样的结构来捕捉外在世界的呢?”为了回答这个问题，康德对范畴进行了分类，建立了康德的范畴框架，这个范畴框架包括 4 个大范畴：quantity(数量)，quality(质量)，relation(关系)，modality(模态)。每一个大范畴又分为 3 个小范畴。Quantity 又分为 unity(单量)，plurality(多量)，totality(总量)3 个范畴；quality 又分为 reality(实在质)，negation(否定质)，limitation(限度质)3 个范畴；relation 又分为 inherence(继承关系)，causation(因果关系)，community(交互关系)3 个范畴；modality 又分为 possibility(可能性)，existence(现实性)，necessity(必要性)。根据这个范畴框架，我们的心智就可以给事物进行分类，从而获得对于外部世界的认识。例如，本文作者冯志伟属于的范畴是：unity，reality 和 existence。这样，我们就认识到：冯志伟是一个“单一的、实在的、现实的”人。因此，康德的范畴框架是帮助我们捕捉外在世界的有力手段。在数据库中，我们可以根据康德的方法给事物建立一些范畴，从而根据这些范畴来管理数据。例如，我们给人事管理数据库建立“姓名，性别，籍贯，职业”等范

此为试读，需要完整PDF请访问：www.ertongbook.com

畴，使用这些范畴进行人事管理。可以看出，康德对于范畴框架的研究，为知识本体的研究奠定了坚实的基础。我们在设计《新词林》的语义代码体系时，也注意到了康德的这个范畴框架，不过我们更多地参考了亚里士多德的十大范畴系统。

在20世纪末和21世纪初，知识本体的研究开始成为计算机科学的一个重要领域。它主要的任务是研究世界上的各种事物（例如物理客体、事件等）以及代表这些事物的范畴（例如概念、特征等）的形式特性和分类。计算机科学对于知识本体的研究当然是建立在上述经典的知识本体研究的基础之上的，不过，有了很大的发展。因此，我们有必要重新给知识本体下定义。下面，我们介绍在计算机科学中对于知识本体的定义。

在人工智能研究中，格鲁伯（Gruber）在1993年给知识本体下的定义是："知识本体是概念体系的明确规范"（An ontology is an explicit specification of conceptualization）。

这个定义比较具体，也比较便于操作，在知识本体的研究中广为传布。

1997年，波尔斯特（Borst）对格鲁伯的定义作了很小的修改，提出了如下的定义："知识本体是可以共享的概念体系的形式规范"（Ontologies are defined as a formal specification of a shared conceptualization）。

1998年，施图德（Studer）等在格鲁伯和波尔斯特的定义的基础上，对于知识本体给出了一个更加明确的解释："知识本体是对概念体系的明确的、形式化的、可共享的规范"（An ontology is a formal explicit specification of a shared conceptualization）。

在这个定义中，所谓"概念体系"是指所描述的客观世界的现象中有关概念的抽象模型。所谓"明确"是指对于所使用的概念的类型以及概念用法的约束都明确地加以定义，所谓"形式化"是指这个知识本体应该是机器可读的，所谓"共享"是指知识本体中所描述的知识不是个人专有的而是集体共有的。

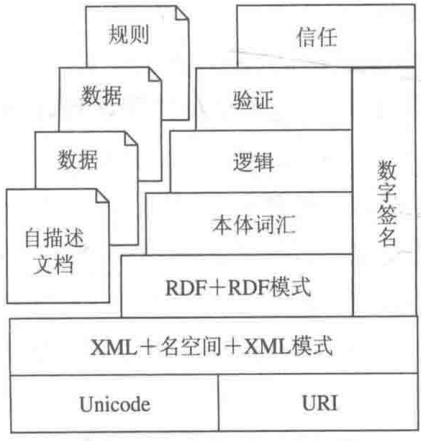
具体地说，如果我们把每一个知识领域抽象成一个概念体系，再采用一个词表来表示这个概念体系。在这个词表中，要明确地描述词的涵义、词与词之间的关系，并在该领域的专家之间达成共识，使得大家能够共享这个词表，那么，这个词表就构成了该领域的一个知识本体。知识本体已经成为了提取、理解和处理领域知识的工具，它可以被应用于任何具体的学科和专业

领域。知识本体经过严格的形式化之后，借助于计算机强大的处理能力，可以对人类的全部知识进行整理和组织，使之成为一个有序的知识网络。

知识本体的研究还与国际万维网的研究有着密切的关系。

2000 年，国际万维网联盟 W3C 总裁蒂姆·伯讷斯-李(Tim Berners-Lee)提出了下一代万维网——“语义网”(semantic web)的理念。现在，“语义网”已经成为计算机科学讨论与研究的热点。

2001 年，蒂姆·伯讷斯-李又进一步提出如下的语义网的体系结构：



语义网的体系结构

其中，Unicode 是国际统一的编码字符集；URI 是英语 Uniform Resource Identifier 的缩写，就是“统一资源定位符”，也被称为“网页地址”，是因特网上标准的资源的地址；XML 是英语 Extensible Markup Language 的缩写，就是“可扩展标记语言”；RDF 是英语 Resource Description Framework 的缩写，就是“资源描述框架”。在 RDF+RDF 模式的上面就是“本体词汇”(ontology vocabulary)，它处于语义网的关键层，用于表示语义网各种信息的概念和语义。由此可见，“本体词汇”在语义网的建设中起着承上启下的联系作用，处于举足轻重的重要地位。采用“本体词汇”来描述语义网中各种资源之间的联系，可以克服目前万维网上的信息格式的异构性、信息语义的多重性以及信息关系的匮乏和非统一性等严重问题。建

立在“知识本体”基础上的《新词林》就是这样的“本体词汇”，它对于语义网的建设具有关键性的作用。

2006 年 5 月，蒂姆·伯讷斯-李又宣布，经过十年的努力，W3C 已发布 W3C 推荐标准 80 余份，语义网已经具备了为达到成功的目标所需要的所有标准和技术，包括作为数据语言的 RDF、本体语言、查询和规则语言。2006 年 4 月，万维网联盟中国办事处成立并召开了 WWW 技术研讨会。

可以看出，我们在《新词林》中提出的基于知识本体的语义代码体系，是有深刻的科学根据的，是人类关于知识本体的研究在信息网络时代的新发展，它的理论意义和实用价值，已经远远地超出了郭绍虞先生在三十多年前所强调的“修辞和文法”的领域，它将会在自然语言处理和网络信息处理中发挥巨大的作用，对于国际万维网以及语义网的建设也是很有帮助的。

《新词林》的语义代码体系可以提供单词的概念类别特征，这些特征有助于提高机器翻译系统的质量以及歧义消解的能力，可以作为高质量的机器翻译词典编制的基础。

在机器翻译中，如果我们根据《新词林》中具有普遍性的语义代码来标注英语机器词典中单词的固有语义特征，由于汉语和英语的词汇都使用同样的语义代码，它们彼此之间的对应关系将变得非常清晰。这是一种新型的机器翻译系统，可以从根本上改善机器翻译系统的质量。

例如，汉语的“我/用钢笔/写/信”这个句子，使用《新词林》中的语义代码可以标注如下：

“Ab01(人·第一人称)/Cf07(器具·文具)/ Ka10(语言活动·书写)/Ck01(创作物·文书)”。

这个句子的英语译文为“I/write/a letter/with a pen”，使用《新词林》中的语义代码体系可以标注如下：

“Ab01(人·第一人称)/Ka10(语言活动·书写)/Ck01(创作物·文书)/Cf07(器具·文具)”。

不难看出，汉语句子和相应的英语句子在《新词林》中的语义代码完全是一样的，只是由于汉语和英语的语法结构的差异，这些标记排列的顺序不尽相同，而这种差异可以通过设计强大的句法语义自动分析软件来解决。显而易见，在《新词林》语义代码这个层面上，同一个句子在不同语言中的标记得到了高度的统一，达到了完美的和谐，这就为多语言机器翻译系统的开发提供了有力的语言知识资源的支持。

从数学的角度来看，把A语言翻译为B语言的过程，就是把A的“显拓扑”空间（符号空间），通过“概念”还原到A的“潜拓扑”空间（语义空间），由于“潜拓扑”空间有一个绝对坐标系，使得A和B在“语义空间”上有统一的描述。只需把语言A的语义通过等价的变换，转成语言B的语义，再与语言B的“显拓扑”空间（符号）产生对应，就完成了翻译。传统机器翻译只注重可表达的“显拓扑”（符号）部分，“潜拓扑”（语义）部分非常薄弱，因此在逻辑上难以解决各种类型的歧义问题，很难获得精确的机器翻译效果。《新词林》中的语义代码体系，为在潜拓扑空间上表达语义提供了巨大的可能性，这是我们进一步开发机器翻译系统重要的保证。

动态的人类知识库是一种智能数据库，人们梦寐以求的智能搜索引擎离不开智能数据库，智能数据库是智能搜索引擎的基础。在搜索中，关键词越多，限制条件越多，搜索的范围越准确。但是，如果数据库没有足够的概念储备来解读所有的关键词，就难以有效地进行这样的搜索。《新词林》中的语义代码体系是在知识本体的基础上构建的，它可以通过语义空间的逻辑转换，把任何语种下的相关内容从网络中搜索出来。因此，《新词林》语义代码体系的研究成果还可以在跨语言信息检索、文本自动分类、搜索引擎等系统中得到应用，提高系统的召回率（recall）和准确率（precision）。在搜索引擎中，可以根据《新词林》中语义代码的概念关联，给用户提供智能搜索提示，从而提高搜索引擎的效率。这样的研究方向有着非常广阔的市场前景和发展潜力。

总而言之，在多语言的信息网络时代，《新词林》中语义代码体系对于机器翻译、文本处理和搜索引擎等实用系统的开发，有着良好的发展趋势和广阔的市场前景。

当然，《新词林》也完全保留了传统的同义词词典的全部功能。当读者在写作和翻译中发生词穷的情况而难以恰当地表达意思的时候，《新词林》可以帮助读者从语义查询有关词汇，以便读者从中挑选恰当的词语，这对于写作和翻译仍然是很有帮助的。

参考文献

1. 冯志伟，计算语言学基础，商务印书馆，2001 年。
2. 冯志伟，从知识本体谈自然语言处理的人文性，《语言文字应用》，2005 年，第 4 期。
3. 梅家驹等，《同义词词林》(第二版)，上海辞书出版社，1996 年。
4. Asuncion Gomez-Perez, Ontological Engineering with examples from the areas of Knowledge Management, e-Commercial and Semantic Web, Springer, 2004.
5. T.R.Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199—220, 1993.
6. G.Miller, R.Beckwith, C.Fellbaum, D.Gross, K.Miller, Introduction to WordNet: A on-line lexical database, International Journal of lexicography 3(4), 235—244, 1990.
7. G.Miller, WordNet: a lexical database for English. Communication of the ACM, 389(11), 39—41, 1995.
8. R.Studer, V.R.Benjiamins, D.Fensel, Knowledge Engineering: Principle and Methods, 1998.
9. P.M.Roget, Thesaurus of English Words and Phrases, London, 1851.
10. L. V. Berrey, Roget's International Thesaurus, Third edition, New York, 1962.
11. R.Hallig & W. von Wartburg, Begriffssystem als Grundlage für die Lexikographie(Versuch eines Ordnungsschemas), Berlin, 1963.

前　言

侧重于对同义词进行归类而不进行精确释义的辞书，其鼻祖当属两千多年前的《尔雅》，但训诂式的目的决定了其与现代语言学中严格意义的义类型词典有着本质的区别。20 世纪 80 年代，中国的语言学家才开始大规模编纂同义词类词典，如《简明同义词典》（张志毅，1981）、《同义词词林》（梅家驹，1983）、《简明类语词典》（1984，王安节等）、《现代汉语同义词词典》（刘叔新，1987）、《类语大词典》（1988—2000，北京语言学院）。但它们多属于释义型而非义类型词典。无论是写作、翻译，还是当今语言信息处理中的语义处理问题，都迫切需要义类型词典的编纂及出版。

1981 年出版的《朗曼当代英语词典》标志着英语义类型词典的成熟，而几乎同时出版的中国的《同义词词林》，尽管是现代汉语中第一部义类型词典，却从一开始就以较为成熟而著称。它研究了世界上类似的词典，如英国的《英语词语宝库》与日本的《分类词汇表》等，根据汉语的特点，将七万个汉语语词进行归纳分类，初步提出一个汉语语义分类体系，导夫先路，具首创之功。其后，各种义类型词典才日渐增多。

语义学时代的语言信息处理领域，分词和语法分析已相对成熟，我们可以设计功能强大的分词软件和语法分析软件，但要想让计算机彻底读懂人类的自然语言，实现语义信息的处理已是当务之急。目前大多数自然语言处理系统还没有对于概念语义信息进行过全面的、科学的、系统的研究。但是，随着语言信息处理的发展，既有的义类型词典已不能适应社会需求，有的义类型词典语义分类不够科学细致、词汇面貌不能及时更新、人机两用的兼容度低的弊端日益显露，与语言信息处理的要求逐渐脱节，以致出现难以应付

此为试读，需要完整PDF请访问：www.ertongbook.com