

军事英语 课程设计

■ 王玉云 胡绍兵 主编



随书附光盘一张



国防工业出版社

National Defense Industry Press

军事英语课程设计

主编 王玉云 胡绍兵

副主编 王永斌 刘 曦 曾琰萍

国防工业出版社

· 北京 ·

内 容 简 介

本书首先分析了军事英语的词汇、语法和句型特征。在此基础上,结合任职教育的实战性、针对性特点,提出将军事英语教学分为军事共核英语、军种专业英语和兵种技术英语三个层次,参照这三个层次所对应的词频、搭配和分级教学语料库,分别对三个层次军事英语的教学目标、内容和教学大纲的设计进行了尝试。随书所附光盘提供了军事英语语料库总库词频表、陆海空和军事共同项目四个一级子库词频表、军事英语缩略词词表、陆海空三军和军事共同项目四个子库与 ANC 的对照词表、军事共核英语与初高中词汇比较词表、陆军独有词表、通信兵技术词表,以及军事英语三个层次的教学大纲和教学内容设计示例。

本书对于想在最短的时间里学到最常用也最有用的军事英语的广大官兵以及从事军事英语教学或研究的教员和学者,具有重要的参考价值。

图书在版编目(CIP)数据

军事英语课程设计 / 王玉云, 胡绍兵主编. —北京:
国防工业出版社, 2015. 4
ISBN 978 - 7 - 118 - 09982 - 9

I. ①军... II. ①王... ②胡... III. ①军事 - 英语 -
课程设计 - 高等学校 IV. ①H31 - 41

中国版本图书馆 CIP 数据核字(2015)第 074098 号

※

国防工业出版社出版发行
(北京市海淀区紫竹院南路 23 号 邮政编码 100048)
北京奥鑫印刷厂印刷
新华书店经售

*

开本 787 × 1092 1/16 印张 11 1/4 字数 252 千字
2015 年 4 月第 1 版第 1 次印刷 印数 1—2000 册 定价 38.00 元(含光盘)

(本书如有印装错误,我社负责调换)

国防书店:(010)88540777

发行邮购:(010)88540776

发行传真:(010)88540755

发行业务:(010)88540717

前 言

伴随我军院校的任职教育转型,将有越来越多的军官和士官进入院校培训。基于其培训的岗位指向性、课程选择的实用性和我军日趋频繁参与国际维和、救灾等军事行动的现实,任职教育学员有必要在最短的时间内掌握军事指向的、最核心的外语知识和最基本的交流能力,以应对可能的紧急作战之需。对此“核心”和“基本”的理解,不能出自感觉和经验,而必须源于科学的频数、概率和分布信息。

本书基于国家社科基金“基于语料库的军事英语培训核心内容研究”(11XYY011)的成果之一“军事英语语料库”编写而成。作为面向教学的应用研究,本书的主要观点包括以下几方面:第一,面对院校任职教育转型而带来的教学对象、程序和环节的变化,军事英语培训必须以应对作战岗位需求的英语语言能力为核心,其培训核心内容的研究,比培训模式和方法的探讨更重要、更不容忽视和延误。第二,用于界定任职教育军事作战英语培训核心内容的标准,必须有频数、概率、分布等基于大量军事领域真实语料的数据作为参照。照搬、引借地方高校或商业机构的培训模式,或仅凭有限的经验和语言直觉而制定出的教学内容,其针对性和实用价值都有极大的局限。第三,基于分析军事英语语料库而得出的军事通用英语词汇、军种独有词汇和兵种技术词汇,以及这三级词汇的类联接模式、典型搭配和多词词丛,应是我军任职教育军事英语培训的核心内容,也是我军在英语培训课程设计上与联合国、北约等国际组织接轨,成功执行与外军的各项联合行动及军事交流的重要保证。

基于以上观点,我们首先分军、兵种对 452 本美军陆、海、空三军最新作战手册和技术手册文本进行扫描、文本转换和分词、赋码的艰巨工作,建成近 2000 万词次的军事英语语料库。在建成的军事英语语料库基础上,通过运用专门的语言索引工具,如 Wordsmith 和 Antconc 等,统计分析了军事英语语体在词频、词汇密度、词长、覆盖面、句长、少用词、词汇搭配、语法搭配和句型等方面的语言特征,并在一级子库和二级子库分析了实际使用中高频词的搭配规律和多词词丛搭配强度。为了准确计算陆、海、空三军和军事共同项目的语料与普通英语语料相比,多用了哪些词,有哪些技术性词汇,我们以美国国家语料库中的书面语子库为参照,计算出军事共同项目、陆军、海军和空军等四个一级子库的正核心词。然后以各军种一级子库正核心词与军事共同项目子库前 2000 词比较,排除两者共有词,剩下的分别定为:①Level 1 军事通用词汇,共 2320 词型;②Level 2 军种独有词汇,如陆军共 405 词型;③Level 3 兵种技术词汇,如通信兵共 520 词型。各级培训的核心内容即这三级词汇的词型,这些词型的主要类联接模式、典型搭配、搭配强度最高的 3~4 词词丛,以及以课文形式出现的高频且分布面广的词型的索引行或索引段。最后,我们设计了以词汇为中心的军事英语培训三级教学大纲。以军事通用、军种和兵种为标准,将军事英语分为三级,每一级由几个模块组成。如第一级军事通用英语分为军事基础、武器装备、单

兵训练、作战指挥以及战场求生五个模块。第二级军种英语,按各军种下的兵种数量分为相应数量的模块。以上两级是所有学员都要求掌握的内容。第三级兵种英语,模块划分与第二级相同,只是各兵种只掌握、了解本兵种模块的内容。将各级核心词汇分布于各个模块。在各个模块的教学实施中,根据各模块词汇的频数和分布两个信息,抽取最常用的10个词型为节点词,到所在子库索引,找出最典型的类联接模式,设定跨距内左右两边的高频搭配词,以及搭配强度最高的3~4词词丛、索引行或索引段。

本书所提供的军事英语语料库,在一定程度上缓解了军队任职教育转型过程中,军事英语教学和培训长期以来极度缺乏真实数据的困难,可以为我军任职教育英语培训的顶层设计提供科学的、系统的、理性的数据参考;教学分级语料库可以为各级各类英语培训提供不同容量和难度的真实语料;军事英语语体分析可以帮助各类阅读军事英语的教员和学员了解这种语体的语言特征;而书中提出的以词汇为中心的分级教学大纲,可以为各兵种的各种英语短训课程设计提供一个参考的样例。

本书共分6章。各章的编者分别是:第1章、第6章,王玉云;第2章,胡绍兵;第3章,王永斌;第4章,刘曦;第5章,曾琍萍。王玉云同时确定了全书的基本结构和主要内容安排,对全书各章进行了审读,并对参考文献进行了汇编。胡绍兵提供全书所有语料库的分析、对比和制表。

由于编者的知识结构和能力有限,各章中难免存在疏漏和不足,甚至谬误,欢迎读者不吝指正。

编者

2014年10月于重庆歌乐山

目 录

第 1 章 基于语料库的军事英语语体特征	1
1.1 军事英语概述	1
1.2 任务与方法	3
1.2.1 基本统计	3
1.2.2 关键度方法	3
1.2.3 MF/MD 模型	4
1.2.4 搭配词计算	4
1.2.5 词丛提取	4
第 2 章 军事英语词汇特征	5
2.1 词汇基本统计特征	5
2.1.1 词频	5
2.1.2 词汇密度	6
2.1.3 词长	7
2.1.4 覆盖面	7
2.1.5 多用词	8
2.1.6 少用词	11
2.1.7 一次性词	15
2.1.8 语族	16
2.1.9 词类	17
2.2 词项典型特征分析	17
2.2.1 名词与名词化	17
2.2.2 介词	19
2.2.3 人称代词	21
2.2.4 情态词	21
2.3 词汇搭配分析	23
2.3.1 搭配的层次	23
2.3.2 搭配类型	24
2.3.3 搭配强度	28
2.3.4 典型军事英语核心词搭配分析	29

第3章 军事英语句型特征	57
3.1 军事英语基本句型特征分析	57
3.1.1 简单句	57
3.1.2 并列句	63
3.1.3 军事英语复合句	64
3.2 军事英语特殊句型	66
3.2.1 祈使句型	66
3.2.2 被动句型	68
3.2.3 存在句型	69
3.2.4 比较句型	69
第4章 军事英语课程目标设计	72
4.1 现状分析	72
4.1.1 教材出版现状	73
4.1.2 个案分析	73
4.1.3 问题与措施	76
4.2 军事任职教育英语课程目标设计	77
4.2.1 应该学什么	78
4.2.2 可能学什么	78
4.2.3 必须学什么	79
4.3 需求分析	79
4.3.1 学员需求分析	79
4.3.2 教员情况调查	80
4.3.3 管理情况调查	80
4.3.4 岗位需求	80
4.3.5 需求分析的标准	80
第5章 军事英语课程内容设计	82
5.1 军事英语课程内容的选择原则	82
5.1.1 以词汇为中心	82
5.1.2 抓住核心知识	83
5.1.3 控制相关知识	84
5.1.4 以频率为标准	84
5.2 以词汇为中心的军事英语课程结构	85
5.2.1 军事核心词汇	85
5.2.2 军种独有词汇	85
5.2.3 兵种技术词汇	86

第6章 军事英语分级教学大纲设计 88

6.1 军事英语分级教学大纲设计 88

6.1.1 Level 1 军事共核英语教学大纲 88

6.1.2 Level 2 军种英语教学大纲 89

6.1.3 Level 3 兵种英语教学大纲 91

6.2 军事英语分级教学语料库 93

6.2.1 Level 1 军事共核英语教学语料库 93

6.2.2 Level 2 军种英语教学语料库 93

6.2.3 Level 3 兵种英语教学语料库 93

6.3 军事英语分级课程设计 94

6.3.1 Level 1 军事共核英语课程设计示例 94

6.3.2 Level 2 军种英语课程设计示例 140

6.3.3 Level 3 兵种英语课程设计示例 163

参考文献 170

第1章 基于语料库的军事英语语体特征

1.1 军事英语概述

伴随我军院校的任职教育转型,将有越来越多的军官和士官进入院校培训。基于其培训的岗位指向性、课程选择的实用性和我军日趋频繁参与国际维和、救灾等军事行动的现实,任职教育学员有必要在最短的时间内掌握军事指向的、最核心的外语知识和最基本的交流能力,以应对可能的紧急作战之需。对此“核心”和“基本”的理解,不能出自感觉和经验,而必须源于科学的频数、概率和分布信息。语料库方法为梳理感性的、零碎的、未能形成理性的判断和认识提供了很好的手段。国内目前所建英语语料库中,英语学习者语料库7个,平行语料库14个,规模达百万词次的13个。但这些语料库多是监控语料库,不公开发布,很难获取其资源。且这类通用语料库(General Corpus)虽然能提供丰富而真实的语言材料,但其采样范围远远超出军事作战英语文本,文体、话题和语域种类繁多,无法与任职教育所要求的用于作战准备的军事英语培训直接接轨。

近年来,借助专门用途语料库(Domain Specific Corpus)辅助专门用途英语(ESP)课程设计呈现出快速发展的趋势,已经建成并投入使用的有石油、法律、医学、海事等英语语料库。也有一个在建的军事外语语料库,但其中的军事英语子库的入库语料主要来自国外权威网站的军事新闻与报道,对于任职教育军事英语培训内容的开发,缺乏岗位关键能力的针对性和作战任务的适切性。已经出版的军事英语教材,从编者前言看,其素材来源庞杂,选材原则也主要依据编者的主观认知和经验。文献检索没有发现相关的研究专著,已有的学术或学位论文反映出两种主要倾向:一是将军事英语培训等同于地方高职英语课程,主张完全搬用地方高职英语的课程标准和教材;二是探讨如何将一些新的教学方法和手段引入军事英语课堂教学。存在的问题十分明显:第一,没有以科学的视角,在清晰地分析军事作战英语语体特征(包括词汇特征和句型特征)的参照数据的基础上,探讨培训内容的定位,研究结果不能有效帮助军队学员在有限的培训时间内,最先学到最常用、最有用的语言知识;第二,缺乏针对任职教育实用性特征所要求的围绕作战需求的培训内容,无法让英语水平确实成为任职能力的一种维度。没有对任职教育转型后军事英语培训目标“是什么”的深刻认识和培训内容“教什么”的科学分析,而直接跨向教学研究的最后一个问题“怎么教”,不仅本末倒置,违背了教育研究的基本规律,也不可能指望高效率的培训效果的出现。

正是基于这样的背景,我们启动了国家社会科学基金项目“基于语料库的军事英语培训核心内容研究”,期望能够以美军陆、海、空三军作战手册和训练手册为入库语料,建成一个千万词次的大型军事英语语料库,也期望基于这个语料库分析军事英语的语法特征、词汇特征(包括词汇的类联接、搭配和多词序列)和句型特征。另外,还期望能够根据

这些特征,研制不同层次英语培训的大纲设计,尤其是核心培训内容的提炼和组织,将语料库纳入到任职教育英语培训的整个过程,贯穿于军事英语培训的目标、内容、过程、评价等各个环节。

经过两年时间的艰苦工作,我们以美军陆军的作战手册和海军、空军的技术手册为入库语料,建成了一个近 2000 万词次的军事英语语料库(English Corpus of Military Manuals, ECOMM)。ECOMM 分为陆、海、空、军事共同项目四个一级军种子库和陆军的步兵、装甲兵、炮兵、工兵、防化兵、卫生兵、通信兵、情报兵、特种兵等九个兵种二级子库,总词次达 19816772,共有 5275 个入库样本。这些子库的覆盖面不同,故样本数也不同。其分布的情况如表 1-1 所列。

表 1-1 ECOMM 的样本分布

军兵种分布		词型	词次	样本数
ECOMM		65431	19816772	5275
军种一级子库	兵种二级子库			
AR(陆军)		32316	5020448	1416
	arm(装甲兵)	7533	545902	124
	art(炮兵)	9230	536490	126
	eng(工兵)	9991	575274	151
	inf(步兵)	9497	560932	132
	int(情报兵)	11016	536528	153
	med(卫生兵)	11766	566366	161
	nbc(防化兵)	11628	573223	183
	sig(通信兵)	9780	553335	227
	spf(特种兵)	12230	572398	159
AF(空军)		29720	4916850	1556
NA(海军)		32143	4937164	1003
CI(军事共同项目)		29869	4942310	1300

为了观察军事英语语体和通用英语语体的差异,这里选择了美国国家语料库(American National Corpus, ANC)作为参照语料库。其样本分布如表 1-2 所列。

表 1-2 ANC 的样本分布

ANC 样本分布	词型	词次	样本数
ANC 总库	134380	14755303	8824
Spoken(口语)	19940	3285739	2400
Face-to-face(面对面交流)	6427	199917	93
Telephone(电话交流)	18748	3085822	2307
Written 1(书面语-1)	73412	5049934	4809
Fiction(小说)	9668	62042	1
Journal(期刊)	71274	6895981	4563
Letter(信函)	5447	91911	245
Written 2(书面语-2)	89406	6419690	1615

(续)

ANC 样本分布	词型	词次	样本数
Non - fiction(非小说类读物)	10058	629375	76
Technical(技术资料)	59203	5037699	1319
Travel - guide(旅游指南)	39550	1041527	179

ANC 的语料来自 20 世纪 90 年代以来的美国英语,共有 14755303 词次,8824 篇样本,分为书面语和口语两大类。口语部分包括两类:面对面谈谈和电话交流。书面语部分又根据体裁分为两个子集:小说、期刊、信函,以及非小说类读物、技术资料 and 旅游指南。在军事英语语体分析阶段,我们保留了口语部分,一是因为该部分的样本除了一般意义上的日常会话,也有部分访谈和辩论,并非完全即兴的口头表达;二是因为美军陆军作战手册中,很多武器装备类操作性手册中,也有很多面对面的指令性语言,其语言特征介于正式与非正式、口语与书面语之间。

1.2 任务与方法

本书的目的是通过自建大型军事英语语料库,对军事英语的词汇、语法和句型进行客观描述,并提供有详细数据参照和语境的高频词表和分级语料库,为我军各军、兵种根据各自不同的培训目标和要求,研制开发高效、适切的培训材料。

采用的主要方法是国际上已逐渐盛行,而外军和我军在语言课程设计过程中都尚未触及的语料库方法,分析军事英语语体特征,再基于这些特征,设计各军兵种语言培训课程教学大纲。

不同语体的产生源于人们可以自主选择不同的表达方式。一种语体的特征从某种意义上讲,取决于文本中对某些表达方式的多用程度、偏爱程度。这一特性决定了对不同语体的研究方法往往是基于量的对比。本书就是通过将自建军事语料库(ECOMM)与 普通英语语料库(ANC)对比,找出军事英语的语体特征,主要是词汇特征、语法特征和句型特征。采用的具体方法有基本统计、关键度分析、MF/MD 模式、计算搭配词和词丛搭配强度计算等。

1.2.1 基本统计

交叉使用 Wordsmith 3.0(Scott 1999)、Antconc 和北京外国语大学开发的类联接分析器,对 ECOMM 和 ANC 做出基本统计,包括词频、词汇密度、平均词长、覆盖面、词类频数等,观察军事英语语体和通用英语语体的基本差异。因为两个语料库的大小略有不同,所以在比较各种频数时,一般都做了归一化处理,把它们转换成 1000 词的频数或 10000 词的频数,即每 1000 个词或 10000 个词中出现的频数。

1.2.2 关键度方法

关键度方法(Keyness Approach)的使用是为了找出 ECOMM 和 ANC 相比的多用词,以了解 ECOMM 和通用英语语料库在多用词使用方面的差异。主要是以关键词为中心,以语料库索引证据为依据,参照其所在的类联接框架,概括词汇的搭配行为。

1.2.3 MF/MD 模型

MF/MD 是 Multi - Feature (多特征)、Multi - Dimensional (多维度)的缩略形式。Biber(2009)^①分析了大量研究成果后,确定共有 6 个维度、67 项可能具有语体区分意义的语言特征。我们为此使用 Treetagger 词类赋码器(POS tagger)对两个语料库都做了赋码。

1.2.4 搭配词计算

搭配词的计算方法就是从语料库中将关键词的所有搭配词提取出来,然后用统计手段(Z 值、MI 值、log - log 值等)测量各搭配词与关键词共现的显著程度,以确定词项之间在多大程度上存在着“相互期待和相互吸引”(卫乃兴)^[1],从而概括、描述共现词在多大程度上反映了词项的典型搭配情况。由于我们的研究主要是为军事英语培训内容提供支撑,在选择搭配计算工具时,采用 log - log 值,因为它在计算搭配时,预先排除了功能词,呈现出来的主要是实义词。

1.2.5 词丛提取

通过计算每个多词词丛的期望频数和期望频数与观察频数之比,可以判断该词丛在语料库中使用是否显著,从而确定该词丛是否进入教学内容。本书提取和计算的是 3 ~ 4 词词丛。

① 转引自桂诗春《基于语料库语言学的英语语言学语体分析》。北京:外语教学与研究出版社,2009:35.

第2章 军事英语词汇特征

2.1 词汇基本统计特征

2.1.1 词频

词频分析首先是统计出每一个词型的频数(词次),其次是这个词型在各个子库中的分布值(Dispersion,简称D值),然后折算成100万词的理论频数(简称U值)和标准频率指数(Standard Frequency Index,SFI)。整个 ECOMM 语料库的综合词频表,参见附录 I-1(附录见光盘)。

下面以陆军 AR 子库的一个片断为例,稍加说明。

从表 2-1 可见,analytical 等 8 个词在 AR 的 9 个子库中的频数都是 119 次,但是它们的 D 值却不一样。最大的为 concurrently,其 D 值为 0.000876876,它在 9 个子库中均有出现,分布面最广。且在各库的频数相对比较均匀,说明它在陆军各兵种手册中使用都很广泛,是陆军各兵种共享词。而另一个词形 causeway,虽然频数也是 119,也在 9 个子库中出现,但其 D 值只有 0.000693730,远小于 cohesive,可以推断其分布面不会很大。查表 2-1 可发现,cohesive 虽在 9 个子库都有出现,但 119 次中有 64 次出现在步兵 inf 子库,而在通信兵 sig 子库只出现了一次。D 值最小的是的 decking,为 0.000675692,只出现在 9 个子库中的一个,即工兵 eng 子库,意味着它应该是工兵的兵种专业技术词(Sub-technical Term)。

表 2-1 AR 词频统计片段一

Word	Total	Texts	arm	art	eng	inf	int	med	nbc	sig	spf
analytical	119	9	13	8	6	1	46	8	10	16	11
anchors	119	5	1	0	113	0	1	0	1	0	3
ASPS	119	7	1	22	10	0	71	1	1	0	13
causeway	119	3	0	0	114	0	4	0	1	0	0
cohesive	119	9	3	10	18	64	2	12	2	1	7
concurrently	119	9	18	13	18	16	13	5	6	14	16
decking	119	1	0	0	119	0	0	0	0	0	0
digits	119	9	4	11	1	1	16	5	48	30	3

U 值是折算的百万分比频数,用以比较不同语料库的频数。在表 2-2 中,U 值最大的是 decking,为 2.318080753;最小是 ASPA,为 2.162620064。SFI 是从 U 值派生出来的一个指数,90 表示在 10 个词中出现一次,80 表示在 100 个词中出现一次,70 表示在 1000 个词中出现一次,60 表示在 10000 个词中出现一次,50 表示在 100000 个词中出现一次,

40 表示在 100 万个词中出现一次。表 2-2 中这几个词的 SFI 约在 43.3 和 43.7 之间,意味着它们大致是在 100 万词中出现一次。

表 2-2 AR 词频统计片段二

Word	Total	Texts	D 值	U 值	SFI
analytical	119	9	0.000864835	2.189358844	43.40316950
anchors	119	5	0.000697901	2.314388091	43.64436186
ASPS	119	7	0.000817937	2.162620064	43.34980228
causeway	119	3	0.000693730	2.311453677	43.63885194
cohesive	119	9	0.000818597	2.239955560	43.50239402
concurrently	119	9	0.000876876	2.221743593	43.46693936
decking	119	1	0.000675692	2.318080753	43.65128561
digits	119	9	0.000837172	2.217887153	43.45939445

2.1.2 词汇密度

我们采用标准型/次比 (Standardized Type/Token Ratio, STTR) 的方法计算词汇密度 (Lexical Density)。即对文本的每 1000 个词的型/次比重新计算,最后计算各 1000 词的平均型/次比。实际上是 1000 词的词块的型/次比。型/次比和词型数目成正比:型/次比越高,使用的词型也越多。

从表 2-3 可见,ANC 的标准型/次比为 45.52,词型数为 134380,而 ECOMM 的标准型/次比为 40.64,而词型数仅 65431。这说明 ANC 的内容覆盖面较广,需要较多的词型,而 ECOMM 的语料为军事专业类样本,用词较为集中。由于在任何文本中词型和词次都不太可能按比例增加,因此,可以对经过对数转换后的型/次比数值进行比较。从表 2-3 可见,这两种办法所得的结果不尽相同,但其走势是一致的:ECOMM 的两种型/次比都略低于 ANC,说明军事英语词汇比通用词汇应用范围要窄一些,也意味着任职教育英语培训,在有限的时间里,必须以军事专业的核心词汇为主要目标。

表 2-3 ECOMM 与 ANC 的词长比较

	ECOMM	ANC
词次	19816772	14755363
词型	65431	134380
标准型/次比	40.64	45.52
平均词长	5.50	4.89
平均句长	13.85	8.40
一次词	23997	57642
一次词的百分比	36.68	42.89
1-字母词/%	5.11194523	4.971737239
2-字母词/%	15.81345361	17.10702631
3-字母词/%	18.1044108	19.24070685
4-字母词/%	11.88260088	16.4703855

(续)

	ECOMM	ANC
5 - 字母词/%	9.131326327	10.36560763
6 - 字母词/%	8.069594474	7.847504162
7 - 字母词/%	9.132675117	7.497431333
8 - 字母词/%	7.810519573	5.77534479
9 - 字母词/%	5.628829019	4.159776629
10 - 字母词/%	4.303367234	2.969344135
11 - 字母词/%	2.35837724	1.6946321
12 - 字母词/%	1.369621284	0.901158876
13 - 字母词/%	0.609726347	0.570754196
14 - 字母词/%	0.483928504	0.239607323
15 - 字母词/%	0.189624363	0.188982921
15 ⁺ - 字母词/%	5.11194523	4.971737239

ECOMM 的词次大于作为参照语料库的 ANC,这里把以上绝对数值做了归一化处理,得出的结果如图 2-1 所示。

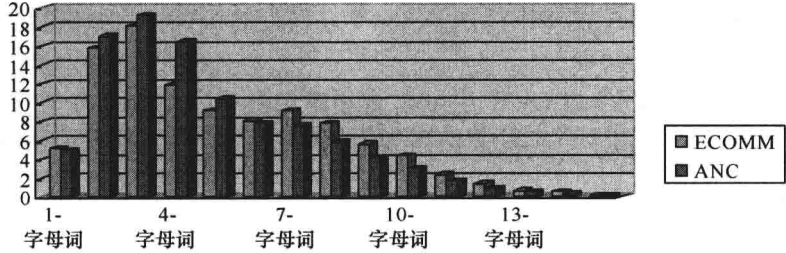


图 2-1 ECOMM 与 ANC 词长比较直方图

2.1.3 词长

从平均词长来看, ECOMM 与 ANC 分别为 5.50 与 4.89, ECOMM 明显长于 ANC。从图 2-1 可以看出,这是因为 ANC 的 3-字母的短词特别多。为了比较,把词次转换成百分比。图 2-1 展示了两个语料库的词长(按字母计算)的情况。不难看出,ANC 和 ECOMM 的走势不一样,2-字母到 5-字母的短词都多于 ECOMM,而 ECOMM 的 7~13 个字母的词要多于 ANC。这正是军事英语的另一特点:与普通英语相比,它的短词较少,而长词较多。

2.1.4 覆盖面

两个语料库的词汇覆盖面有较大差异。ECOMM 因为专业性强,覆盖面比 ANC 更集中。从前 1000 个词型到前 5000 个词型都是如此。ECOMM 的前 3000 个词型在语料库总词次的覆盖率就已经超过 ANC 的前 5000 个词型(92.33:90.72),而前 5000 个词型更是

覆盖了总词次的 95.72% (见表 2-4)。

表 2-4 ECOMM 与 ANC 词型词次分布

前 × × × 词型	ECOMM_Tokens	ECOMM /%	ANC_Tokens	ANC/%
1000	15782524	79.64	11027510	74.74
2000	17541817	88.52	12196469	82.66
3000	18296309	92.33	12782547	86.63
4000	18710347	94.42	13140496	89.06
5000	18968087	95.72	13386387	90.72

2.1.5 多用词

我们用 Antconc 的 Keyword 功能对 ECOMM 和 ANC 进行关键度分析,生成核心词表,其中包括正核心词汇和负核心词汇。正核心词是 ECOMM 与 ANC 相比使用频率偏高的词汇,而负核心词则是 ECOMM 使用频率偏低、而 ANC 中使用频率较高的词汇部分。得到 ECOMM 的正核心词 3331 个,负核心词 1905 个。词表按关键度高低排列,排在最前的词意味着多用情况最大。从表 2-5 中可见,关键度最高的前五个词分别是 be、operation、use、support、unit,它们在 ECOMM 里的关键度分别为 329256.3、69413.51、57982.18、56631.69、56181.87,而在 ANC 中的关键度则只有 232717.37、13163.88、348.58、48364.55、30326.05。

值得注意的是,3331 个多用词只占 ECOMM 总词型的 5.09%,但却覆盖了 78.47% 的词次。

表 2-5 ECOMM 与 ANC 的多用词比较(前 100 个)

No.	WORD	FREQ.	ECOMM/%	FREQ.	ANC/%	KEYNESS	P
1	operation	96341	0.45	2098	0.01	63450.14	3.14×10^{-24}
2	support	89438	0.42	5741	0.04	49833.85	6.48×10^{-24}
3	unit	78352	0.37	2876	0.02	48666.23	6.96×10^{-24}
4	force	78396	0.37	3764	0.02	46574.71	7.94×10^{-24}
5	or	216506	1.01	66348	0.43	38856.40	1.37×10^{-23}
6	commander	53817	0.25	248		37813.23	1.48×10^{-23}
7	mission	52465	0.25	1032		34792.34	1.91×10^{-23}
8	area	67600	0.32	6367	0.04	33309.77	2.17×10^{-23}
9	the	1463446	6.84	822750	5.39	32259.71	2.39×10^{-23}
10	must	61239	0.29	5232	0.03	31277.06	2.62×10^{-23}
11	frie	47105	0.22	1855	0.01	28926.62	3.32×10^{-23}
12	plan	58362	0.27	5779	0.04	28178.74	3.59×10^{-23}
13	train	48887	0.23	2873	0.02	27803.81	3.74×10^{-23}
14	enemy	39511	0.18	542		26799.28	4.17×10^{-23}
15	aircraft	38465	0.18	576		25963.28	4.59×10^{-23}
16	equipment	37184	0.17	950		24093.40	5.74×10^{-23}

(续)

No.	WORD	FREQ.	ECOMM/%	FREQ.	ANC/%	KEYNESS	P
17	commanded	34118	0.16	716		22498.71	7.06×10^{-23}
18	require	50134	0.23	5893	0.04	22387.16	7.16×10^{-23}
19	may	77924	0.36	16927	0.11	22185.84	7.36×10^{-23}
20	provide	56957	0.27	8501	0.06	22178.53	7.37×10^{-23}
21	personel	32474	0.15	530		21802.33	7.76×10^{-23}
22	air	41679	0.19	3327	0.02	21768.17	7.79×10^{-23}
23	will	85835	0.40	22235	0.15	19823.58	1.03×10^{-22}
24	position	39025	0.18	3539	0.02	19492.43	1.09×10^{-22}
25	combat	25442	0.12	304		17366.87	1.54×10^{-22}
26	use	110447	0.52	36379	0.24	17284.81	1.56×10^{-22}
27	requirement	31233	0.15	2074	0.01	17218.29	1.58×10^{-22}
28	procedure	31511	0.15	2331	0.02	16848.46	1.68×10^{-22}
29	system	50121	0.23	9538	0.06	16201.44	1.89×10^{-22}
30	control	49119	0.23	9570	0.06	15544.22	2.14×10^{-22}
31	should	50865	0.24	11258	0.07	14182.49	2.82×10^{-22}
32	target	30099	0.14	3151	0.02	14180.16	2.83×10^{-22}
33	flight	23835	0.11	1195		14018.69	2.93×10^{-22}
34	tactical	19721	0.09	73		13890.39	3.01×10^{-22}
35	maintenance	20927	0.10	440		13792.25	3.07×10^{-22}
36	battalion	19552	0.09	183		13476.83	3.29×10^{-22}
37	corps	19661	0.09	255		13365.72	3.38×10^{-22}
38	load	21362	0.10	840		13110.70	3.58×10^{-22}
39	communication	22736	0.11	1275		13059.67	3.62×10^{-22}
40	movement	21788	0.10	1096		12805.40	3.84×10^{-22}
41	element	25146	0.12	2211	0.01	12697.87	3.94×10^{-22}
42	conduct	24112	0.11	1886	0.01	12668.79	3.97×10^{-22}
43	division	20379	0.10	767		12589.94	4.04×10^{-22}
44	grind	22292	0.10	1363		12547.93	4.08×10^{-22}
45	supply	21716	0.10	1277		12338.36	4.29×10^{-22}
46	team	24386	0.11	2188	0.01	12223.01	4.42×10^{-22}
47	attack	24575	0.11	2262	0.01	12200.43	4.44×10^{-22}
48	capability	18670	0.09	437		12190.97	4.45×10^{-22}
49	ground	23170	0.11	1849	0.01	12095.16	4.56×10^{-22}
50	army	19366	0.09	963		11407.00	5.44×10^{-22}
51	vehicle	18324	0.09	678		11347.82	5.52×10^{-22}
52	weapon	20078	0.09	1207		11347.40	5.52×10^{-22}