

# 俄语军事缩略语知识库的构建

徐进

解放军外国语学院 河南洛阳 471003

Email: xujin2410@yahoo.com.cn

**摘要:** 缩略语是几乎所有的语言中都存在的语言现象, 它把必要的信息压缩到在接触的一瞬间就能立刻了解的程度, 使语言的使用和表达更加的便捷和快速。由于缩略语本身具有的简洁、快速等特点, 且具有一定的保密性, 所以在军事领域中的应用更为广泛。缩略语的广泛使用, 也成为自然语言处理中无法回避的问题, 日益受到自然语言处理的研究的重视。本文试图从理论上研究建立一个收词量大、内容全面、便于查询的俄语军事缩略语知识库, 以方便计算机对缩略语的自动处理。

**关键词:** 军事缩略语, 知识库, Access 数据库

## The Construction of a Repository for Russian Military Abbreviations

Xu Jin

PLA UFL, Luoyang, Henan, 471003

Email: xujin2410@yahoo.com.cn

**Abstract:** Abbreviations exist in almost every language. They compress necessary information to the extent that, once we see it, we understand it, enabling the rapid use of a language and its expression. Featuring concision and, to some extent, secrecy, abbreviations are widely used in the military field. The wide use of abbreviations is also drawing more and more attention from researchers in the field of natural language processing. The present study tried to carry out a theoretical study on the construction of a repository for Russian military abbreviations that is characterized by large vocabulary, extensive content, and convenient query, in order to facilitate the automatic processing of abbreviations by computer.

**Keywords:** military abbreviations, repository, Access database

### 1 引言

缩略语是几乎所有的语言中都存在的语言现象, 它把必要的信息压缩到在接触的一瞬间就能立刻了解的程度, 使语言的使用和表达更加的便捷和快速。随着社会的发展变革、各国对外交往的日益频繁及受到全球经济一体化的影响, 使得各个语言中都出现了大量的缩略语。缩略语的广泛使用, 也成为自然语言处理中无法回避的问题, 日益受到自然语言处理的研究的重视。

由于缩略语能在不影响内容的前提下, 大幅度缩减素材的字数, 减少文件的篇幅, 提高信息传输的速率, 且具有一定的保密性, 所以在军事领域中的应用更为广泛。现代战争要求快速准确的对大量的信息进行处理。但是在信息化战争的条件下, 军事斗争中的大量信息已经不是现有的人力所能够胜任的。在此背景下, 为了满足现有需求, 就必须使用更多的机器辅助手段, 甚至是由机器单独处理大量的军事信息, 以满足信息化条件下的军事斗争需要。

但是由于军事缩略语广泛存在于各种军事素材中, 且涉及多军种、多领域的专业术语; 由军事科技推动的新军事变革日益深入, 相应的技术装备和作战理论也层出不穷, 新的术语和由此产生的缩略语大量出现, 成为军事交流和翻译中的难题之一。

要解决上述问题, 首先就必须加强在军事缩略语方面的理论研究并在此基础上构建一个面向计算机的, 且能够供军事翻译人员快速查询, 不断更新的俄语军事缩略语知识库。但是就目前来看, 俄语军事缩略语的理论研究还比较缺乏, 知识库构建更是鲜有人涉及。

## 2 国内外缩略语词典、词库、知识库研究现状

俄罗斯方面对于缩略语库的建设还比较少,但是在线的有一个专门的缩略语库(网址为www.sokr.ru)。这个缩略语库收词全面,查询方便,对每个缩略语都提供俄文全称及解释,且经常更新,是笔者见到的目前最大的俄语缩略语库。但缺点在于无法下载,而且没有汉语释义。

国内对俄语军事缩略语的词典建设相对较为重视,出版了数本军事缩略语词典,主要有:《俄英汉军语缩略语词典》(温金权等,1987)、《俄汉军事缩略语大词典》(董文周、孙凯等,2001)及《俄汉军事缩略语大词典》(张汉宽、张洋等,2008)等。

这些词典的优点在于收词量大,释义准确,但是内容相对较为简单,仅仅只有俄语的全称及汉语释义等内容,且为传统的纸质词典,查询麻烦,扩充更新慢。

而对于中文缩略语知识库来说,已经有研究者开始了缩略语库的建设工作,并取得了一些成就。

北京大学语言所在87年建立《现代汉语语法信息词典》(以下称《语法信息词典》)时就已经开始重视缩略语的现象并建立了《简称略语库》。这个库文件包括了608条简称略语以及这些词语的拼音,所属子类,是否可以做主、谓、宾、定、状、补等成分,还说明了这个简称所对应的全称和一些使用的例子。支流等人在《简称略语库》的基础上对其进行了扩充,建立了一个中文缩略语知识库,协助机器自动还原中文缩略语。他们研究建立了面向信息处理的中文缩略语分类体系,完成了8000个缩略语的归类,建立了大规模的缩略语知识库,其中的每个记录主要包括缩略语、其对应的完整语汇或全称以及缩略语的同形、种类等属性。

鲁东大学的鲍明凌和亢世勇在《基于数据库的现代汉语新词语缩略语的研究》一文中介绍了他们的工作,他们建立了一个关系数据库用来存放缩略语的一些属性。有原词语,类型,缩略方式、构成方式、结构。他们所建立的库中共收录了2957条缩略语。这个缩略语库的规模和质量堪称上乘,但是由于建库的初衷并非是为了计算机自动处理,所以数据库的组织方式以及属性设置不适合计算机自动处理。

虽然以上词典或词库并不都是面向计算机自动处理的,但是却可以给我们提供优质的资源和可借鉴的经验。但总体来说,军事缩略语词库在国内还非常缺乏。而面向计算机自动处理的外语军事缩略语库就更是少之又少。

## 3 本文中“军事缩略语”所指的对象

本文中所指的“军事缩略语”,是广义上的。它面向军事理论、信息技术、作战指挥、武器装备、体制编制、教育训练与后勤保障等学科领域,不仅仅反映军事技术领域的用语,还包括展示信息化战争、联合军事演习、国际维和与反恐等方面的语词,甚至还包括适应驻外人员、外事工作人员的缩略语词汇。

## 4 俄语军事缩略语的特点

俄语军事缩略语的特点与一般的俄语缩略语的特点基本相同,但是也有本身的一些特色。

### 4.1 俄语军事缩略语的构成方式

俄语军事缩略语的组成结构按照其组成方式来说主要有以下几种:

#### (1) 首字母型

这种类型的缩略语主要是由每个单词的首字母组成的,这种缩略语在整个缩略语中占大多数,如:OPP(отдельная разведывательная рота, 独立侦察连)、OC(основные силы, 基本兵力)、OPT(область рассеяния тракторий, 弹道散射范围)等。

## (2) 抽取型

这种类型的缩略语是将一个单词或一个词组中的部分音节抽取出来组合在一起构成的,其组成相对较为复杂,大致可分为以下两种情况:

1) 抽取单个单词或词组中每个单词的部分音节或字母,其中的字母主要为辅音字母。如: ПГ (пламегаситель, 灭火罩, 消焰器〈炮兵〉)、макс (максимум, 最大值)、тех (техника, технический, 技术, 技术的)、комбат (командир батальона, 营长)等。

2) 抽取部分音节及整个单词或单词的变格形式组合在一起构成。如: комвзвод (командир взвода, 排长)、авиаразведка (авиационная разведка, 航空侦察)、авиасвязь (авиационная связь, 航空通信)等。

## (3) 符号数字型

这种类型的缩略语是由词或词组中的字母或音节加上特定的符号或数字组成的,使用什么符号大多没有什么规律,有时仅仅是为了书写方便。如: ЗПУ-1 (одиночная зенитная пулеметная установка, 单管高射机枪)、У.Хок (усовершенствованный Хок, 改进型“霍克”导弹)、Об. п. (объездной путь, 迂回线路)等。

## (4) 混合型

这种类型的缩略语是由上述两种构成方法一起作用组合而成的。如截成型和首字母型混合: СаК (санитарный катер, 医务艇)、ТуркВО (Туркестанский военный округ, 土耳其斯坦军区)、ПД ТСРП (противодействие техническим средствам разведки противника, 对敌技术侦察对抗行动)等。

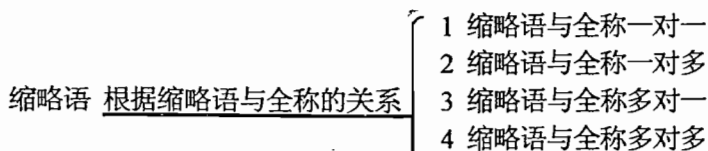
## (5) 音译型

这种类型的缩略语大都是直接由英语音译而来,如: ФТАХ (英文FATX的音译, 巴勒斯坦解放组织)、НАТО (英文НАТО的音译, 北约, 北大西洋公约组织)、НОРАД(英文缩写NORFD的音译, (美国、加拿大)北美联合空天防御指挥部)等。

## 4.2 存在大量歧义

引起俄语军事缩略语歧义的主要原因有以下两点:

一是由于缩略语与全称之间不是严格的一对一关系,而是存在着大量的一对多、多对一甚至是多对多的现象。如缩略语“АА”所对应的全称多达十个(авиакрыло 空军联队, авианосец 航空母舰, авианосная авиация 舰载航空兵, артиллерийская академия 炮兵学院等)。再比如“авиационная разведка (航空侦察)”可以缩略为“А”,同时也可以缩略为“авиаразведка”。这样的例子在俄语军事缩略语中比比皆是。按照缩略语与全称之间的对应关系如下图:



正是这种缩略语与全称之间对应关系的不唯一,使得俄语军事缩略语出现了大量的歧义,使机器难以处理。

二是由于缩略语本身与俄语中已经存在的单词相同引起的歧义。如“драп”一词,既可以是对“дальний разведывательный авиаполк (远程侦察航空兵团)”的缩略,也可以指“драп (厚呢子)”这一俄语中本身就存在的词。

## 4.3 区分大小写

俄语军事缩略语大部分都是大写字母组成，但是也有全部或部分由小写字母组成的。如：“дска (дивизион сторожевых катеров, 护卫艇大队)、Руз (Разведывательное управление (министерства обороны) Узбекистана), 乌兹别克斯坦国防部情报局”。俄语军事缩略语的特点在于同样的字母用大写或小写字母写出时，所表示的意义可能不一样。如“ОРРО”表示“отдельный радиорелейный отряд (独立无线电中继通信队)”而“орро”表示“отдельная рота ранцевых огнемтов (独立背囊式喷火连)”。

但是这种情况并不是绝对的，有时大小写所表示的意义是一样的，如：“оатр”和“ОАТР”都表示“отдельная автотранспортная станция (独立汽车运输连)”，“ПР”和“пр”都表示“пехотная рота (步兵连)”。

#### 4.4 军事缩略语的词形变化复杂

众所周知，俄语具有丰富的词形变化，俄语军事缩略语的词形变化也同样复杂。俄语军事缩略语基本上可以分为变化和不变化两类：

(1) 变化的俄语军事缩略语主要是截成型和混合型中以字母音节结尾的缩略语。其变化规则如下：

如果结尾是有音节组成的，则按照俄语语法的一般规则来变化。如：военвоздух (военное воздухоплавание, 军事浮空飞行)即按照阳性单数名词处理。

如果缩略语音节的最后一部分是由某个单词构成的，则按照该单词的变化规则来变化。如：Наштафлот (начальник штаба флота, 舰队参谋长)则按照单词флот来进行变化。

除了以上两种，还有极个别的首字母型缩略语由于经常使用，开始具有了普通词汇的词形变化。如：ОМОН(отряд милиции особого назначения, 特种警察部队)，这个词就按照普通的阳性单数名词处理。但是这样的缩略语数量极少。

(2) 不变化的俄语军事缩略语主要是首字母型和音译型的缩略语。如：ЗП(заатмосферный перехват, 大气层外截击)、ИБМ(ИВМ, International Ballistic Missile, 洲际弹道导弹)。

## 5 俄语军事缩略语知识库的构建

### 5.1 参照理论及模型

俄语军事缩略语库的构建主要参照了俄国学者А.Е.Кибрик关于词汇信息结构的学说及北京大学计算语言学研究所支流、段慧明、朱学锋、俞士汶等人设计的中文缩略语知识库。

“俄国学者А.Е.Кибрик认为俄语词汇知识库应设有七个部分的信息，这七个部分分别为：条头词；词法区；句法区；词汇区；义类区；语境区；百科区”。(傅兴尚，许汉成，易绵竹，李向东，2009：44)

北京大学计算语言学研究所支流、段慧明、朱学锋、俞士汶等人设计的中文缩略语知识库在上文中已有所介绍，在这里就不在赘述。

### 5.2 知识库构建流程简介

第一步，收集词条。词条收集的主要来源是各种俄语军事词典。目前数据库中的词条来自于军事网站，俄罗斯国家语料库等，但大多来自较为权威的军事词典，如：《俄汉军事缩略语大词典》、《俄汉军事大词典》等。收集到的词条必须至少包括以下三项内容：“俄语缩略语”、“俄语全称”及“汉语解释”。

第二步，统一文档格式。由于所得到的词条来源多样，在这里统一存入txt文档中，并以字母顺序排列。

第三步，统一内容格式。将存有收集到的缩略语词条的文档（txt文档）整理为每个词条单独一行，这一行包括“俄语缩略语”、“俄语全称”及“汉语解释”三个部分。使用EmEditor软件打开存有缩略语词条的文档，利用EmEditor软件中的“查找/替换”功能及正则表达式，用“#”号将每项内容隔开，整理为统一规则的形式。之后必须进行人工检查，以确保准确无误。具体内容格式如下图：

```
# А # авианосец атомный # 核动力航空母舰；
# А # авнаразведка, авиационная разведка # 航空侦察；
# А # авиационный # 航空的，飞行的，飞机的；空军的。
```

第四步，导入数据库。利用Access数据库中的数据导入功能，以“#”号为分隔符，将文本中的三项内容分别导入到Access数据库中的三个字段中，并由数据库自动编号。如下图：

ID	缩略语	全称	汉语释义
1	А	авианосец атомный	核动力航空母舰
2	А	авнаразведка, авиационная	航空侦察
3	А	авиационный	航空的，飞行的，飞机的；空军的

第五步，补充数据库内容。以此数据库为基础，再向库中加入“同形”、“形成方式”、“类别”等其他属性字段，并向其中填入属性值，构建俄语军事缩略语知识库。

### 5.3 知识库属性及组织方式介绍

每个缩略语的属性包括ID号、同形、全称、形成方式、汉语释义、类别、语境等。全称，汉语释义这些属性十分容易理解，在这里就不再赘述。下面就对其余的属性加以介绍：

(1) ID号：每个缩略语在数据库中都有唯一的一个ID号，它是每个缩略语各个属性之间相互联系的最可靠的纽带，在检索和信息获取时起着重要作用。

(2) 形成方式：本项属性的值基本上是上文中对于俄语军事缩略语构成方式的分类，即“首字母型，截成型，混合型，符号数字型，音译型”。但是由于“混合型”过于模糊，所以取消了“混合型”。对于“混合型”的缩略语则是在这一属性中将所用到的形成方式都标出，中间用“&”号连接。如：ТуркВО (Туркестанский военный округ, 土耳其斯坦军区)则标注为“抽取&首字母”。

(3) 类别：“类别”按照《俄汉军事大词典》中的“专业分类标注略语”进行分类，并参照《中国军事百科全书》及《俄联邦军事基本情况（2004年版）》，实际上类似于该词的百科知识。从“大军事”的理念出发，广泛包含了气象学、侦察学、情报学甚至于军事有关的组织机构等各类与军事有关的类别。以下是部分标注略语：测（测绘、大地测量）、船（舰船及船艺）、弹（枪弹、炮弹、炸弹）、对抗（电子对抗）、防化（防化兵）、飞（飞机构造、飞行技术）、海（海军、航海）……

(4) 语境：“语境”是指常常与该缩略语同时出现的词语或搭配。库中所举出的例句都拟从在线的俄罗斯国家语料库（<http://ruscorpora.ru>）中检索。

(5) 同形：“同形”是用来描述一个缩略语对应多个全称的情况。对于同一个缩略语对应多个全称的情况，在库中用“同形”这一属性加以区分，其属性值为“1, 2, 3……”等。在这里需要特别指出的是，由于上文提到的部分俄语军事缩略语对于字母大小写的区分，为了处理方便，所有大小写不一样的缩略语都视为不同的词条，而不视为同形。

知识库采用 Access 数据库建立，一条记录即是一个词条，每个词条都具有以上属性。以下是一条记录实例：

ID	缩略语	全称	图例	形成方式	汉语释
9687 П В		пушечное вооружение		首字母型	加农炮 炮

## 6 总结

研究军事缩略语, 建立军事缩略语知识库, 对扩大我们的军事知识、了解军事技术的发展、正确解读俄语军事素材都具有重要的意义。北大的俞士汶教授说过: “语言知识库是自然语言处理系统不可或缺的组成部分, 语言知识库的规模和质量在很大程度上决定了自然语言处理系统的成败”。(俞士汶, 段慧明, 朱学锋: 2004) 由此可以看出缩略语知识库对于缩略语自动处理技术的重要性。事实也已经证明, 仅仅依靠现有的统计或规则的方法而缺少的语言知识的支持, 自然语言处理技术是难以有更大的发展的。

但是总体来说, 目前的军事缩略语知识库的建设还十分缺乏。知识库建设是一项长期且艰巨的任务, 我们应该将更多的精力投入到知识库建设中去, 为自然语言处理技术的发展提供更多的语言资源。

## 参 考 文 献

- [1] 俞士汶, 段慧明, 朱学锋等. 综合型语言知识库的建设与利用[J]. 中文信息学报, 2004. (5)
- [2] 傅兴尚, 许汉成, 易绵竹, 李向东. 俄罗斯计算语言学与机器翻译[M]. 北京: 语文出版社, 2009.
- [3] 常明. 俄语军事缩略语的语法特征, 军事外语教学与研究[J], 北京: 军事科学出版, 2009.
- [4] 张汉宽, 张洋. 俄汉军事缩略语大词典[K]. 北京: 中央编译出版社, 2008.
- [5] 封敏. 俄汉军事大词典[K]. 上海: 上海外语教育出版社, 2007.
- [5] 中国大百科全书编辑委员会. 中国大百科全书[K]. 北京: 中国大百科全书出版社, 1989.
- [7] 中国军事百科全书编审委员会. 中国军事百科全书[K]. 北京: 军事科学出版社, 1997.
- [8] 军事科学院世界军事研究部. 俄联邦军事基本情况(2004年版)[M]. 北京: 军事科学出版社, 2004.