

看语料库用于词典修订的重要性和急迫性 从『摩托』与『马达』的释义

刁晏斌

摘 要 本文利用《人民日报》图文数据库,对照《现代汉语词典》中的释义等,对“摩托”与“马达”的释义及相关问题进行了考察和分析,在此基础上,阐明了语料库用于《现汉》以及其他词典编纂和修订的重要性和迫切性。

关键词 《现代汉语词典》 释义 数据库

笔者曾利用《人民日报》(1946—2005)图文数据库考察和研究现代汉语词汇及其使用的发展变化。在研究过程中,较多地利用了《现代汉语词典》(以下简称《现汉》),并将某些词的释义等与数据库中所反映的实际情况进行对照,结果发现了一些问题。本文就以“摩托”与“马达”的释义为例,作一些分析和探讨,在此基础上,再谈一谈《现汉》修订中语料库使用的重要性和急迫性。

一、“摩托”的释义及使用情况

《现汉》第1—4版“摩托”的释义都是“内燃机。[英 motor]”,到第5版又增加了一个义项“摩托车”。

为了了解“摩托”的表义情况,我们首先考察了《人民日报》1946年10月31日至1950年7月7日的共100个用例,其所表达的意思及用例数如下:

1. 发动机/内燃机,共有15例。如:

(1) 于是临(沂)峰(县)公路上,便发出了坦克和卡车凄厉的摩托

声,他们用强烈火力掩护其残兵溃退。(1947.1.12)

(2) 此外,并没收各该犯非法取得之各种炸药一万一千七百八十五斤,汽车摩托一架,车轮一个,掷弹筒一个,匣枪一枝。(1949.3.27)

按,这里用的是英文原词 motor 的“本义”(详后)，“摩托”既可以说是发动机,也可以说是内燃机,因为人们在指称坦克和汽车等的动力机时通常会用“发动机”,而这些发动机实际上都是内燃机。那么,发动机和内燃机究竟用哪一个更好?我们在下文中再讨论。

2. 以“摩托”为动力的,共 32 例。这里的“摩托”都是做修饰成分的,其所修饰的中心语主要是指车辆等交通运输工具的,我们所见有“车、车辆、卡(车)、脚踏车、自行车、两轮车、运煤车、汽船、工具、机”等。这一意思和用法源于英语的形容词 motor,义为“driven by an engine”,汉译为“由马达推动的”(见《英汉双解大词典》),即 motorbike(摩托车)和 motor dory(摩托艇)中的 motor。下边我们举一个中心语为“车辆”的用例:

(3) 在十八和十九世纪,为征服者轻易而迅速征服的亚洲,现在显示出可惊的能力,顽强地抗拒着有飞机、坦克、摩托车辆及野战炮等新式装备的军队。(1950.1.25)

3. “摩托化”的用例共 33 个。“摩托化”几乎都用于军事,即“摩托化步兵”,指搭乘车辆实施机动作战的部队(见《中国大百科全书》“步兵”条)。《现汉》第 1 版收“摩托化部队”,释义为“全部装备摩托车辆的部队”,第 3 版以后取消该条。英语中与“摩托化”相对应的是 motorized,其动词原形是 motorize,用于军事方面,义为“使(部队)摩托化”(见《全息英汉词典》),比如“摩托化部队”是 motorized troops。

有时“摩托化”只用“摩托”来表示,比较以下二例,这一点就十分清楚了:

(4) 在这里受检阅的包括野战步兵、警卫部队、坦克、榴弹炮、山野炮、高射炮和摩托化部队。(1949.3.28)

(5) 装甲部队、摩托部队、机械化部队、炮兵、骑兵、工兵、步兵都以

所向无敌的英雄气概,要求在最后解放全中国的进军中一显身手,为人民立功。(1949.4.11)

此后也不乏二者互用的例子:

(6) 意大利当局调动大批骑警和摩托警察镇压示威群众,制造了大规模的流血惨案……当罗马示威者昨天在圣保罗广场集会以表示他们对法西斯主义的憎恨的时候,摩托化和头带钢盔的警察就向人群发动了野蛮的袭击(1960.7.8)

“摩托”等于“摩托化”的例子一共有 5 个。

4. 以“摩托”为动力的机械(主要是交通、运输工具),共有 12 例。这与英语中 motor 的一个名词用法相当,但是 motor 一般指汽车,而汉语则主要指摩托车。例如:

(7) 看! 每一个炮手、每一个摩托驾驶手,都是那样精壮有力地举手致敬。由数十辆汽车构成的检阅行列蜿蜒而进。(1949.3.28)

5. 其他 4 例,其中 3 例指发电机,1 例指电动机。

到了 1960 年的《人民日报》中,“摩托”的使用情况略有变化,主要是指称发电机和电动机的用例消失,其他意义保留,全部 36 个用例从第一义到第四义的数量分布分别是 1、16、10、9。

再过了 40 年,到了 2000 年,“摩托”的使用情况有了更大的变化,总体上可以概括为“简化”,即意义和用法趋于单一。《人民日报》2000 年 1 月 1 日到 6 月 1 日的 276 个用例中,以“摩托”为动力义的有 240 例(其中“摩托车”230,“摩托艇”10),义指摩托车的 34 例,另有 2 例是“摩托化”。

此外,我们还在新浪网的“爱问”搜索引擎上进行检索,2006 年 8 月的前 100 个用例也只有以上三种类型,三种类型的数量分布分别是 85、14 和 1。

根据以上事实,我们对《现汉》中“摩托”的释义提出以下两点意见:

第一点是关于“内燃机”的释义,即“内燃机”与“发动机”用哪一个更好。先看《现汉》中“发动机”的释义:“把热能、电能等转化

为机械能的机器,用来带动其他机械工作,如电动机、蒸汽机、涡轮机、内燃机、风车。”由此可知,二者是种属关系,前者的指称范围小,而后的概括范围大。

在早期的工具书中,有以“发动机”来解释“摩托”的,比如《辞海》1936年版中,“摩托”的释义就是“motor之音译,即发动机”。如此看来,在汽车、摩托车等范围内,以“内燃机”或“发动机”来解释“摩托”,都是正确的。那么,哪一个释义更准确、更符合语言运用的实际呢?笔者认为,把“摩托”解释为“内燃机”可能会遇到以下几个困难:

其一,这样的释义没有源语言的根据。英语中 motor 的含义是“a machine that changes power, esp. electrical power, into movement”,汉译为“发动机、马达”(见《朗文英汉双解词典》),而英语中与汉语“内燃机”对应的是词组 internal combustion engine。

其二,“内燃机”不能涵盖所有的用例,因为有的用例中,“摩托”所指并不是内燃机,而是其他类的机器。例如:

(8) 天空闪耀着将要熄灭的晨星,机场上已经发出摩托响声。机械师们比太阳起得还早,他们已经把飞机准备好了。(1952.6.17)

这里的“摩托”显然是飞机上的,下文提到,这些飞机都是喷气式的,所以它们的动力机应当是涡轮喷气发动机(见《中国大百科全书》“飞机”条),而不是内燃机。

其三,在实际运用中,“摩托”与“发动机”的关系比其与“内燃机”更近,或者说“摩托”与“发动机”的关联程度更高。如前所述,现在“摩托”最常见的用法是用于“摩托车”中,而人们在谈到摩托车的动力机时,习惯于用“发动机”而不是“内燃机”,比如在《人民日报》数据库中,“摩托车发动机”的组合形式有 40 例,而“摩托车内燃机”则为 0;“摩托车的发动机”与“摩托车的内燃机”之比为 15:1。其他使用内燃机的车辆基本也是如此,比如“坦克发动机”

与“坦克内燃机”为 7 : 0,“汽车发动机”与“汽车内燃机”为 310 : 15。可以设想,如果“摩托”专指内燃机的话,那么上述数据基本上应当是颠倒过来的。

以下举一个实际的用例:

(9) 春兰生产空调,自己生产空调压缩机;生产冰箱,自己生产冰箱压缩机;生产摩托车,自己生产摩托车发动机。(2000. 4. 3)

如果按《现汉》的释义,此例就不应当用“发动机”,而要用“内燃机”。

鉴于以上三点,笔者认为,“摩托”准确而合理的释义应当是“发动机”而不是“内燃机”。

第二点意见是,《现汉》“摩托”的释义不能涵盖所有的用例,其中最明显的就是不能解释“摩托化”以及与之意思相当的“摩托”,因为这里的“摩托”既非“发动机”或“内燃机”,更不是“摩托车”。时下,在我军和一些外军的军兵种中,“摩托化步兵”(有时简称“摩步”,如“摩步师”)依然存在,所以,词典应当增补词条“摩托化”,释义为“装备各种机动车辆的”,或者是恢复第 1 版的“摩托化部队”这一词条及其释义。

二、“马达”的释义及使用情况

《现汉》第 1—4 版“马达”的释义是“电动机的通称。[英 motor]”,第 5 版改为“电动机。[英 motor]”。

然而,有时“马达”并非指电动机,而是指发动机,并且这样的例子还非常多,如:

(10) 刚挖了一半,坦克的马达声就从背后响起来了。(1954. 1. 24)

(11) 六时,飞机的马达声远远传来。接着,敌人的侦察机在公路上空出现。(1966. 8. 7)

(12) 刚到 8 时 10 分,记者就听到隆隆的汽车马达声,只见一队叙军大卡车装载着各式装备开过来。(2005. 4. 12)

以上三例中,因为“马达”前分别有“坦克”、“飞机”和“汽车”,所以意思非常明显,而以下一例中“马达”与“发动机”互用,前者的所指就更清楚了:

(13) 美国“奋进”号航天飞机上的两名宇航员 10 日通过太空行走为国际空间站太阳能电池板的一对马达罩上了隔热毯……美国宇航局希望这种隔热毯能够在太空夜晚寒冷环境中减少发动机动力的流失并防止工作中断,从而使发动机保持太阳能电池板始终对准太阳,使之产生电力。(2001.12.12)

《人民日报》1946 年 5 月 23 日至 1949 年 12 月 9 日全部 100 个“马达”的用例中,表义的分布情况为:发动机 30,电动机 43,发电机 21,所指不明 6。

所指不明的例子如:

(14) 过去电滚坏了或出了毛病时,非送到上海修理不行,修一次所用的往返运费、修理费,每个马达要数亿元。(1949.2.24)

单就这一句话,看不出“马达”是指电动机还是发动机或发电机。此例“马达”与“电滚”互用,查《现汉》,“电滚子”既指发电机,也指电动机,还是没法确定。再查上下文,本文讲的是发电厂的事情,因此“马达”可能是指发电机。

有时为了准确地表达,就要在“马达”前加上修饰语,或者是采取其他的补充措施,例如:

(15) ……获得发电马达则要费两年半的时间。(1947.7.7)

(16) 同时华北水利推进社领导的机器凿井队在平汉沿线等地也打了井六十七眼,并利用电力马达抽水,增加了水田三万三千五百亩。(1949.5.20)

(17) 马达(电机)嗡嗡地响,牵动天轴飞也似地转,皮带哗啦啦地传动每个机器工作。(1949.10.11)

以上是“马达”50 多年前的使用和表义情况,50 年后的情况怎样,与原来相比有无变化? 截取从 1999 年 11 月 12 日到 2005 年 12 月 24 日的 100 个用例来进行对比分析,结果是:发动机 55,电

动机 35, 发电机 1, 所指不明 9。

“马达”50 年间的变化是相当明显的: 由多指电动机到多指发动机, 而指发电机的用法基本消失(仅有一例为“手摇马达”, 指发报机的手摇发电机, 属于“历史词语”)。

前引《现汉》“发动机”的释义显示, “电动机”与“发动机”也是种属的关系, 与“内燃机”和“发动机”不同的是, 后两者在使用中所指事物绝大多数都是重合的, 如用于摩托车上的既是内燃机, 也是发动机; 而前两者所指多不相同, 如飞机上所用的, 一般只能是发动机而不会是电动机。所以, 《现汉》“马达”释义的问题就更明显一些: 释为“电动机”与语言实际距离太大, 属于比较明显的失误。如果立足于当代, 且取其主要义项, 那么正确的释义应当是“发动机, 有时特指电动机”。

这样, 源于同一个英语词 motor 的“摩托”与“马达”, 也就不像《现汉》释义所显示的那样, 各有自己的表义范围, 区分得清清楚楚了, 实际上, 二者互用的例子并不少见。

三、语料库用于《现汉》修订的重要性和紧迫性

当今, 语料库语言学方兴未艾, 语料库方法在现代语言研究中已经越来越显示出其重要性, 其应用的一个重要方面, 就是词典学的研究, 主要是“通过对词项的搜集、比较、注释和分类, 进行辞书的编辑”^[1]。近年来, 已经不断有人提出利用计算机辅助辞书编纂和修订的问题, 如荣毓敏^[2], 郑恩培、陆汝占^[3], 高永伟^[4]等, 并且也已经有了比较成功的实践。^[5]具体到《现汉》的编纂与修订, 也有人从这一角度提出问题并进行讨论,^[6]并且据说社科院语言所词典室的同仁们也在做与此相关的工作^[7], 但是, 就《现汉》第 5 版修订后的情况来看, 这方面暂时还没有大进展。

时至今日, 在词典的编纂及修订中使用语料库的必要性, 已经不是一个理论问题, 而是一个实践的问题, 正因为各类工具书中存

在那么多相关的可以指摘之处,所以这个问题才变得如此的重要和迫切。所以,以下继续从这方面入手来谈语料库用于《现汉》修订的重要性和紧迫性。

释义方面的情况已经如上所说,而所讨论的“摩托”与“马达”释义的问题并不是个别的现象,如果用大规模真实文本语料库来对《现汉》中的每一个词条进行考察,一定会发现很多类似的问题。

以下简单地谈一下词条的增删,主要是当增不增和删减不当的问题。

先说当增不增的问题。比如“癌”,《现汉》第1版释义为“上皮组织生长出来的恶性肿瘤,如胃癌、肝癌、食道癌、皮肤癌等。也叫癌瘤或癌肿”;第3版保留上述释义,增加了“癌变”一条;第5版释义改为“上皮组织生长出来的恶性肿瘤,常见的有胃癌、肺癌、肝癌、食管癌、肠癌、乳腺癌等”,下边在“癌变”的基础上又增加了一个“癌症”。由第5版看,修订者显然认为“癌瘤”或“癌肿”已经不再作为癌症的别名而使用了,但实际的情况是,这两个别名并未消失,而是一直都在使用,在《人民日报》数据库中,前者有93个用例,后者有124个用例,其中有相当一部分还是近几年的。对《人民日报》的检索还显示,由“癌”在前构成的词还有“癌病”和“癌细胞”,前者有57例,后者有626例,后者的用例数远多于“癌变”的202例。另外,《现汉》所收含有“细胞”的词条有“白细胞、刺细胞、红细胞、卵细胞(第5版删除)、神经细胞(第3版删除、第5版恢复)、吞噬细胞(第3版删除)、腺细胞(第5版删除)”等,考虑到以上两个因素,“癌细胞”也应当立条。至于“癌瘤”或“癌肿”,或者应当在“癌”的释义中恢复“也叫癌瘤或癌肿”,或者可以单立词条。此外,在含有“细胞”的词中,各版《现汉》均未收“体细胞”,而它在《人民日报》中也有200个以上的用例。

再说删减不当的问题。晁继周先生在谈到词条的删减时说:“在词典修订中,我们还应该借助电脑做另一项统计工作,那就是

统计规范型词典已收的词有哪一些一段时间(如2~3年)以来未在报刊上出现过,这个数据可以为修订词典时删减条目提供参考。”^[8]如果真正这样做了的话,上述“摩托化部队”或许就不会在第3版及以后删除了,仅《人民日报》就显示,无论是“摩托化”还是“摩托化部队”,都未在实际的使用中消失。此外,以“摩托化”为关键词在国内几大搜索引擎上进行检索,新浪“爱问”搜索显示,从1998年7月10日到2006年9月3日,共有10600个用例,而在“百度”上则搜索到124000篇相关新闻,Google上的这一数字是122000。

此外,《人民日报》数据库还显示,《现汉》中的一些词条没有用例,如果在一个更加适用于《现汉》等工具书修订的大型语料库中也是如此的话,那就应当考虑删除这些词条了,A字母下这样的词条有音译词“爱美的”、“阿兰若”等。

附 注

[1] 黄昌宁,李涓子.语料库语言学.北京:商务印书馆,2002.

[2] 荣毓敏.计算机·数据库·辞书编纂.辞书研究,1991(1).

[3] 郑恩培,陆汝占.汉语词典编纂计算机化的若干问题.语言文字应用,1999(2).

[4] 高永伟.因特网与词典编纂.辞书研究,2000(4).

[5] 如亢世勇、刘海润主编的《新词语大词典》(上海辞书出版社2003年)。作者说,从词目的确立、意义的考察、词性的确定到典型例句的选择,都是在一个超大规模的动态语料库中进行的,见该词典前言。

[6] 王伟.从《现汉》修订谈计算机在词典编纂中的应用及展望.辞书研究,1997(1).

[7] 这是该室的陆尊梧先生告诉笔者的。

[8] 晁继周.关于汉语规范型词典收词问题的思考.//韩敬体主编.《现代汉语词典》编纂学术论文集.北京:商务印书馆,2004.

(辽宁师范大学文学院 大连 116029)

(责任编辑 刘 凌)