

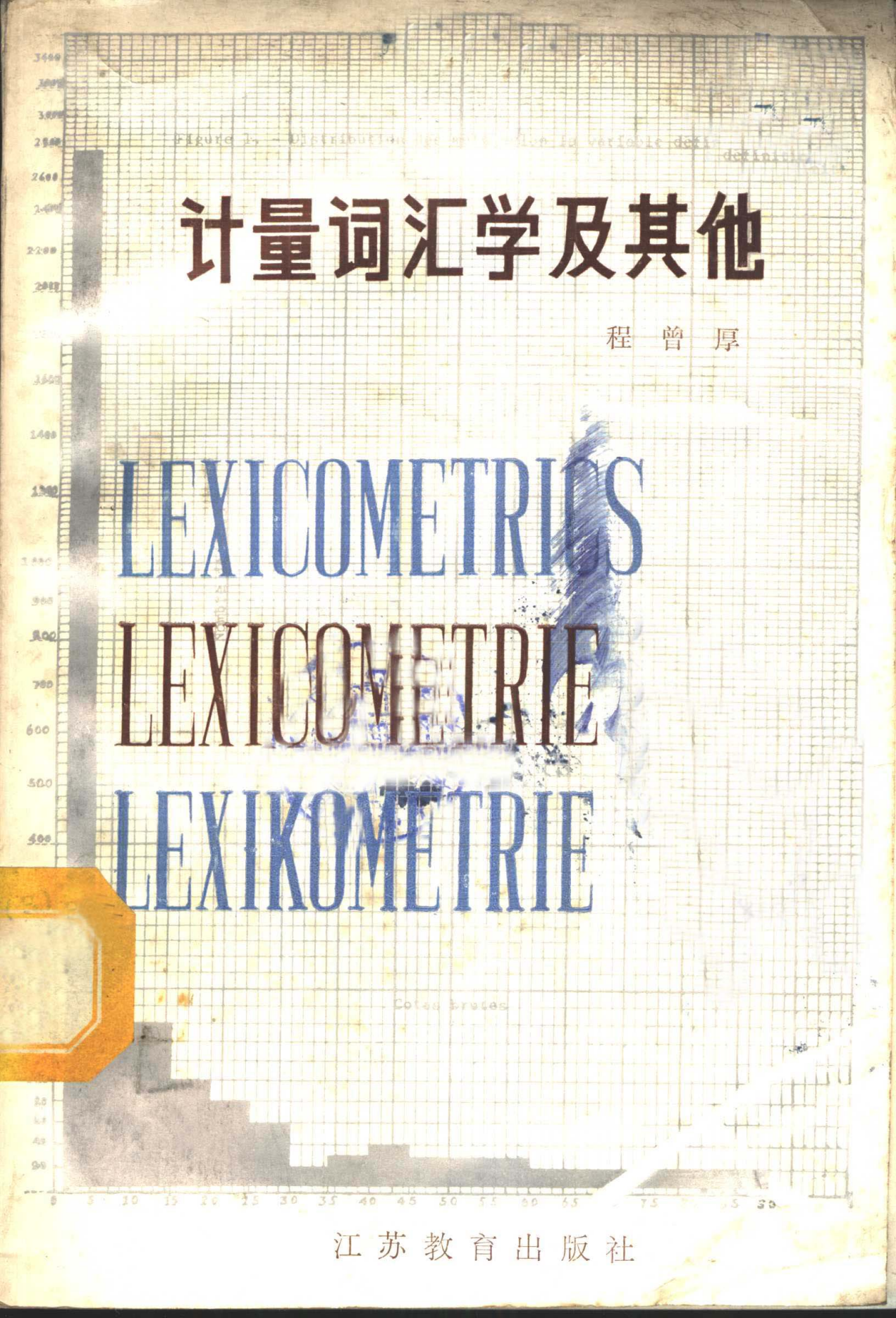


Figure 1. - Distribution des mots selon la variable des...
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
500
400
300
200
100
0

计量词汇学及其他

程 曾 厚

LEXICOMETRICS

LEXICOMETRIE

LEXIKOMETRIE

江苏教育出版社

南京大学语言与语言工程丛书

计量词汇学及其他

程曾厚

江苏教育出版社出版

江苏省新华书店发行 淮阴新华印刷厂印刷

开本 850×1168 毫米 1/32 印张 8.25 插页 3 字数 170,000

1987 年 1 月第 1 版 1987 年 1 月第 1 次印刷

印数 1-3,000 册

书号: 13351·027 定价: 1.65 元

责任编辑 缪咏禾

《语言与语言工程丛书》出版缘起

随着现代科学技术的提高，语言学的研究领域扩大了。近年语言工程学在某些发达国家得到迅速发展。我国起步虽较晚，但少数研究机构和高校在某些领域中，已取得一定成就。近年来，南京大学在汉语、外语（英、法、德、俄、西、日）、声学、计算机科学等专业，分别从事语言学和语言工程学的研究，并取得一些成绩。特别在语言工程方面，进展较快。在这个基础上，我们建立了相应的跨系科的研究机构——语言与语言工程研究所，以加强横向联合研究。为了促进研究工作的发展和展示研究成果。出版这套语言与语言工程丛书。这套丛书范围较广泛，从高技术的学术专著到一般普及读物，都在收录之列。作者一般是南京大学成员，兼收校外专家有关专著。希望得到国内外专家教授的指教。在此我们要特别感谢江苏教育出版社、江苏古籍出版社、南京大学出版社等单位的大力支持和协助。

南 京 大 学

序

程曾厚同志1978年去加拿大拉瓦尔大学研究法语，同时走访该国语言学家，探询语言学的各种新发展，内容有计量语言学、术语学、双语现象研究等。回国后把所见所闻写成介绍文章，在杂志上发表，使读者耳目一新。最近又集合旧作，增加新篇，汇成一书。我得到先睹之快。

近三十年来，我国闭关建设，学术界跟国外往来极少。但是，现在的世界已经进入一个新的时期，跟二次大战以前完全不同了。人们说，现在是地球缩小、知识爆炸，这样的形容并非夸大。我们一方面取得了明显的进展，另一方面又因为跟外界隔阂，在某些部门扩大了差距。今后要使我们的社会主义现代化建设顺利前进，必须改变方针，开门建设。多多了解世界，放大眼界；好好学习外国，取长补短。以不断加快的步伐，缩短历史的差距，追回失去的时间。

世界各国永远处在竞赛之中。一时先进，一时落后，本不足奇。但是如果不能以前车之覆为后车之鉴，那就危险了。我国在清代以前，长期站在世界的前列。“乾嘉盛世”，自以为“天下万物皆备于我”。不知道就在这个“盛世”，别国突飞猛进，我国慢步小进，相对速度变化，落在后头了。鸦片战争以后的丧权辱国，正是闭关政策的惩罚。今天我们还在为“慈禧太后”补课呢！

我国人口众多，经济薄弱，不可能有大量的人出国考察。弥补的办法是，出国的同志回国以后，把所见所闻写成

文章，详细介绍，如实说明。读者阅读以后，可以补益知识，改进工作。在这方面，程曾厚同志可说是已经做了模范了。

程曾厚同志的文章，说理透彻，行文晓畅。他把国外的学术空气、对话情景、以及个人的印象和感受，一一生动描述，写在纸上，使读者不感到是在阅读专门著作，而是神游学海、漫观新奇。这更是难能可贵的！

周 有 光

1982年儿童节在北京。

自序

这是两组我在国外进修回来后陆续写出的文章，其中一部分是在调查访问的基础上整理而成的。这些文字既是我进修收获的一部分，也是我向国家、向同行汇报的一种方式。

进修和访问的部分目的，是想打开一扇窗子，看看外面的情况。所以，我关心的是“闻未所闻，见未所见”的内容。笔者希望根据个人所接触到的专业内容，向我国语言学界和外语教学界介绍一些国内尚未介绍过的语言学新学科、新学派和一些科研机构，这里着重介绍了一个学派（“心理机械论”语言学），两个研究中心（“国际双语现象研究中心”和“纪尧姆著作保存会”），三个学科（“计量词汇学”、“术语学”和“双语研究”），以及麦基教授等几位有成就的语言学家。从这些介绍中可以看到西方语言学新发展、新动向之一斑。凡是访问材料，都有谈话录音、访问记录及有关书面材料作为依据。

本人希望，这些粗浅的介绍除了能有助于了解国外语言学的动态外，还能促进国内外的学术交流，对我国语言学界和外语教学界起到一些有益的借鉴作用，以此为我国的四化建设添上一块小小的砖瓦。

笔者特别感谢语言学界老前辈周有光先生，热心扶掖后进，并在百忙之中，抽出宝贵时间，审阅全书，并撰写序言。这对笔者是莫大的鼓励。

介绍新知识，需要介绍者本人对新知识有一个很好的消

化过程，而笔者对此做得还很不够。有些反映新概念的新术语如何恰当地译成汉语，也颇费推敲，目前的译法并没有很大把握。介绍中错误和疏漏之处，切望有识者不吝指正。

程曾厚 1983年11月

补记：集中各文均成稿于1983年前。极少数地方根据近两年收到的材料有所补充或修正。只有《西方国家计量词汇学的历史回顾》和《历时语言学家马尼埃教授和他的〈语言学引论〉课》两文是在1985年完成的。

作者 1986年1月又识

目 录

周有光先生序

自序

第一部分

计量词汇学 and 基础词汇	1
西方国家计量词汇学的历史回顾	31
萨瓦尔的《词汇价》	58
附录 1: 基本词汇的概念	69
附录 2: “基础英语”词表	75
附录 3: “基础法语”调查词表	94

第二部分

加拿大拉瓦尔大学语言系的教学和科研	141
“国际双语现象研究中心”介绍	154
记著名双语研究专家麦基教授	166
“心理机械论”语言学和“纪尧姆著作保存会”	175
《魁北克法语宝典》——罗曼语方言词典的典范	198
历时语言学家马尼埃教授和他的《语言学引论》课	208
什么是“术语学”?	220
听隆多教授谈术语学	243

计量词汇学和基础词汇

一、外语教学是编制基础词汇表的主要目的

为了语言教学的目的，西方最早制定词表的人可推著名的盎格罗-撒克逊学者艾尔弗里克(Aelfric)(955?—1020?)。他于公元十世纪出版《拉丁语法》，其中包括一部拉丁语-英语小词典。词典按“主题”列出词汇，所以，这就是一张“分类词汇表”。

到了二十世纪，随着各种交流的频繁发展，外语教学的重要性日益明显。为外语教学需要而制定的词表，经常是由外国的外语教师编制的。本国教师给外国学生编制词表反而是稍后的事情。

语言教学用的词表主要是“基础词汇”(英语basic vocabulary, 法语vocabulaire fondamental, 德语Grundwortschatz)。基础词汇不仅是词汇中最稳定的部分，也是最有用的部分。一个学生，以最少的时间和精力，掌握了最少量的基础词汇，能借以阅读大量的作品和表达大量的思想。这无疑为使外语教师深感兴趣的事情。

经过各国学者们几十年来的努力和研究，发现如下一个事实。掌握了一种欧洲语言最常用的1,000个词，可理解任何一篇正常文字的80%的内容，掌握了常用的2,000个词，可理解89%左右的内容，掌握了常用的3,000个词，可理解93%左右的内容，掌握了常用的4,000个词，可理解95%

左右的内容，掌握了常用的5,000个词，可理解97%左右的内容。由此可以推断出，4,000个常用词能构成任何一篇正常文章或谈话的95%的内容，8,000个常用词能包括98%左右的内容，每种语言其他成千上万的词汇总加起来，也仅仅只能解决1%到2%的问题。

基础词汇性质上，特别是数量上不同于一般词汇和专业词汇，这就有一个“选择”的问题。哪些词是最有用的词？选多少词？如何选？这些就是基础词汇表制定工作中碰到的问题。经过半个多世纪各国教育学家、语言学家和心理学家的探索和实践，可以说，大致理出了一些头绪。现在，制定基础词汇表有严密的方法，非但要有专家学者的指导和参予，而且经常动用大量的人力和物力。二十世纪语言学的发展，以及统计方法的日臻完善，更大大地推动了基础词汇的研究工作。这几十年来，取得的成绩是引人注目的。

二、主观的经验标准和“基础英语”词表

从历史上看，基础词汇的选择标准大致有主观标准和客观标准两大类。大体上说，二十世纪以前的各种词表，不管目的和用途如何，事实上都是使用传统的主观经验标准。用主观标准制定的词表，由作者根据自己的学识、修养、经验和兴趣，主观判断一个词是否有用？是否是基础词？用主观标准制定的词表，未必就没有可取之处。事实上，专家的判断力经常不是没有道理的。我们下文还将谈到，时至今日，在基础词汇的选定工作中，主观标准始终是有用的，甚至还是少不了的。

完全用主观标准制定的基础词汇表，影响最大的莫过于

“基础英语”(Basic English)的词表。“基础英语”的概念是由两位英国哲学家提出来的，一个是奥格登(C.K.Ogden)，一个是理查兹(I.R.Richards)。

奥格登和理查兹注意到，在一部词典成千上万条定义所使用的数量更加庞大的词中，某些词一再出现，想到可以用数量极有限的少数词，就足以给一切词下定义，进而凭一张数量很少、但表达力极强的词表，可以表达思维的各种内容。

他们于1928年正式提出“基础英语”词表。“基础英语”本名是“Basic”，原意是一个首母缩略词，指 British American Scientific International Commercial，意思是“英美国际科学贸易”语言。可见，Basic English的本意不是指“基础英语”或某种“简易英语”。两位作者的“抱负”是将Basic作为一种以英语为基础的世界性通用语言。但后人在实际使用中还是把它作为“基础英语”来对待和利用的，而原作者也接受了 Basic English 的提法。

“基础英语”的作者以主观的经验方法，提出一张只有850个词的词表。其中指称事物(things)的名词600个，形容词(qualities)150个，余下100个词为“功能词”(operations)。

600个名词中包括200个可数的“可图解名词”(picturable things)，如 angle(角)，apple(苹果)，knife(刀)等；400个表示各种概念的一般名词(general names)，包括抽象名词，集体名词和亲属词等，如 building(建筑物)，metal(金属)，blood(血)，bread(面包)，knowledge(知识)，father(父亲)。

150个形容词中，有50对“反义词”，如 good-bad(好-坏)，thick-thin(厚-薄)。

100个“功能词”中包括15个表示简单动作的动词，如go(去)，got(得到)，put(放)，take(取)，加上be(是)，have(有)，may(可以)三个动词。20个方位词主要是前置词和后置词，如from(从)，to(向)，down(向下)等。其余为副词、人称代词、连接词及语法词。

这850个基础词能否解决问题？两位作者以及别人用“基础英语”写成各种书籍及文章、题材包括柏拉图的《理想国》，莎士比亚的《朱利叶斯·凯撒》，及论述日常科学问题的叙述文。据说，文章写得清清楚楚，似乎作者的本意基本达到了。我们看到中华书局1949年10月出版的《基本英语文库》书目就有《鲁宾孙飘流记》、《莎氏乐府》、《圣经故事》、《贵妇之死》、《小人国游记》、《凯撒大将》、《蜡烛制造史》及《天文学概论》等八种。

1932年，奥格登出版《基础英语词典》(The Basic Dictionary)，收常用英语词20000个，完全用“基础英语”的850个词释义。

虽然由于第二次世界大战等一些特殊的历史条件，“基础英语”三、四十年代在世界若干地区得到广泛的推广，但今天从外语教学的角度看，“基础英语”词表是不足取的。

“定义词”是判断词汇有用与否的一个因素，但并不是最重要的因素，更不是唯一的因素。把基础词汇的数量定在一千个以内，也不无独到之处，但这是个假象。有人计算过，“基础英语”可以利用“形容词+ly”构成副词等办法，使实际使用的词汇量在二千个以上。

“基础英语”最受批评的一点，是因为很多常用词（特别是动词）不收，提倡用“委婉语”表达思想。这样使语言变得

矫揉造作，极不自然。

由哲学家编制基础词汇表，这在近代是个例外。完全用主观标准制定基础词汇表而获得一定成功，又是一个例外。但事实毕竟是，这条道路以后再也无人问津了。

三、客观标准

二十世纪以来，大家越来越倾向于使用客观标准选择基础词汇。客观标准强调客观和严格，不以作者个人意志为转移。大多数学者认为，这样制定的基础词汇表才能具有科学的价值。这个趋势到了五十年代，就成了不可逆转的大方向了。

这儿要说明的是，历史上的词表都是使用经验的标准，近代还有“基础英语”的先例。但今天完全用客观标准选定的基础词汇表反倒没有先例。今后是否会有，也很难说。一般的做法是，以客观标准为主，主观标准为辅，尽量增加客观标准的比重，减少主观标准的内容。从现有的各种基础词汇表来看，用客观标准选词是一个努力的方向。

今天，人们找到并使用的客观选词标准，有“频率”，“分布率”，“易联想性”，“扩散率”，“覆盖率”等。

四、“频率”标准和克定的德语词表

人们最早发现、最早使用的客观标准是“频率”(英语frequency, 法语fréquence, 德语Frequenz)。虽然在“频率”之外又找到了许多别的标准，但“频率”仍然是最重要和最基本的客观标准。“频率”是指一个词在一定长度的语言取样中出现的次数，而不问语言材料是录音的口语，还是书面

文字。

据说，早在公元九世纪，评注《圣经》的希伯来学者就已经利用“频率”来计算不同版本的经典中的用词情况。捷克人文主义学者夸美纽斯(Comenius, 1592-1670)于1631年写成《通向语言之门》，也利用“频率”概念制定基础词汇表，为语言教学服务。

大规模使用“频率”对词汇进行统计工作是从德国人克定(F. W. Kaeding)开始的。他为了创制一种合理的速记法，对德语词汇作了大量、周详、细致的统计。他于1898年出版《德语词频词典》(Häufigkeitswörterbuch der deutschen Sprache)，这是词汇统计领域中的一部“经典”作品。

克定准备了110份每份为10万词汇量的材料，总词汇量达10,910,777个。材料主要取自报刊杂志(占总词汇量的近40%)等14个不同的领域，留取“频率”在4(包括4)以上的不同词共79,716个。

统计工作在当时是手工进行的。数量巨大的语言材料据说动员了5,000名速记人员及800名合作者才得以完成。

克定的统计中以词形为统计单位，不考虑词义。一词多义以一个词论，单复数却以两个词论。这样处理当然不合理。但是不要忘记，克定的研究工作是为了速记学，不是语言教学。

即使如此，克定的《德语词频词典》为后世德语基础词汇的研究工作打下了良好的基础，对一般的词汇计量工作也富有启发意义。克定搜集的语言素材，注意内容广泛，各种来源要注意比例得当，使统计工作据以进行的语言素材具有代表性。克定搜集的语言材料全是书面语言，这对速记学来说

是可以理解的。

“频率”标准有局限性。不同的语言材料中，词的“频率指数”是不相同的。所以，法国学者米谢阿 (R. Michéa) 说：“一个词的频率并没有人们有时赋予它的固定性。一批材料中的第一个词的频率就未必可靠。频率越往下，这不可靠性就迅速增加。常识告诉我们，如果词的频率之间真的存在一个固定的毋庸置疑的等级，那这张理想的频率词表早就为大家一致通过了。”(Les Vocabulaires fondamentaux, 《基础词汇表》，1964) 法国另一位学者吉罗 (P. Guiraud) 说得更干脆：“整理一批 300,000 个词的语言材料时，第 500 个词以下就不再可靠了；有 1,000,000 个词，才能定出一张大约 1,000 个词的词表；而一张大约有 2,000 到 3,000 个词的词表，要以 5,000,000 个词的取样作为基础。”(Problèmes et méthodes de la statistique linguistique)《语言统计学的问题和方法》，1960)。

五、分布率标准和范德·贝克的法语词表

谁也不否认“频率”是有效的客观选词标准。但是，大家也知道，频率的相对性也是很大的。关键是所使用的语言材料的数量和性质。时代不同，地域不同，材料的长度不同，对“词”这个概念的理解不同，材料是书面语言还是口头语言，都会影响语言材料的性质，从而影响词汇的频率。概括地说，语言素材不同，统计出来的频率也就不同。

每有一批语言材料，就可以求得一个词的不同频率。法国学者缪勒 (C. Muller) 不无道理地发问道：“那问题马上就来了，即在另一批材料中，是否会找到相同的频率呢？”他接

着自己回答：“有些词的使用概率很少变化，不论说话人是谁，不论他谈什么题目，也不论他谈话的环境如何。而也有些词与此相反，它们在某些环境里频率是高的，但又可能长时间完全不出现。”缪勒的结论是：“这就意味着，频率概念如果不马上和分布率概念相结合，那频率概念的价值是很低的。”(Quelques méthodes d'analyse du vocabulaire 《词汇分析方法种种》，1964)

这样，人们想到一个词在若干篇语言素材的取样中出现的篇数多少，也是衡量一个词重要与否的标准，这就是“分布率”(英语 range, 法语 répartition, 德语 Verteilung)。人们对分布率概念的认识比较早。范德·贝克(E. Vander Beke)于1935年出版的《法语词手册》(French Word Book)是引进“分布率”标准进行词汇研究的有代表性的作品。

范德·贝克统计法语词汇所使用的语言素材的总词量超过一百万，达1,147,748个。他共搜集语言材料88篇，每篇词数原则上为一万个，实际平均达13,600个词。88篇材料中，题材相当广泛，甚至包括统计开始时的1926年9-10两月的报刊文章，但以小说、故事和剧本为主，小说、故事共34篇，剧本12篇，共占总词量的56.4%。

范德·贝克词表最主要的特色，是除“频率”外，还采用“分布率”作为标准，而且在选词时以“分布率”标准为主，“频率”标准为副。范德·贝克认为，一个词如果有五位作家各用一次，那也比另一个仅仅被一位作家使用十次的词更重要。

整个统计调查工作是手工进行的，所以动员了大量法语教师，从1926年9月开始，至1927年7月15日结束。88篇材料共收不同词19,253个。最后的词表只收“分布率”指数

在 5 以上的词，客观上“频率”指数最低的也是 5，共 6,067 个。分布率在 5 以下的词有 13,186 个，达 68.5%，其中不少词频率很高，只是因为分布率指数不足 5，所以尽在淘汰之列。

如此强调“分布率”的重要性，不是没有可商榷的地方。但对以前只以“频率”为唯一的选词标准说来，无疑是一大进步。

范德·贝克的词表以书面语言为材料，尤其突出了文学作品的比重。作品写作的平均年代在这次调查统计之前将近半个世纪。所以词表反映的是十九世纪后半期法国小说中的世界，而不是当代的社会内容。词表另一个不足，是和动词相对而言，出现的具体名词很少。不过，这不是范德·贝克词表所特有的缺点，这几乎是任何频率和分布率词表的通病。

“分布率”是一个客观标准，它弥补了“频率”标准的不足。“分布率”给“频率”增加了稳定性的内容。只有同时掌握了一个词的频率和分布率，才能更好地判断一个词的重要与否。大多数学者都把“分布率”看成是“频率”的一个附属标准。以后进行词汇统计时，“频率”和“分布率”就牢牢地结合在一起了。

六、“易联想性”标准和《基础法语》词表

在任何一张“频率”和“分布率”词表中，语法词、工具词很齐全，常用动词和抽象名词也有，而具体名词总是不足，出现得也不稳定，这是一个很大的缺陷。为了解决这个问题，人们继续寻找新的选词标准。人们发现，有些表达具体