

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与，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)。

from to down

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《鲁滨孙飘流记》、《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,选取‘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 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

事物的具体名词，在生活中十分有用，但频率和分布率始终很低。所以，这些词是“频率”标准找不出来的。可是，没有这些词，一张基础词汇表将是不完整的。

原来很多人，包括语言学家自己，常常把“频率高的词”和“常用的词”等量齐观。其实不然，很多“常用的词”频率未必高；有些常用词频率低得叫人吃惊。“嘴”这个词太重要了，连一、二岁的婴孩也知道“嘴”有什么用，但在成年人的日常生活里，“嘴”这个词并不是常常挂在我们嘴上的，所以频率未必很高。

法国学者米谢阿最早提出了“易联想性”（英语 *availability*，法语 *disponibilité*，德语 *Anwendbarkeit*）的概念，并提出了一系列寻找“易联想词”（英语 *available Word*，法语 *mot disponible*，德语 *Verfügbares Wort*）的具体方法，还亲自作了一系列试验性的调查统计。米谢阿下的定义为：“易联想词频率并不特别高，但却是随时都能使用、一有需要立即在思想中出现的词。”（*Limitation et sélection du vocabulaire dans l'enseignement actif des langues vivantes* 《现代语言的积极教学中词汇的限量和选择》）也就是说，一定的语境条件，一定的上下文，会使我们很容易地联想起某些词，主要是指称具体事物的名词。谈到吃饭，我国南方人容易联想起“米饭”北方人容易联想起“馒头”或“面条”。谈到交通工具，城市居民当然会立即联想起“公共汽车”，但即使天天乘“公共汽车”上下班的人，也许在生活里会接连好几天不向任何人说“公共汽车”这个词，更不会写这个词。这些词只有当谈到有关的特定话题时，才会自然而然地在脑海里联想起来，这就是“易联想词”。

最早利用“易联想性”作为选词标准的词汇研究工作，是法国古根内姆 G. Gougenheim) 米谢阿 里旺克 (P. Rivenc) 和索瓦若 (A. Sauvageot) 完成的“基础法语”(Français élémentaire, 1958年改名为 Français fondamental) 词表。米谢阿主要负责“易联想性”标准的理论设想和方法设计工作。

“基础法语”词表由法国政府提供经费、委托国内著名学者具体完成的。为此专门设立“基础法语研究中心”，1959年改名“法语传播研究中心”(Centre de Recherche et d'Etude Pour la Diffusion du Français，简称CREDIF)。动员的人力之多，物力之大，在当时是没有先例的，特别是首次利用了当时称得上是先进设备的录音器材。调查工作从五十年代初开始，1954年完成，同年7月出版《基础法语》，1956年出版《基础法语的制定》(L'Elaboration du Français élémentaire)。

“基础法语”词汇调查所根据的语言材料是当代法语口语，而不是书面语言，这是一大特色。“基础法语”调查共在现场采录了163篇访问材料，访问前考虑到被访问者在地区、职业、年龄、性别、文化程度等方面应具有充分的代表性。访问后由采访的工作人员按统一标准将录音材料转写成文字，共得1,090页打字页，每页300词左右，共312,135个词，包括7,995个不同的词。

从7,995个词中选出频率为20或20以上的词共1,063个。这只是以频率标准选出的词表。最后定稿的《基础法语》(第一阶段)词表的实际词数为1,475个。

从1,063个词的频率词表到1,475个词的正式词表，要

经过一番相当复杂的加工和筛选过程。问题不是简单地加一部分词，而是经过了一番几乎是脱胎换骨的改造。

第一步，从 1,063 个词中只留取频率指数在 29 以上、分布率指数在 5 以上的词，从而淘汰了频率不很高及由于偶然因素获得高频率的词。这样，从 1,063 个词减少至 805 个词。

第二步，从 805 个词再削减 104 个“俗词（如 *machin*，“某物”）；时髦词（如 *formidable*，“厉害”）；不必要的“同义词（如 *bouquin*，“书”）及由于调查环境而频率高的词如 *micro*，“扩音器”等。事实上，1,475 个词中只有不足一半的 701 个词是由“频率”和“分布率”标准提供的。其余 774 个词是“易联想词”和“教学用词”，主要是根据“易联想性”标准选出来的“易联想词”。

统计表明：在上述 805 个高频率的词中，有 253 个语法词，即没有词汇意义的工具词或小品词，这些词频率异乎寻常地高。表上最前面的 38 个词，可占全部语言材料的 50% 的内容，最前面的 278 个词竟占全部语言材料的 80%。但事实上，掌握了这 38 个词，并不能理解 50% 的语言材料，掌握了这 278 个词，并不能理解 80% 的内容。那是为什么呢？语言是交际工具，传递语言信息最多的不是频率极高的语法词，而恰恰是剩下来的 20% 的词。可见易联想词的重要性是很明显的。

寻找“易联想词”的方法是按“主题”（英语 *topics*，法语 *centres d'intérêt*，德语 *Themenkreis*）进行调查。《基础法语》调查人员主要向若干组 9-12 岁的小学生分 16 个主题进行调查，要他们对每个给定的“主题”写下最先想起来的 20 个名

词。每个“主题”进行四组调查，共有四份调查结果。然后，在每个主题下将搜集到的名词按“频率”选出名列前茅的词作为“易联想词”。

调查共有 16 个“主题”。前面五个“主题”为 1、身体各部位(les parties du corps); 2、衣著(les vêtements); 3、住房(la maison); 4、家俱(les meubles de la maison); 5 食品和饮料(les aliments et boissons du repas)。

最后，1,475 个词的总词表中有“词汇词”(mots lexicaux) 1,222 个，语法词(mots grammaticaux) 253 个。其中名词有 692 个，占 46.9%，大部分是以“易联想词”入选的具体名词；动词 339 个，占 22.9%，语法词 253 个，占 17.1%。

《基础法语》名词的收词量显著增多。频率词表共有词 1,063 个，其中名词 395 个，仅占 37.16%。两张词表的差别表现在名词所占的比例由 37.16% 上升为 46.9%。作者们认为，这样的比例才是正常和合理的。《基础法语》词表和以往各种频率词表的主要差别也在这一点。

原作者又以同样的精神和方法选出《基础法语》的第二批词表。1958 年，古根内姆出版《基础法语词典》，包括第一阶段和第二阶段两批词汇共 3,000 余个。

《基础法语》词表的制定，是西方为教学目的用的词汇统计研究中一件兴师动众的大事。这对以后同类的工作，在理论上和实践上，无疑都是有启发意义的。“易联想性”标准的提出，一般评论都予以充分肯定，认为是一个有价值的发现。按“主题”调查“易联想词”的方法也是很可取的，但主题调查的具体方法，还有不少地方有待改进。“主题”应该有

多少个“**是哪些‘主题’**”?易联想词“仅限于名词吗?调查对象为什么限于小学生?这些都是问题。调查对象的人数,地区分布,甚至性别、年龄,也都会对“易联想词”的调查结果产生各种影响。

七、计量词汇学

二十世纪以来,随着外语教学的蓬勃发展,对基础词汇的调查研究越来越多,调查的语言素材范围越来越大,选词标准的设计越来越严格。另一方面,统计学的进步又为词汇统计提供了更多更新的手段和方法。电子计算机的问世更为词汇的调查统计工作打开了新的天地。大量繁琐、单调、机械的手工统计工作可以请机器代劳,又快速,又准确。电子计算机还为发现和使用更新的客观选词标准提供了物质基础,使过去手工统计时代不敢想象的事情有了可能,不敢问津的事情变成了现实。

在种种有利因素的促进下,词汇研究和数学、统计学,特别是和电子计算机的结合,酝酿产生了一门新的学科。这个学科就是“**计量词汇学**”(英语 *lexicometrics*, 法语 *lexicométrie*, 德语 *Lexikometrie*)。法国学者米谢阿早在 1949 年就曾谈到过“**一门年青而富有前途的科学,即统计词汇学 (la lexicologie statistique)**”(Introduction pratique à une statistique du langage,《语言统计的实用导论》),法国学者吉罗于 1960 年出版了《统计语言学的问题和方法》,1968 年缪勒出版的《统计语言学入门》(Initiation à la statistique linguistique)称之为“**定量语言学**”(linguistique quantitative)还有的学者叫作“**数学语言学**”(linguistique

mathématique).

德国学者哈特曼 (R. Hartmann) 提出的名称是“语言学中的数学统计方法 (die mathematisch-statistische Methode in der Sprachwissenschaft)”，又称作“词汇频率研究 (Worthäufigkeitszählung)” (Bedeutung und Entwicklung der Worthäufigkeitsforschung, 《词汇频率研究的意义和发展》), 总之, 对这个新学科建议的名称很多。德国学者凯尔 (R. D. Keil) 在《国际语言教学的应用语言学杂志》1965年5月号发表《计量词汇学中统一的方法》(Einheitliche Methoden in der Lexikometrie)。他在文章中写道: “在 de Tollenaere 的鼓励下, 我建议将“计量词汇学”作为词汇学的一个分支的专业术语。这一分支用统计方法给一切可测定的词汇现象提供数据”。这样, “计量词汇学”不是作为一门独立的学科, 而是词汇学的一个分支, 目的是给词汇的各种可以测量的特性提供数据。

“计量词汇学”的提法还是比较新的。六十年代出版的大部份语言学词典都不载这个词。美国纳什 (Rose Nash) 于1968年出版的《英、俄、德、法语言学和语文学词汇》中不见这个词。1973年法国杜布阿 (J. Dubois) 主编的《语言学词典》中也不收这个词 却收了“统计词汇学”的词条。国内翻译过来的《语言和语言学词典》(哈特曼和斯托克著 上海辞书出版社, 1981年) 也不收“计量词汇学”一词 只收“词汇统计”的条目。

当然, “计量词汇学”不仅对制定为语言教学用的基础词汇表有用, 它也能词汇学和词典学服务, 对低年级的教材编写能提供有关词汇的统计材料。计量词汇学对文学语言的

分析也很有用，可借以比较不同时代、不同风格和不同作家的用词特点。此外，计量词汇学有助于揭示语言中词汇结构的复杂性，有助于为语法学家确定不同词类的性质和功能。

八、寻找新标准——“扩散率”

计量词汇学一方面努力使已有的选词标准日臻完善，另一方面借助电子计算机和现代统计方法寻找新的选词标准。

1964年，美国斯坦福大学教授久兰德 (A. Juilland) 及其合作者出版《西班牙语词频词典》(Frequency Dictionary of Spanish Words)。久兰德首先提出并使用了“扩散率”(英语 dispersion, 法语 dispersion)。他认为“分布率”标准还嫌粗糙，因为“分布率”掩盖了确定一个词汇是否重要时有参考价值的某些事实真相。久兰德把“分布率”称作“简单扩散率”(simple dispersion)，而进一步提出还有“复杂扩散率”(Complex dispersion)。单有“简单扩散率”是不够的，因为在词汇的调查统计中，频率指数相同、分布率指数也相同的词，实际上在各篇语言材料中的“分布”情况可以是很不相同的。

为了说明问题，不妨举一个假设的例子。假定在需要统计的七篇语言材料中，有 A 和 B 两词，频率指数都是 210，分布率指数都是 7。如果由此得出结论认为，A 和 B 两词同

语言材料	1	2	3	4	5	6	7
A -	1	1	1	1	1	1	204
B -	30	30	30	30	30	30	30

等重要，那只是个假象，实际情况可能远非如此。

任何人都不会毫不犹豫地得出结论：B词比A词更重要，原因呢？因为B词在每篇材料中分布均匀，A词的分布是极不均匀的。

用什么方法把这个明显的事实在数量上反映出来呢？按照繆勒提供的计算公式，要进行好几道数学运算，相当复杂，好在计算工作请计算机完成，作者只须设计方法，编制程序，不要自己动手逐词演算。

运算过程大致如下：

第一步，基本数据

A词 - 1 1 1 1 1 1 204

B词 - 30 30 30 30 30 30 30

A和B频率指数 F 相同， $F = 210$

A和B分布率指数 R 相同， $R = 7$

理论平均数 $\bar{x} = 210/7 = 30$

第二步，求扩散率指数 (D) (扩散率指数取决于出现该词的材料篇数 (n) 和与理论平均数 (\bar{x}) 相比较的每篇实际频率指数 X_i)

求离散差 (σ^2) (即与理论平均数 \bar{x}) 相比较的偏差平方的平均值)

	A 词			B 词		
	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
x_1	1	-29	841	30	0	0
x_2	1	-29	841	30	0	0
x_3	1	-29	841	30	0	0