

## 术语形成的经济律——FEL 公式\*

冯志伟

(教育部语言文字应用研究所, 北京 100010)

**摘要:** 本文分析了单词型术语和词组型术语在术语数据库 GLOT-C 中分布, 试图从理论上解释为什么在术语系统中词组型术语占大多数的这个重要的术语现象, 在此基础上提出了“术语形成的经济律”, 并且用 F E L 公式来描述这个定律。

**关键词:** 单词型术语; 词组型术语; 术语形成经济律; FEL 公式

## Economic Law of Term Formation——FEL Formula

FENG Zhiwei

(Institute of Applied Linguistics, the Ministry of Education, Beijing, 100010)

**Abstract:** In this paper, the author analyzes the distribution of word-term and phrase-term in terminological database. He tries to theoretically explain why the phrase-terms are the majorities in a terminological system. Based on this analysis, the author proposes the economical law of term formation, and he uses FEL formula to describe this law.

**Keywords:** word-term; phrase-term; economic law of term formation; FEL formula

现代科学技术日新月异地发展, 新的科学概念层出不穷, 人们不可能给每一个新出现的概念都用一个新的单词来命名它。在大多数情况下, 会采用由原有的单词组合起来构成词组来表示新的概念, 这样, 就会形成许多以词组为形式的术语, 它们叫做词组型术语 (phrase-term)。从目前发展的趋势来看, 词组型术语越来越多, 在整个术语系统中占了很大的一部分。这几乎成了现代术语发展的一个规律。当然, 单词型术语 (word-term) 仍然是术语系统中的重要组成部分。

如果我们从语言学的观点来看术语系统, 那么, 我们可以看出, 单词只不过是术语的构成材料 (它既是词组型术语的构成材料, 也是单词型术语的构成材

料)，而术语(包括词组型术语和单词型术语)则是由这些构成材料形成的产品。因此，可以说，一切术语都是由单词构成的。在我们设计的数据处理中文术语数据库 GLOT—C 术语数据库中的 1 510 条单词型术语和词组型术语，都是由 858 个不同的单词构成的。这 858 个单词，与 1 510 条术语的数量比较起来，只是一个较少的数目。这种由少量的单词构成大量术语的语言现象，反映了语言使用中的经济原则，我们把它叫做“术语形成的经济律”(economic law of term formation)。

早在十九世纪初年，德国杰出的语言学家和人文学者洪堡德(Von Humboldt, 1767—1835)就观察到“语言是有限手段的无限运用”。但是，由于当时尚未找到能够证实这种论断的技术工具和方法，这种论断只是停留在科学假设阶段。

究竟如何来理解“语言是有限手段的无限运用”这个科学假设呢？

美国语言学家乔姆斯基(1928— )认为：

“一个人的语言知识是以某种方式体现在人脑这个有限的机体之中的，因此语言知识就是一个由某种规则和原则构成的有限系统。但是一个会说话的人却能讲出并理解他从未听到过的句子以及和我们听到的不十分相似的句子。而且，这种能力是无限的。如果不受时间和注意力的限制，那么由一个人所获得的知识系统规定了特定形式、结构和意义的句子的数目也将是无限的。不难看到这种能力在正常的人类生活中得到自由的运用。我们在日常生活中所使用和理解的范围是极大的，无论就其实际情况而言还是为了理论描写上的需要，我们完全有理由认为人们使用和理解的句子范围都是无限的。” [1]

乔姆斯基以“句子”的使用和理解为例进一步说明了“语言是有限手段的无限运用”这一科学假设。

如果我们把乔姆斯基这段话中的“句子”改为“术语”，可以类推地得到如下的假设：

“一个人的语言知识是以某种方式体现在人脑这个有限的机体之中的，因此语言知识就是一个由某种规则和原则构成的有限系统。但是一个会说话的人却能讲出并理解他从未听到过的术语以及和我们听到的不十分相似的术语。而且，这种能力是无限的。如果不受时间和注意力的限制，那么由一个人所获得的知识系统规定了特定形式、结构和意义的术语的数目也将是无限的。不难看到这种能力在正常的人类生活中得到自由的运用。我们在日常生活中所使用和理解的范围是极大的，无论就其实际情况而言还是为了理论描写上的需要，我们完全有理由认为人们使用和理解的术语范围都是无限的。”

通过这样的类推我们可以看出，“语言是有限手段的无限运用”这个科学假设也可以应用到术语学中，我们不妨把这个科学假设称为“术语生成性假设”(hypothesis on term generation)。

根据“术语生成性假设”，我们可以得到如下三个命题：

命题 1：任何一个会说话的人在他的知识范围内都有能力猜测或理解他从未听到的术语，从而认识有关的科学概念；

命题 2：任何一个会说话的人在他的知识范围内都有能力创造或说出他从未听到的术语，从而表达有关的科学概念；

命题 3：任何一个会说话的人在他的知识范围内都有能力在少量的单词型术语的基础上创造出大量的词组型术语，从而使得在一个术语系统中，词组型术语的数量大大地超过单词型术语的数量。

命题 1 和命题 2 是显而易见的，它们可以由“术语生成性假设”直接推导出来，而命题 3 则不十分明显，需要我们通过科学实验和数学计算来加以检验。

命题 3 在术语学中究竟是否存在呢？这是本文需要通过科学实验和数学计算来检验的内容。

今天，我们有了电子计算机这个有力的技术工具，通过科学实验和数学计算来检验命题 3 这个大胆的科学假设的时机已经成熟。

我们的实验数据充分地说明，本文提出的“术语形成的经济律”验证了命题 3。因此，“术语形成的经济律”正是洪堡德的“语言是有限手段的无限运用”这一假设的一个科学实例。术语系统中单词型术语的数目是有限的，而由单词型术语构成的词组型术语的数目却是无限的。由少量的、有限的单词构成大量的、无限的术语，这正是“有限手段的无限运用”这一假设在术语学中的具体表现。可见，“术语形成的经济律”是一个有着深刻的语言学和哲学背景的普遍性规律。

在本文中，我们将讨论术语形成的经济律的三个基本概念：术语系统的经济指数、单词的术语构成频率和术语的平均长度。并且提出“FEL 公式”来描述这三个基本概念之间的关系。

### 一 术语系统的经济指数

为了说明什么是术语系统的经济指数以及术语系统的经济规律，需要先定义如下的初始概念：

1. 系统的术语数：在一个术语系统中，不同的术语的总数，也就是术语系统的容量。系统的术语数用  $T$  表示，它的单位是“条”。

2. 单词的绝对频率：在术语系统中，某一个词的出现次数(或使用次数)。词的绝对频率用  $\alpha$  表示，它的单位是“次”。

3. 不同单词数：具有同一频率的不同单词的数目。不同单词数用  $\nu$  表示，它的单位是“词”。

4. 不同单词的总数：在术语系统中，具有不同绝对频率的不同单词的总数。不同单词总数用  $W$  表示，它的单位是“词”。不同单词总数的计算公式是：

$$W = \sum v$$

5. 运行单词数：具有同一绝对频度的不同单词  $v$  和它的绝对频度  $\alpha$  的乘积。运行单词数用  $\rho$  表示，它的单位是“词次”。运行单词数的计算公式是：

$$\rho = \alpha v$$

6. 运行单词总数：具有不同绝对频率的运行单词的总数。运行单词总数用  $R$  表示，它的单位是“词次”。运行单词总数的计算公式是：

$$R = \sum \rho = \sum \alpha v$$

术语系统的经济指数就是系统的术语数  $T$  被不同单词总数  $W$  来除所得的商。术语系统的经济指数用  $E$  来表示。这样，我们有如下公式：

$$E = T/W \cdots \cdots \cdots (1)$$

$E$  的单位是“条/词”，读为“每词多少条”。

在大多数术语系统中， $E > 1$ 。如果  $E \leq 1$ ，则说明术语系统设计的经济效应不高。例如，在术语系统  $GLOT-C$  中， $T=1510$ ， $W=858$ ，则该系统的经济指数  $E$  为：

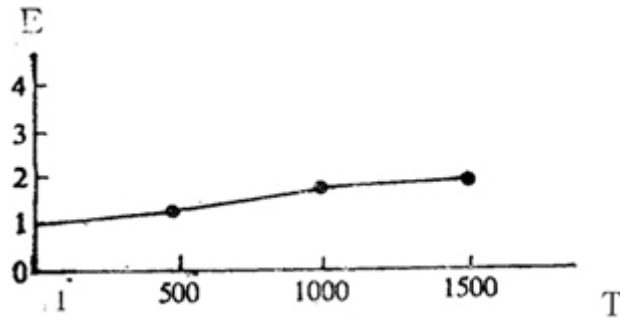
$$E = T/W = 1510/858 = 1.76$$

这说明，当术语系统有 1 510 条术语时，每个单词平均可构成 1.76 条术语。可见，这个术语系统具有较高的经济效应，也就是说，在该系统中，每个单词构成的术语条数较多。

术语系统的经济指数的高低，受到系统中术语数的强烈影响。随着系统的术语数的增加，术语系统的经济指数也逐渐升高，在我们设计的“数据处理中文术语数据库” $GLOT-C$  中，当系统的术语数为 500 条，不同单词数为 342 个词时，其经济指数为 1.46；当系统的术语数增加到 1 000 条，不同单词数增加到 588 个词时，其经济指数也增加到 1.70；当系统的术语数进一步增加到 1 510 条，不同单词数进一步增加到 858 个词时，其经济指数也进一步增加到 1.76。如下表所示：

T	W	E
500	342	1.46
1000	588	1.70
1510	858	1.76

这种情况，可图示如下：



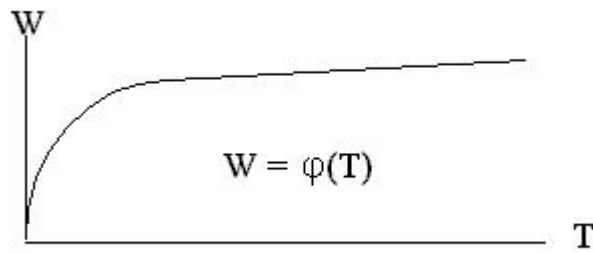
在一定的学科领域内，如果具有大量术语条目的术语系统具有较高的经济指数，那么，这个系统必定具有大量的由少数基本单词构成的词组型术语，而这些词组型术语构成了该术语系统的主要部分。

## 二 单词的术语构成频率

在术语系统中，每个单词的绝对频率并不是一样的。有的单词经常使用，叫做高频词，有的单词不常使用，叫做低频词。随着术语条目的增加，高频词的数目一般来说也相应地增加，而新词出现的可能性越来越小。这时，尽管术语的条数还继续增加，不同单词总数增加的速率却越来越小，而高频词则反复地出现。在术语数 T 与不同单词总数 W 之间，存在着如下的函数关系：

$$W = \phi(T)$$

这种函数关系可粗略地用下图表示：



术语系统的高频词越多，则由这些高频词构成的术语也越多。单词构成术语的这种能力，叫做单词的术语构成频率。

单词的术语构成频率就是在一个术语系统中运行单词的总数  $R$  被不同单词数  $W$  来除所得商。单词的术语构成频度用  $F$  表示。这样，我们有下面的公式：

$$F=R/W \dots\dots\dots (2)$$

$F$  的单位是“次”。事实上，因为  $R$  的单位是“词次”， $W$  的单位是“词”，所以  $F$  的单位就是“词次/词”，它恰恰等于“次”。

$F$  的值不能小于 1，即  $F \geq 1$ 。对于同一个术语系统来说，单词的术语构成频率  $F$  不能小于术语系统的经济指数  $E$ ，即  $E \leq F$ ，因为我们总是有  $T \leq R$ 。

在 GLOT-C 中，1 510 条术语的运行单词总数为 3 216 个，而构成这 1 510 条术语的不同单词总数为 858 个，即  $R=3216$ ， $W=858$ 。这样，我们有：

$$F=R/W=3216/858=3.75$$

这说明，当 GLOT-C 系统的术语数为 1 510 条时，其单词的术语构成频率为 3.75，也就是说，平均每个单词可以出现 375 次。因此，这个值也可以代表这些单词构成术语的平均频率。

单词的术语构成频率也受到术语系统中术语数的影响。在 GLOT-C 术语数据库中，当术语数为 500 条时 ( $T=500$ )，单词的频率表如下：

$a$	$\nu$	$\rho$	$a$	$\nu$	$\rho$	$a$	$\nu$	$\rho$
1	181	181	8	4	32	16	1	16
2	66	132	9	3	27	19	1	19
3	32	96	10	6	60	20	1	20
4	19	76	11	1	11	26	1	26
5	8	40	12	1	12	27	1	27
6	4	24	13	4	52	37	1	37
7	5	35	15	1	15	49	1	49

此时， $W = \sum v = 342$ 。并且， $R = \sum \rho = 987$ 。因此， $F = R/W = 987/342 = 2.89$ 。

当系统中的术语数为 1 000 条 ( $T=1000$ ) 时，单词的频率表如下：

$a$	$v$	$\rho$	$a$	$v$	$\rho$	$a$	$v$	$\rho$
1	295	295	12	6	72	25	1	25
2	103	206	13	2	26	26	1	26
3	54	162	14	2	28	29	1	29
4	36	144	15	2	30	33	1	33
5	19	95	17	2	34	37	1	37
6	16	96	19	1	19	48	1	48
7	12	84	20	1	20	51	1	51
8	10	80	21	1	21	52	1	52
9	6	54	22	1	22	64	1	64
10	6	60	23	1	23			
11	2	22	24	1	24			

此时， $W = \sum v = 588$ 。并且， $R = \sum \rho = 2072$ 。因此， $F = R/W = 2072/588 = 3.52$

当系统的术语数为 1 510 条 ( $T=1510$ ) 时，单词的频率表如下：

$a$	$v$	$\rho$	$a$	$v$	$\rho$	$a$	$v$	$\rho$
1	411	411	13	5	65	26	2	52
2	150	300	14	3	42	27	2	54
3	73	219	15	2	30	33	2	66
4	52	208	16	3	48	34	1	34
5	44	220	18	1	18	35	1	35
6	24	144	19	3	57	44	1	44
7	14	98	20	1	20	47	1	47
8	14	112	21	2	42	55	1	55
9	13	117	22	1	22	56	1	56
10	5	50	23	2	46	63	1	63
11	8	88	24	3	72	68	1	68
12	7	84	25	2	50	79	1	79

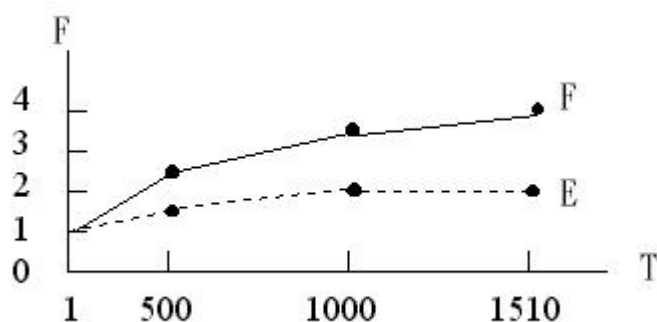
此时， $W = \sum v = 858$ 。并且， $R = \sum \rho = 3 216$ 。因此， $F = R/W = 3216/858 = 3.75$ 。

我们可得到如下的表：

T	W	R	F
500	342	987	2.89
1000	588	2072	3.52
1510	858	3216	3.75

从表中可以看出，随着系统中术语数的增加，单词的术语构成频率也相应地

增加。图示如下：



在上图中，虚线表示系统的经济指数 E 的变化情况，实线表示单词的术语构成频率 F 的变化情况。如果术语数 T 相同，单词的术语构成频率 F 的值不小于系统的经济指数 E 的值，即  $F \geq E$ 。仅当术语数  $T=1$ ，系统中只有一个单词时，F 等于 E，在其它场合，F 永远大于 E。

从上面三个频率表中还可看出，随着单词绝对频率  $\alpha$  的增加，具有同一绝对频率的不同的单词的数目  $v$  相应地减小，这种关系可用下图来描述：



这说明，在一个术语系统中，高频词只占了不同单词总数的一小部分，而它们却能构成大量的术语。例如，在“数据处理中文术语数据库 GLOT-C 中当术语数 T 为 1 510 条时，绝对频率大于 10 的高频词只有 62 个，而它们的出现次数却是 1 342 词次。由这些高频词构成的运行词总数占了全部的运行词总数的 41.4%。术语系统中的高频词越多，则该系统中单词的术语构成频率也就越高。

### 三、术语的平均长度

包含在术语中的单词数，叫做术语的长度。在一个术语系统中，术语的最小长度为 1。单词型术语的长度永远等于 1，每个单词型术语只能包含一个单词。



例如，“程序”这个单词型术语的长度为 1。词组型术语的长度永远大于 1。例如，“程序/设计”这个词组型术语的长度为 2，“数字/字符/子集”这个词组型术语的长度为 3，“条件/控制/转移/指令”这个词组型术语的长度为 4，“平均/无/故障/工作/时间”这个词组型术语的长度为 5，等等。从术语经济原则的观点看来，术语的长度太长，不便于使用和记忆，因而，我们有必要研究术语的长度问题。

从术语系统的整体来看，还应该研究术语的平均长度。在一个术语系统中，术语的平均长度就是运行单词总数 R 被术语数 T 来除所得的商。术语的平均长度用 L 表示。计算公式为：

$$L=R/T$$

L 的单位是“词次/条”，读为“每条多少词次”。

L 的值永远不小于 1，即  $L \geq 1$ 。在每一个术语都只由一个单词构成的术语系统中， $L=1$ ，在其它场合， $L > 1$ 。

在“数据处理中文术语数据库”GLOT—C 中， $R=3\ 216$ ， $T=1\ 510$ ，所以，该系统的术语平均长度为：

$$L=R/T=3216/1510=2.130$$

这意味着，在 GLOT—C 中，当术语数等于 1510 条时，平均每条术语由 2.130 个单词构成，即每条术语中含有 2.130 词次。

随着术语系统中术语数的增加，术语的平均长度也有增加的趋势。在 GLOT—C 系统中，当术语数为 500 条时，术语的平均长度为 1.974 词次/条；当术语数为 1000 条时，术语得平均长度为 2.072 词次/条；当术语数为 1510 条时，术语的平均长度为 2.130 词次/条。当然，术语的平均长度不能太长，每个术语系统都能在其运行过程中，不断地把术语的平均长度调节到最佳值。在这个调节的过程中，某些太长的术语被淘汰了，某些较短的术语变长了，这样，术语的平均长度就可以保持相对的稳定。

#### 四、术语构成的经济律——FEL 公式

前面我们讨论了术语构成的三个主要概念：术语系统的经济指数 E、单词的术语构成频率 F 和术语的平均长度 L。现在我们进一步研究这三个概念之间的关

系。仔细观察 GLOT—C 术语数据库的实验数据，我们可以发现：术语系统的经济指数 E 与术语的平均长度 L 的乘积与单词的术语构成频率之值是近似地相等的。

实验数据如下：

T	E	L	$E \times L$	F
500	1.46	1.974	2.88304	2.89
1000	1.70	2.072	3.52140	3.52
1510	1.76	2.130	3.74880	3.75

当 T=500 时，我们有  $E \times L = 2.88304$ ，而这时  $F = 2.89$ ；当 T=1 000 时，我们有  $E \times L = 3.52140$ ，而这时  $F = 3.52$ ；当 T=1 510 时，我们有  $E \times L = 3.74880$ ，而这时  $F = 3.75$ 。可以看出， $E \times L$  之值与 F 之值几乎是相等的。

根据这些实验数据，我们可以在 E、F 和 L 之间建立如下的数学关系：

或者

$$E \times L = F$$

$$F = EL$$

事实上，由于

$$E = T/W \quad (1)$$

以及

$$F = R/W \quad (2)$$

(2) ÷ (1) 得到

$$F/E = R/T \quad (3)$$

根据术语平均长度的定义，我们有

$$L = R/T \quad (4)$$

比较 (3) 与 (4)，可以得到：

$$F/E = L$$

因此，可有

$$F = EL$$

这就是上面的 FEL 公式。

由此，我们可以得出结论：在一个术语系统中，术语系统的经济指数 E 与术语的平均长度 L 的乘积恰恰等于单词的术语构成频率 F 之值。我们把这个规律，

叫做“术语形成的经济律”。

从 FEL 公式，我们还可得到如下的推论：

1、在一个术语系统中，当术语的平均长度  $L$  一定时，单词的术语构成频率  $F$  与术语系统的经济指数  $E$  成正比。术语系统的经济指数越高，单词的术语构成频率也越高。这时，FEL 公式变为：

$$F=k_1E \text{ (} k_1 \text{ 是一个常数)}$$

这说明，为了提高术语系统的经济指数，应该增加单词的术语构成频率，使得每个单词能构成更多的术语。

2、在一个术语系统中，当系统的经济指数  $E$  一定时，单词的术语构成频率  $F$  与术语的平均长度  $L$  成正比。术语的平均长度越长，单词的术语构成频率越高。这时，FEL 公式变为：

$$F=k_2L \text{ (} k_2 \text{ 是一个常数)}$$

这说明，为了提高单词的术语构成频率，必须增加术语的平均长度，因为系统的经济指数是一定的，每个单词只能被包含到有限数目的术语之中，所以，只有增加术语的平均长度。

3、在一个术语系统中，当单词的术语构成频率  $F$  一定时，系统的经济指数  $E$  与术语的平均长度  $L$  成反比。系统的经济指数的增加将会引起术语平均长度的缩小，而系统的经济指数的减少将会引起术语平均长度的增长。这时，FEL 公式变为：

$$EL=k_3 \text{ (} k_3 \text{ 是一个常数)}$$

这说明，在不改变单词的术语构成频率的条件下，如果我们想提高术语系统的经济指数使得每个单词能够构成更多的术语，那么，我们只好从原有的术语中，抽出一些单词来构成新的术语，这样，术语的平均长度就缩短了。因为在这种情况下，运行单词总数是不变的，我们必须从原有的术语中，一般是从较长的术语中，抽出一部分单词来造成新的术语，而这将引起术语数目的增加。其结果，术语系统的某些术语中所包含的单词数可能会减少，而新术语的长度不可能太长，因而系统中术语的平均长度就缩短了。

由此可见，FEL 公式反映了术语系统的经济指数、单词的术语构成频率以及术语的平均长度之间的相互依存和相互制约的关系。这个公式是支配着术语的形

成和变化的一个经济规律。

从 FEL 公式，我们可得到：

$$E=F/L$$

由此我们可知，提高术语系统的经济指数的方法有两个：

- 1、在不改变单词的术语构成频率的条件下，缩短术语的平均长度；
- 2、在不改变术语的平均长度的条件下，提高单词的术语构成频率。

一般地说，在一个术语系统中，最好不要过大的改变术语的平均长度。术语的平均长度改变过大，往往会使术语系统改变到人们难以辨认的程度。由于这个原因，我们最好不要使用缩短术语平均长度的方法来提高术语系统的经济指数。看来，提高术语系统的经济指数的最好方法，还是在尽量不过大地改变术语的平均长度的前提下，增加单词的术语构成频率。这样，在术语形成的过程中，将会产生大量的词组型术语，使得词组型术语的数量大大地超过单词型术语的数量，而成为术语系统中的大多数。在我们设计的“数据处理中文术语数据库”GLOT—C 中，词组型术语占了 75.17%。这个事实，正是术语形成经济律作用的结果。而术语形成的经济律又是洪堡德提出的“语言是有限手段的无限运用”这一普遍假设在现代术语学中的实际体现和科学证明。

\* 本文原文为英文，22 年前发表于 *Social Sciences in China* (No.4, 1988)。本刊发表前，经作者翻译成中文并做了一些修改。——编者注

#### 参考文献

[1] N. Chomsky, 《乔姆斯基序》，载《乔姆斯基语言理论介绍》，第 1~2 页，1982 年，黑龙江大学出版。