

核心术语集的定义与数学模型研究

杨福义¹, 叶其松²

¹鞍山师范学院, 辽宁 鞍山

²黑龙江大学俄罗斯语言文学与文化研究中心, 黑龙江 哈尔滨

收稿日期: 2021年11月21日; 录用日期: 2021年12月11日; 发布日期: 2021年12月27日

摘要

在分析齐普夫(Zipf's Law)定律的数学曲线及其函数的基础上, 结合二八定律、不平衡定律、冯志伟术语经济率, 将术语频率曲线的连续函数转为离散数据的精准计算, 给出术语系统核心术语的定义和核心术语数据集的计算模型, 计算了一些专业的核心术语词表, 可用于术语审定和研究核心术语的定义问题, 为术语审定和定义的编写方案提供参考。也可以为语言学研究与应用中核心词汇数据集与句子短语核心模式集提供参考。

关键词

齐普夫定律, 二八定律, 核心术语, 数学模型, 计算术语学

Research on the Definition and Mathematical Model of Core Term Set

Fuyi Yang¹, Qisong Ye²

¹Anshan Normal University, Anshan Liaoning

²Center for Russian Language Literature and Culture Studies, Heilongjiang University, Harbin Heilongjiang

Received: Nov. 21st, 2021; accepted: Dec. 11th, 2021; published: Dec. 27th, 2021

Abstract

Based on the analysis of the mathematical curve and function of Zipf's law, combined with the 28 law, the imbalance law and Feng Zhiwei's term economic rate, the continuous function of the term frequency curve is transformed into the accurate calculation of discrete data, and the definition of the core terms of the term system and the calculation model of the core term data set are given. Some professional core term lists are calculated, which can be used for term validation and Re-

search on the definition of core terms, and provide reference for the preparation scheme of term validation and definition. It can also provide a reference for the core vocabulary data set and the core pattern set of sentence phrases in linguistic research and application.

Keywords

Zipf's Law, 28 Law, Core Terms, Mathematical Model, Computational Terminology

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

核心术语是学科的骨架,影响着理论的概貌展现和力量发挥。它一经形成便作为观察、描述、解释中不可或缺的工具发挥效用,其体系建构是其学科发展的一项核心工作。这项工作本身就是一种科学研究,须按照一定的科学原则来进行。核心术语的形成与科学分类是紧密联系在一起。建构成体系且相关联的核心术语为学科确定了研究焦点,勘定了其研究边界。作为一个以经验为基础的从感性到理性的抽象过程,这是学科迈向科学的起点。

核心术语的形成是一个主客观辩证统一的过程。术语将人类活动关系和事件联结到一个具有各种联系的世界里。术语与生活实践紧密相关。术语学属于人文社会科学和自然科学的交叉学科,并通过人类活动而不断完善,这就更加突出了其人文社会科学属性。人类以经验为基础对科学现象和事物进行体系化、术语化、概念化和范畴化,进而抽象形成关于事物和现象的术语。按照本体论的时序关系,学者将核心术语建构起来,从其发生和发展的角度来建立核心术语体系,是学科科学化发展的重要出发点[1]。

任何学科的建立都离不开一套理论系统和可用于阐释相应理论的术语系统,它们是以少数核心术语为基础建立起来的。“术语学”,“术语”,“术语集”,“术语系统”,“术语编纂”,“术语标准化”等正是建立术语学学科必不可少的核心术语[2]。

做好核心术语集中每一个术语的定义,是术语工作的核心。对整个中国科技术语系统的定义起着重要的指导意义。

很多情况,由于各学科交叉,对集中反映中国科技术语的核心术语没有重点研究,因此出现了各学科自行定义,既缺乏术语系统定义问题的协调统一的研究,也缺少国际间不同语言术语的协调与交换。这是当前术语研究工作迫切需要解决的问题之一。

在中国科技术语的研究和审定体系中,核心术语的研究与实践应用,对于术语学的研究与审定,建立中国科技术语与世界科技术语相关联的术语研究架构,都是重要的研究方向。而核心术语集的研究就具有更重要的意义。

对于一个专业领域的术语进行解析,可以得到这个专业的词根构成的术语集合。根据词频分布规律确定的核心术语集合对术语工作的协调和交流具有重要意义。

2. 核心术语的定义

2.1. 术语的定义与释义

科学技术语言中最重要的是基本术语的定义。没有准确的术语定义,定理、推理和结论一系列的证

明无法展开。《证明与反驳》这本书译者的话指出([3], 序 2): “冲突可以促进数学生长, 总有点费解。作者给你细讲了几条线索中的一条: 证明与反驳互为触发剂, 协同作用于数学知识的革新。”“新概念(新方法)的形成是数学史上的里程碑。”。

在定义释义中出现的术语依然需要定义, 这是无穷的递归, 以至于某个环节出现循环定义而成为两个定义的循环映像[4]。

为了解决术语概念的定义递归的终结, 先哲和前贤作了以下论述: 亚里士多德是世界古代史上伟大的哲学家、科学家和教育家。亚里士多德指出: “定义是一个有关本质不可证明的陈述” ([3], p. 129)。帕斯卡(Pascal, B.)是伟大的哲学家、数学家、物理学家。为了纪念它对人类的伟大贡献, 物理学以他的名字命名了国际单位制中表示压强的基本单位帕斯卡, 简称帕, 符号 Pa。计算机科学命名了语法科学严谨, 层次分明, 程序易写, 可读性强, 第一个结构化编程语言为 Pascal 语言。帕斯卡的定义规则([1659], 第 596~597 页): “凡是一目了然的现成术语。不要下定义。凡是有一丝一毫模糊性或多义性的术语, 不准不下定义。在术语的定义里。只使用一目了然或做过解释的词” ([3], p. 128)。

2.2. 定义的原则

在全国科学技术名词审定委员会科学技术名词审定原则及方法(修订稿)中指出[4]:

定义是对于一个概念的确切而简要的表述。科技名词审定工作中主要采用内涵定义, 必要时也采用外延定义等方式。

内涵定义: 说明某一概念的上位概念(即属概念), 确定该概念在其概念体系中的位置, 指出该概念的本质特征, 使其同其他相关概念区别开来。

外延定义: 用一系列众所周知、可穷尽列举的下位概念来说明某一概念。

名词委的定义叙述是科学而严谨的, 众所周知也就是一目了然, 在我国普遍实行中学教育普及的条件下, 可以认为, 在递归循环退出的条件是, 对于小学生所掌握的词语, 在专业术语词典中无须定义或者只需要简要定义。

名词委给出了定义的四项基本原则: 本质性(反映的客体所特有具有区别开来的特征)、科学性(明晰、准确、客观、符合逻辑)、系统性(概念体系中与上位概念及同位概念间的关系)、简明性(只需描述概念的本质特征或一个概念的外延, 不需要说明性与知识性的解说)。

ISO 标准是指由国际标准化组织(International Standard Organization, ISO)制定的标准。关于科技术语系统的一系列定义, 中国国家标准予以等价采用。涉及的主要术语定义如下:

定义【definition】: 描述一个概念, 并区别于其他相关概念的表述。

术语【term】在特定专业领域中一般概念的词语指称。

专业领域【subject field】专业知识领域。

概念【concept】通过对特征的独特组合而形成的知识单元。

内涵定义【intensional definition】用上位概念和区别特征描述概念内涵的定义。

外延定义【extensional definition】列举根据同一准则划分出的全部下位概念来描述一个概念的定义。

上位概念【Superordinate concept】广义概念。表示属概念或整体概念的概念。

下位概念【subordinate concept】狭义概念。表示种概念或部分概念的概念。以上引自 GB/T 15237.1-2000 术语工作词汇, eqv ISO 1087-1:2000

2.3. 核心术语[Core terms]的释义与定义

2.3.1. “核心术语”术语“核心”的释义

本文作者定义了核心术语[Core terms]

核心术语的定义与解释的关键在于中国汉字所独有的汉字本体特征, 在于汉字是客观世界实体图形描述的高度抽象。核心术语的关键解读在于“核”字。“核”是汉字典型的形声字, “核”字的表义部首分类属于“木”, “核心术语”的“核”字与木的关系是什么?

桃树、杏树、枣树等所结的果实之核, 即桃核、杏核和枣核。

按新华字典的解释是: 果实中坚硬并包含果仁的部分;

核: <名>(形声。从木, 亥声。本义: 果核);

心: <名>(象形。据甲骨文和小篆, 中间像心;外面像心的包络。本义: 心脏);

核心: 核之心也, 谓之“仁”。植物代代遗传繁殖之要也。亦有杏仁、桃仁、枣仁之称。食品则有杏仁酥等。核之心对于植物的代代繁殖最为重要。

2.3.2. “核心术语”的定义

核心[core]: 核之心也。

核心术语[core term]: 起到中心作用, 构词性能强的术语。

核心术语集合[core term set]: 核心术语的集合, 简称核心术语集。

3. 核心术语集合计算模型的理论基础

3.1. 齐普夫定律

关于单词在文献中出现频次的齐普夫定律(Zipf's Law)。是 1935 年由美国哈佛大学语言学教授 George K. Zipf 对英语文献中单词出现的频次进行大量统计以检验前人的定量化公式而提出的一个数学公式。也是文献学三大定律之一。1948 年, 时年 46 岁的齐普夫完成了他的专著《人类行为与最省力法则——人类生态学引论》, 1949 年首次出版。“最省力法则”较好地解释了齐普夫定律的内在成因和机制, 是齐普夫定律的理论基础。由于他的贡献, 人们称单参数词频分布定律为齐普夫定律(Zipf's Law)。

中国著名语言学家冯志伟在《数理语言学》中对词频分布的三个公式进行了详细讲解叙述[5]。齐普夫根据中篇小说《尤利西斯》所编写的频率词典得到其函数表达式。在这本频率词典中, 总计使用了 29899 个单词, 一个单词只出现一次, 词次为 1。总计使用 26 万多词次(也称为文本容量 T, 即条数)。按每个单词使用次数的多少进行排序, 可以获得每个单词的序号(正整数)以及该序号单词在书中出现的次数, 即该序号单词出现的频次, 也称为频率。

这样, 以单词序号(例如: n)为该单词横坐标上的点(x), 以该单词的频次为纵坐标点(y)则在直角坐标系下这些点的连接拟合后, 就构成了一条曲线。描述这条曲线的方程, 整个曲线在 1 到不同单词总数 W 区间曲线下的面积, 就构成文本容量 T , 也就是 $1 \rightarrow T$ 之间曲线的积分, 也就构成了词频分布方程中的总词次。

词频曲线方程的参数求解, 在数学上多取对数转换为直线方程, 采用最小二乘法进行最佳拟合求解。从而获得曲线方程的系数数据, 以反映词频分布变化的规律。

齐普夫定律的词频数学表达式如下所示:

$$f = cr^{-1}$$

公式中: f 词频数; r 单词序号; c 系数。

当试验次数增多或样本容量相当大时, 频率 f 转变为概率 P 。就得出如下公式, 即单参数词频概率分布公式, 即齐普夫定律的数学表达式:

$$P_r = cr^{-1}$$

公式中: P_r 为单词出现的概率; c 为系数。

1936 年, 美国语言学家 M. Joos 对 zipf 的公式, 进行修正得出双参数词频分布定律:

$$P_r = cr^{-b}$$

公式中： b 、 c 为系数。

英籍法国数学家 B. Mandelbrot 用概率论和信息论方法通过严格的数学推导，从理论上提出了词的三参数词频分布规律：

$$P_r = c(r+a)^{-b}$$

公式中： a 、 b 、 c 为系数。

词频计算所求出公式的参数，近似反映了词频分布的整体概貌，词频的统计与概率数据，是离散的数据而不是连续的函数，所以具体的计算应根据离散数据求和的方法，而不能采取连续函数的积分运算。

这些公式都是数学的**连续函数**，实际上单词排序的级(序号)，是从 1 开始的**正整数**。在计算核心术语集合的界点时，使用**离散函数**的求和公式来计算。

齐普夫定律已经在语言学、情报学、地理学、经济学、信息科学等领域有了广泛的应用，而且取得了不少可喜成果。中国数学家和语言学家周海中曾经指出：齐普夫定律是描述词频分布规律的强大数学工具，作为经验定律，它仍有不足之处，有待进一步完善。

冯志伟指出：“关于词的频率分布问题是比较复杂的。”“公式本身的性质决定了文本中不能存在频率相同的词，这与语言的客观事实是不符合的。频率的雷同数是随着序号的降低而减少的，越是序号高的单词，频率相同的越多。可见，词的频率分布规律还有必要进一步加以研究” [5]。

齐普夫定律是对词频分布的近似描述，使用了**连续函数**，可以计算并求出特征数据。对于词频分布的规律，反映的是**正整数的离散数据**，**词序**是整数，**频次**(注：有些文章使用频率)也是整数，因此对齐普夫定律在计算上采用离散函数计算的方法计算，**用数列的求和代替积分**，获取不同单词序号的**词集合数**与**频次总和**之间的关系。

这些公式都是数学的**连续函数**，实际上单词排序的级(序号)，是从 1 开始的**正整数**。应该使用**离散函数**的求和公式来计算。

3.2. 帕累托定律(长尾定律, 不平衡定律)

帕累托分布(Pareto distribution)是以是 19 世纪末 20 世纪初意大利经济学家维弗雷多·帕累托命名的，是从大量真实世界的现象中发现的幂次定律分布，这个分布在经济学以外，也被称为布拉德福分布。帕累托因对意大利全国 20% 的人口拥有 80% 的财产而 80% 的人口只有 20% 的财产的统计观察而著名，后来被约瑟夫·朱兰和其他人概括为**帕累托法则**又称为**二八定律**或**80/20 法则**，后来进一步概括为**帕累托分布**的概念。帕累托法则(Pareto's principle)也叫帕累托特定律、朱伦法则(Juran's Principle)、关键少数法则(Vital FeRule)、不重要多数法则(Trivial Many Rule)最省力的法则、不平衡原则等，被广泛应用于社会学及企业管理学等。帕累托分布具有较长的尾部，也称为**长尾分布**。

帕累托分布在许多领域有着大量的应用。例如：学习一门外语，应该先掌握哪些单词最省力？科学普及文献应该选用哪些科技词汇？小学语文教材识字如何优选常用字的学习序列问题？中文部首的学习教育如何选择顺序？这些，都需要数学模型。

3.3. 幂律分布

齐普夫分布和**帕累托分布**都是**幂律分布**，具有长尾的特点。

幂律分布是自然界与社会生活中存在各种各样性质迥异的而具有相同规律的现象，因而对它们的研究具有广泛而深远的意义。借助于有效的物理和数学工具以及强大的计算机运算能力，科学家们对幂律分布的本质有了进一步深层次的理解。

不平衡现象是客观世界的规律。银河系星球的大小和数量不平衡,地球上植物的大小和高度不平衡,不同人的智力发育不平衡等等。分析不平衡现象,正确用于科技术语的研究审定、定义和发布,可以加强术语工作的科学性,提高工作效率。

常用汉字是汉字总数的一小部分,为广大的人民群众社会生活使用,非常用汉字多用于各行各业的科技术语。因此,研究科技文献术语的汉字分布对术语定义也具有重要作用。

在长尾分布的水平方向上。越向右其**专业深度越强,普及性越差**。越向左其**普及性通用性越强,而专业深度性越弱**。在中国的汉字体系中,所谓的“死字”正是专业性极其深邃的科学技术用字。使用人数虽然极其稀少,对科学研究却极其重要。

术语也是这样,例如:数学术语:子环,子域,同构。化学专业的术语:二棕榈酰磷脂酰胆碱,孕酮。气象术语:旗云,冷涡等。

4. 核心术语集合的数学模型计算公式

4.1. 计算过程

核心术语集合的过程如下:

1) 首先获取专业术语的全部词表,进行标准化的预处理(半角到全角字符的转换,夹杂标点符号与英文字母的处理),而后进行分词与词性标注。

2) 对分词后的术语集合汇总统计,提取术语部件,即术语基本词汇,对基本术语频次进行排序。获得术语词频表。

3) 在术语词频表的基础上,使用计算机进行计算,获取核心术语与非核心术语的分界点。从而区分核心术语集和与非核心术语集。由计算机输出核心术语集。

4.2. 计算公式

核心术语的数量和分布主要是计算核心术语的数量,确定核心术语集合的范围,核查审定最基本的、需要重点优先定义的术语核心词汇。

计算方法如下:

计算各词序号的累计频次数与总词次百分比。命名为 A

计算各词序号与总序号(词总数)的百分比。命名为 B。

当 A + B 最接近于 100% 的点的词序号。就是**分界点**,简称**界点**。

界点的图示,参见图 1。

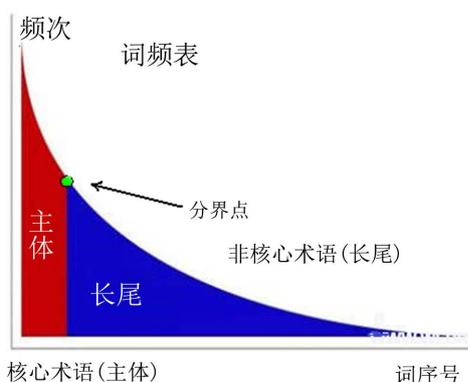


Figure 1. Diagram of dividing points of long tail model

图 1. 长尾模型分界点示意图

我们编写了计算软件, 使用核心术语集和的软件工具可以方便地计算出核心术语, 核心汉字, 核心部首, 以及各国语言中的核心单词等。

全部计算过程见表 1, Unicode10.0 汉字数(按部首分类统计): 87,849

部首采用康熙字典部首, 汉字编码由国际标准化组织和 UNICODE 决定。

Table 1. Classification and statistics of 87,849 Chinese characters (by radical)

表 1. 87,849 个汉字(按部首)分类统计表

序号	部首字	部首号	汉字数量	占总数百分比	字数累计百分比(A)	A + B 的和	序号累计百分比(B)
1	艸	140	3641	4.1446%	4.1446%	4.6119%	0.4672%
2	口	30	3274	3.7268%	7.8714%	8.8060%	0.9345%
3	水	85	3206	3.6494%	11.520%	12.922%	1.4018%
4	木	75	3087	3.5139%	15.034%	16.904%	1.8691%
5	手	64	2481	2.8241%	17.859%	20.195%	2.3364%
6	金	167	2344	2.6682%	20.527%	23.331%	2.8037%
7	心	61	2219	2.5259%	23.053%	26.324%	3.2710%
8	人	9	1892	2.1536%	25.206%	28.945%	3.7383%
9	火	86	1850	2.1058%	27.312%	31.518%	4.2056%
10	虫	142	1741	1.9818%	29.294%	33.967%	4.6728%
11	糸	120	1735	1.9749%	31.269%	36.409%	5.1401%
12	竹	118	1681	1.9135%	33.183%	38.790%	5.6074%
13	土	32	1679	1.9112%	35.094%	41.169%	6.0747%
14	女	38	1616	1.8395%	36.933%	43.475%	6.5420%
15	言	149	1607	1.8292%	38.763%	45.772%	7.0093%
16	鳥	196	1508	1.7165%	40.479%	47.956%	7.4766%
17	山	46	1434	1.6323%	42.112%	50.055%	7.9439%
18	魚	195	1375	1.5651%	43.677%	52.088%	8.4112%
19	肉	130	1343	1.5287%	45.205%	54.084%	8.8785%
20	玉	96	1236	1.4069%	46.612%	55.958%	9.3457%
21	日	72	1194	1.3591%	47.972%	57.785%	9.8130%
22	目	109	1167	1.3284%	49.300%	59.580%	10.280%
23	足	157	1159	1.3193%	50.619%	61.367%	10.747%
24	石	112	1107	1.2601%	51.879%	63.094%	11.214%
25	辵	162	1054	1.1997%	53.079%	64.761%	11.682%
26	衣	145	1031	1.1736%	54.253%	66.402%	12.149%
27	疒	104	964	1.0973%	55.350%	67.967%	12.616%

Continued

28	禾	115	906	1.0313%	56.381%	69.466%	13.084%
29	馬	187	904	1.0290%	57.411%	70.962%	13.551%
30	犬	94	874	0.9948%	58.405%	72.424%	14.018%
31	食	184	816	0.9288%	59.334%	73.820%	14.485%
32	車	159	780	0.8878%	60.222%	75.175%	14.953%
33	刀	18	779	0.8867%	61.109%	76.529%	15.420%
34	宀	40	763	0.8685%	61.977%	77.865%	15.887%
35	邑	163	751	0.8548%	62.832%	79.187%	16.355%
36	米	119	709	0.8070%	63.639%	80.462%	16.822%
37	阜	170	708	0.8059%	64.445%	81.735%	17.289%
38	貝	154	700	0.7968%	65.242%	82.999%	17.757%
39	頁	181	692	0.7877%	66.030%	84.254%	18.224%
40	門	169	667	0.7592%	66.789%	85.481%	18.691%
41	示	113	656	0.7467%	67.536%	86.695%	19.158%
42	广	53	631	0.7182%	68.254%	87.880%	19.626%
43	雨	173	625	0.7114%	68.966%	89.059%	20.093%
44	支	66	624	0.7103%	69.676%	90.237%	20.560%
45	巾	50	566	0.6442%	70.320%	91.348%	21.028%
46	酉	164	556	0.6329%	70.953%	92.448%	21.495%
47	穴	116	528	0.6010%	71.554%	93.517%	21.962%
48	革	177	512	0.5828%	72.137%	94.567%	22.429%
49	田	102	498	0.5668%	72.704%	95.601%	22.897%
50	走	156	491	0.5589%	73.263%	96.627%	23.364%
51	牛	93	461	0.5247%	73.787%	97.619%	23.831%
52	彳	60	449	0.5111%	74.299%	98.598%	24.299%
53	毛	82	446	0.5076%	74.806%	99.573%	24.766%
54 界点★	彡	190	434	0.4940%	★75.300%	★100.5%	★25.233%
55	大	37	428	0.4871%	75.787%	101.48%	25.700%
56	耳	128	396	0.4507%	76.238%	102.40%	26.168%
57	尸	44	391	0.4450%	76.683%	103.31%	26.635%
		中间	58-	211 行数据	(略)		
212	艮	138	15	0.0170%	99.973%	199.03%	99.065%
213	辶	204	13	0.0147%	99.988%	199.52%	99.532%
214	玄	95	10	0.0113%	100%	200%	100%

核心部首汉字数据集的确定, 对我国对外汉语教学有着重要的指导作用, 在汉语汉字教学改革的体系中, 推进汉语汉字教学的科学化、规范化的系统设计和教材全面改革方面具有重要的意义。表 1 界点的位置表明汉字总数的 75% 使用的是 25% 的部首, 25% 的汉字使用的是 75% 的部首。因此占总数 25% 的 54 个部首汉字是首先重点学习的对象。

5. 应用计算实例

5.1. 计算实例一

5.1.1. 中国科技术语动物学专业核心术语的计算

界点位置 2509-误差: 0.0116。

长尾定律分界点值: 40.0/60.0 界点: 2508。

中国科技术语动物学专业核心术语词数: 2508。

5.1.2. 中国科技术语电学专业核心术语的计算

界点位置 1309-误差: 0.0148。

长尾定律分界点值: 26.9/73.1 界点: 1308。

中国科技术语电学专业核心术语词数: 1308。

两个专业比较, 可以得出, 电工专业核心术语较为集中, 构词能力强。经济指数较高。

表 2 所示是中国科技术语部分专业核心术语集合的统计。

Table 2. Statistics of core terms and components of scientific and technological terms

表 2. 科技术语部分专业核心术语部件统计表

专业	术语总数	术语部件总数	核心术语部件数	界点值	经济指数
电工学	7296	4871	1308	27/73	1.50
动物学	6699	6276	2508	40/60	1.07
天文学	2275	2172	789	36/64	1.04
电子学	4725	3826	1165	30/70	1.23
计算机	14,286	6272	1510	24/76	2.27
化学	5861	4868	1498	31/69	1.20
数学	8856	5359	1354	26/74	1.65
物理学	8259	6078	1795	30/70	1.36
自动化	1898	1290	352	27/73	1.47
农学	3070	2961	1162	38/62	1.03
化学工程	2161	2003	725	36/64	1.07

注: 各专业术语数依据名词委曾经发布的光盘版统计资料。

5.2. 计算实例二

中国科技术语 120 个最重要的核心术语。

2014 年杨福义从 20 万科技术语数据库中的抽取获得 10 万条的术语部件[6], 经过计算, 其核心术语有 22,705 个, 约占术语总数的 10%, 排序得到中国科技术语最核心的 120 个词汇如表 3 所示(按频数依次排列)。

Table 3. Core glossary of Chinese scientific and technological terms (top 120)
表 3. 中国科技术语核心词汇表(前 120 个)

系统	试验	细胞	控制	资源	系数	结构	作用	模型	装置
分析	生物	效应	测量	材料	时间	蛋白	数据	函数	土壤
空间	环境	反应	设计	曲线	因子	基因	方程	理论	信息
化学	技术	管理	植物	神经	电流	信号	强度	设备	复合
综合征	电子	纤维	自动	温度	保护	工程	免疫	处理	电压
生态	组织	卫星	通信	动脉	分布	指数	程序	数字	辐射
合金	运动	过程	静脉	激光	速度	自然	城市	安全	密度
网络	动物	中心	机械	压力	应力	循环	质量	地图	气候
状态	离子	海洋	计算机	地理	条件	连续	表面	频率	脉冲
传感器	电路	故障	轨道	测试	特性	DNA	动态	业务	振动
交换	方法	雷达	模式	混凝土	语言	平均	光学	运输	金属
序列	土地	噪声	规划	效率	地球	功能	操作	图像	分子

6. 结语与讨论

我们认为：核心术语集合的计算依据理论基础扎实，可以在不同的专业词频表上予以验证和计算使用，是术语审定与科研工作的重要工具。建议开展以下工作。

6.1. 出版各种语言相对应的核心术语词典

在术语定义中，需要参考 ISO 国际标准和国家标准，对基本术语做出严谨科学的定义。在此基础上作出复合科技术语的定义。已经发布的术语定义，有必要进行清理。以保证定义的科学性与逻辑性。例如：“计算机断层扫描”术语，进行结构分析，可以得到三个术语，计算机、断层、扫描。也只能在这三个术语的定义与解释的基础上，定义“计算机断层扫描”。当然通俗的释义是使用计算机用断层的方式进行的扫描。说明扫描使用的工具与方法。

根据我国多民族术语使用的需要，可以编写中国各民族核心术语翻译词典，解决各民族科技文献急需的术语翻译问题，并为国际化的各国语言核心术语的翻译工具库的建设作出探索性贡献。以利于世界各国的术语查询互操作。

由于汉字反映的音义结合体，汉字构词反映了意合的重要属性，中国研究的各科核心术语集合可以翻译成为英语、俄语、日语等世界各国语言对应的核心术语集和，有利于科技术语核心术语的计算方法推广试用，以促进世界术语科研的交流。

6.2. 开展关于核心术语云模型的研究

核心术语与其相关联的术语，构成了**术语云团**，例如：以“系统”相关联的术语有 3000 多个，是科技术语系统中，最大的**术语云团**。结构大小不同的**术语云团**及其相互的关联则构成**科技术语云**的系统模型。

按使用频率排序后的中国科技术语的前 6000 条术语部件占整个术语系统的重要地位，这些术语及其关联，则构成整个科技术语系统云结构的基本骨架，从而把科学化、系统化和结构化的理念深入地运用到术语科学研究中。

裴亚军指出“科技术语云力图建设一个面向服务的体系结构, 为一切抽象、自治和无状态的用户需求建立一系列易于理解并执行的组织原则和复杂的系统模型” [7]。任重道远, 需要吾辈努力践行之。

6.3. 开展多学科交叉协作研究核心术语集之间相关关系

核心术语集之间的相关关系, 构成语义网, 需要建设术语知识库的链接和查询检索系统以进行属性的深入研究。这项工程, 不仅需要语言学家, 也需要数理逻辑学家的参与, 使用类似于几何证明与数学推理一样的严谨论述。只有这样, 才能使中国科技术语的反映概念本体映像的汉字本体, 进入国际术语研究的前列。

致 谢

本文得到冯志伟导师的理论指导和帮助, 在此表示感谢。

参考文献

- [1] 句云生. 加快发展中国特色术语学[N]. 中国社会科学报, 2020-08-11(1986).
- [2] 叶其松. 术语学核心术语研究[D]: [博士学位论文]. 哈尔滨: 黑龙江大学, 2010.
- [3] (英)伊姆雷·拉卡托斯. 证明与反驳-数学发现的逻辑[M]. 康宏奎, 译. 上海: 上海译文出版社, 1987.
- [4] 全国科学技术名词审定委员会. 科学技术名词审定原则及方法[EB/OL]. <http://www.cnctst.cn/sdgb/sdyzjff/index.html>, 2021-11-14.
- [5] 冯志伟. 数理语言学[M]. 北京: 商务印书馆, 2012: 270-271.
- [6] 杨福义. 大型术语部件库建设的实践与思考[J]. 中国科技术语, 2014(4): 9-13.
- [7] 裴亚军. 科技术语云搭建 21 世纪科技知识管理平台[N/OL]. 中国社会科学报, 2012. http://blog.sina.com.cn/s/blog_48e1bd7301010q7t.html, 2021-11-14.