

หน่วยคำความถี่สูงในภาษาจีนปัจจุบันและปัจจัยที่ส่งเสริมให้ เกิดมีความถี่สูง

现代汉语高频语素及其形成原因分析

The study on High Frequency Morpheme in Modern Chinese

安美娜¹

บทคัดย่อ: งานวิจัยได้สร้างฐานข้อมูลหน่วยคำในภาษาจีนปัจจุบันขึ้น ซึ่งหน่วยคำเหล่านี้ได้มาจากพจนานุกรมจีนปัจจุบัน ฉบับปรับปรุง ค.ศ. 2007 และคำนวณหาหน่วยคำที่มีความถี่สูงในการนำมาสร้างคำประสมจำนวน 300 หน่วย เพื่อใช้เป็นหน่วยคำเป้าหมายในการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาค่าความถี่ในการนำมาสร้างคำประสม ลักษณะทางไวยากรณ์ และลักษณะทางความหมายของหน่วยคำเป้าหมาย นอกจากนั้นยังวิเคราะห์ปัจจัยที่ส่งเสริมให้เกิดปรากฏการณ์ทางภาษาเหล่านี้ จากการวิจัยพบว่า หน่วยคำที่มีต้นกำเนิดเก่าแก่ มีหลากหลายความหมายขอบข่ายความหมายกว้าง และลักษณะทางไวยากรณ์ที่หลากหลาย ล้วนเป็นปัจจัยหลักที่ส่งเสริมให้หน่วยคำมีความถี่สูงในการนำมาสร้างคำประสมในภาษาจีนปัจจุบันทั้งสิ้น

คำสำคัญ: ภาษาจีนปัจจุบัน หน่วยคำความถี่สูง การนำมาใช้สร้างคำ หน่วยความหมายของหน่วยคำ

¹ อาจารย์อรอนงค์ อินสอาด สาขาวิชาภาษาจีน คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น และนักศึกษาปริญญาเอก

สาขาวิชาภาษาและอักษรจีน สถาบันวิจัยภาษาและอักษรจีน มหาวิทยาลัยซีหนาน สาธารณรัฐประชาชนจีน

安美娜, 泰国孔敬大学人文和社会科学学院中文系汉语教师泰国留学生; 西南大学文献所博士生。

On-anong Insa-art, Lecturer, Chinese Section, Faculty of Humanities and Social Science, Khonkaen University,

Thailand; Ph.D. candidate, The Institute of Chinese Language Documents, Southwest University, P.R. China.

摘要: 文章以《现代汉语词典》(2007年修订版)作为选取语料的范围,建立一个现代汉语语素库。在现代汉语语素库的基础上,统计出最高频的300个语素作为研究对象。着重考察其在构词能力、语法属性、语义分布的情况,并分析与描写形成各种语言现象的原因。分析表明现代汉语高频语素的生命力长久、义项数量多、义域以及语法属性的宽泛都是其形成的主要因素。

关键词: 现代汉语; 高频语素; 构词能力; 素位

Abstract: The research takes a sample of 300 high frequency morphemes from *Modern Chinese Dictionary (2007's Edition)* as its research basis. Based on quantitative study, the paper discusses the word-formation ability, Semantic Morphemes and grammatical category of high frequency morpheme in modern Chinese. The range of the reference meaning; the number of semantic morphemes; the grammatical category are the factors that influence morphemes to take a high frequency in modern Chinese.

Keywords: modern Chinese; high frequency morpheme; word-forming ability; semantic morphemes

现代汉语复合词是现代汉语词汇系统的主体,复合词的结构单位是语素。相对于现代汉语庞大的词汇系统而言,语素在数量上是有限的,应该是可以封闭调查的。但是现代汉语的语素究竟有多少,至今没有进行科学的统计。在现代汉语语素中,各语素的构词能力是很不相同的。那些活动能力最强,最活跃,在复合词中复现频率很高的语素,我们称之为“高频语素”。怎样确定高频语素?现代汉语高频语素究竟有多少?语素生成新词能力的强弱究竟受哪些因素的影响?这些问题都值得研究。

以往有学者曾提出以高频语素为核心的现代汉语词语系统研究的基本思路和主要的方法。但没有具体实施。现在还没有学者对现代汉语高频语素进行过全面统计、分析与描写。本文就将这项工作作为我们研究的起点。

一、题目的学术背景

从 20 世纪 50 年代起,不同领域,用于不同目标的各种常用词的词表都陆续研制出来了。据初步统计,已经公布与正在研制并已初步成型的共有 20 余种词表。在已经公布的词表中,最有影响的两种表格即北京语言学院语言教学研究所的《现代汉语频率词典》(1986) 中的 8548 条常用词表与《汉语水平词汇大纲》提出的 8822 条常用词。

在汉语词汇定量研究方面值得一提的是尹斌庸(1984)的《汉语语素的定量研究》对汉语中4871个单音节语素进行了定量分析,统计了语素的词性、独立性和结合性。结果显示,名、形、动三类语素整体的构词能力以及各自的构词能力分布都近似概率论。苑春法、黄昌宁(1998)通过计算机对语素进行了大规模的描写,从宏观上分析了基本语素的素类分布与复合词词类分布的关系、汉语语素成词的位置特点,从微观上描写了名、动、形三类词的构词规律,并注意到语素在构成复合词时所发生的意义转化。苏新春(2002)致力于《现代汉语词典》词汇计量研究,侧重于其所收词目的多角度分类统计分析,建立了“《现代汉语词典》数据库”,把这部有相当权威性、规范性的语文词典作为现代汉语词汇总貌的一个缩影进行计量分析,已完成了10多个专题的研究,对词汇学和词典学上的一系列问题进行了定量的分析研究。符淮青的《语素“红”的结合能力分析》(1983)和《“打”义分析》(1985)亦是较早从语素的具体义项分析语素构词的成功典范。因此,有很多相关应用或计量研究都以这个分法为参照。

胡华(2007)提出以高频语素为核心的现代汉语词语系统研究的基本思路 and 主要的方法,就是以具有相同高频语素的词语群为具体对象的研究和方法,但没有具体实施

语素及其构词的计量研究成果举不胜举。研究者从不同的视角、运用不同的研究方法、采用不同的语料,呈现了丰硕的研究成果。

二、研究的路径

第一道工序是语料的处理。将《现汉》(2007年修订版)中的复合词输入电脑,排除《现汉》中标注有〈方〉的方言词。词缀能产性强,但词汇意义虚化,位置固定,构词不自由,也予排除。例如“儿”,有两种情况:有实在意义,有意义区别作用的是自由语素和半自由语素;意义虚化,没有意义区别作用的是不自由语素。所以,在我们的语素库中,没有意义区

别作用的不自由语素“儿”一律不收录。离析异音同形语素和同音同形语素,把一个字表示的不同语素进行标注区分。异音同形语素,如:jiao3音的“角”和 jue2 音的“角”,se4 音的“色”和 shai3 音的“色”。同音同形语素,如:名词性语素“上¹”和动词性语素“上²”。在完成了上述语料的处理后,我们建立了现代汉语复合词库。然后,在现代汉语复合词库的基础上,使用 Excel 离析出构词语素,最后得到现代汉语语素库。

第二道工序。在现代汉语语素库的基础上,使用 Excel 进行检索构词量在 63 个以上的词,最后得到 300 个语素,就是本研究所谓的“高频语素”,也就是本文的研究对象。

第三道工序。将这些高频语素分别从构词能力、语法属性、语义分布进行数据统计,并且根据数据统计结果,考察形成各种语言现象的原因。

三、现代汉语高频语素的认定

“频率 (frequency)”一词,也常称为频度,统计学中一般给“频率”下这样的定义:单位时间内某种事件发生的次数。词的频率通常指某个词在报刊、杂志、教材等书面材料中出现的次数。本文所谓复现率,是指语素在《现汉》(2007 年修订版)中重复出现的次数。本文将复合词的构成单位(词根语素)在《现汉》(2007 年修订版)的复现频率,也就是说构词能力作为确定高频语素的标准。那么,本文所谓的高频语素,就是构成复合词的能力强、复现频率很高、使用频繁的那些词根语素。

四、现代汉语高频语素的构词能力情况分析

本研究按构词量的多少将高频语素进行分析,以期发现构词能力强弱不同的高频语素的特点。

1. 构词量多于200个的高频语素有16个,按构词量多少排列如下:人、大、头、心、水、气、手、不、小、风、生、口、天、火、学、体。

从语义上看,它们指称的概念都是具体事物,也是人类语言中的核心概念,包括:人体器官、自然现象的名称。许慎(2007)的《说文解字》提到“近取诸身,远取诸物”¹,从此源头上看,近取自身、身体的名词、身动的动词,这是原始语言的基础。可见,这几类词都是远古基本词汇的主

¹许慎,《说文解字》[M],北京:中华书局,2007年版,第15上(序)

要部分，是几千年沿用下来的基本词汇。可以看出，它们生命力长久、所表示的意义原始，基本意义稳定的基本词汇。它们在古汉语中曾经是长期独立使用过，但到了现代汉语，有些变成了构词成分，不再独立使用了；有的在一定的语境中既独立使用，又作为构词语素参与构词。这是原始语言的基础。

2. 构词量在100~199的高频语素有122个，按构词量多少排列如下：

车、发、文、电、外、门、出、工、家、花、动、道、行、事、情、中、法、开、书、眼、线、合、光、名、国、自、流、面、一、地、打、点、主、色、年、分、音、正、商、时、性、金、意、平、通、公、场、物、语、路、定、会、词、油、海、战、热、制、军、病、报、交、笔、明、回、理、马、日、数、教、用、清、实、话、无、品、声、身、机、成、内、红、后、星、作、石、空、神、房、落、上、单、失、保、兵、民、字、重、器、黄、盘、权、山、期、白、节、放、对、血、老、变、转、反、世、度、皮、包、板、过、言、来、收。

这些高频语素是现代汉语交际不能缺少的基本语素。从语义来看，它们包括方位和时令的名称、与手有关的动词、自然物、自然想象、颜色词等。复合词中有一些语素搭配范围很宽泛，并且与其它语素的语义关系表现得较为稳定，有的甚至带有一些类似词缀的特征，这就是在一些论著中被称为“类词缀”的构词单位。强段含有许多可作为类词缀的高频语素是现代汉语中能自由地创造新词的活跃成分。如：“半-”、“-家”、“反-”、“-热”、“无-”等。这些高频语素能构成一百条以上的系列词，形成了以高频语素为类特征的“词类”，还可以在线类推出许多新词，增强了这些高频语素的构词能力。

3. 构词量在100条以下的高频语素有162个，按构词量多少排列如下：

戏、工、解、下、传、台、量、有、业、见、立、死、菜、亲、位、原、月、片、价、号、本、起、标、火、毛、骨、同、说、前、药、负、女、力、化、客、服、黑、论、接、进、食、脚、画、毒、球、目、结、游、表、新、感、鱼、利、货、照、精、视、礼、入、压、嘴、安、告、市、剧、案²、美、集、长、养、领、受、夜、政、职、底、经、首、商、半、酒、木、铁、指、警、连、选、母、套、展、直、果、记、草、求、活、条、步、密、上²、图、杂、考、冷、下²、印、应、衣、错、轮、关、防、务、烟、存、补、彩、私、称、查、绝、本²、取、编、笑、走、质、春、暗、势、阴、毒、望、约、账、沉、纸、全、茶、退、古、形、官、

相、引、证、区、班、改、修、洋、根、便、犯、票、观、角、提、婚、间、守、网

中段高频语素都与人类的生活、生产、经济、科技、文化等活动密切相关,是现代汉语的普遍用字,要构建覆盖面广阔的词汇量。其次,随着新事物的发展,覆盖面广阔的中段高频语素被选中作为构成新词的概率很高。只要一定的社会环境发展起来,频繁使用到其中的某个高频语素,其构词量便会大大增加,例如:有关生产语素的涌出,如网、机、器、金、铁、油、市、商、账、价等语素。这些高频语素从多方面反映了当今时代的特征。当今时代在信息、工业、经济行业语的泛化,反映出经济建设已成为我国社会生活中的重要主题。

从上述分析,可看出超强段的高频语素仍保留汉语古代时期的基本词汇特征,在强段的高频语素及中段的高频语素有些是古代的词被推出了基本词汇系统,有些词进入了基本词汇系统,一个时代的基本词汇之所以会发生变化,确实是社会、文化等因素影响的结果。

五、现代汉语高频语素形成的原因分析

上面我们对复合词的构成语素进行了提取,并得到本文所研究的对象“高频语素”。结合提取结果,我们拟从以下几个方面来分析现代汉语高频语素所形成的原因。

1. 高频语素的多义性。凡高频语素,其包含的义位都很多。多数高频语素除具有本义之外,还有引申义、比喻义等多个义位。从语义学理论上讲,高频语素的每一个语素项都可能作为构词语素,去和别的语素构成不同的复合词,以表达新的概念。一般来说,语素项都具有一定的构词能力,只是每个语素项的构词能力有大小区别罢了。每个语素项的构词数累计到一起,就是一个语素的全部构词。从这个意义上说,某一个语素,若它的义项愈多,它的构词能力也就愈强。

从义项数的统计结果来看,超强段的现代汉语高频语素的义项数平均为 8.56 个、强段为 7.84 个、中段为 5.90 个(以《现汉》(2007 年修订版)所列的义项数为统计依据)。可见,义项数量越多,构词量也就越多。

2. 高频语素语法功能的兼类性。语素大多来自古代汉语中的单音词,经过了漫长的时间,通过意义引申使得义项增生,所有义项的语法功能并不完全相同,故语素存在兼类现象,就是说,语素的语法功能范围大。这些单音的多义语素,一身多兼,一些名词性语素同时也具有动词性、形容词性甚至副词性功能,一些动词性语素也会具有名词性和形容词性功能,

形成一语素多语法属性的显著特性,从而使得其构词模式呈现出多样化的特征。高频语素拥有相当数量的义位,其构成词的能力也很强,可以较自由地创造新词。

3. 高频语素义域宽广,灵活性强,活动范围大,又为复合词的合成提供了充足的条件。

高频语素的含义多,它的每一个义位都可能和别的语素相结合,构成一个新的复合词,往往就具有很强的构词能力。

4. 在类推心理的作用下,以一个高频的语素为词根,配以不同的语素,构成一族词,可称为同根词、同素词,或称同族词。由于同素词含有相同的构词材料,彼此在意义上存在或亲或疏的关系。这种同素词聚是汉语词汇具有系统性的表现形式之一。例如以“地”为语素,可以构成“地面、地底、地下、地洞、地窖、地堡、地坪、地势、地形、地温、地雷、地沟”等复合词,各词的意义自然是不同的,但同时又必然具有或隐或显的联系。

5. 现代汉语高频语素很多在单用时都是基本词汇中的成员。斯大林(1971)《马克思主义与语言学问题》一书中提出了基本词汇的问题,说“语言的词汇中的主要东西就是基本词汇,其中也包括成为它的核心的全部根词,基本词汇比语言的词汇窄小的多,可是它的生命长久的多,它在千百年的长时期中生存着,并且为构成新词提供基础。”¹着力研究汉语基本词汇问题的林焱先生认为,斯大林的理论中包含有“研究基本词汇的最重要的材料”,“问题只在于对斯大林这些天才的指示如何正确的理解和发挥”²后来人们把划分基本词汇和非基本词汇的标准归纳为三性:稳定性、全民性、能产性。基本词汇的三个特征不是平列的,而有主次之分。这要根据研究对象来确定。研究现代汉语词汇,恐怕应以能产性作为首选条件。那么我们研究高频语素,首要依凭的是词语的能产性特征。因此高频语素所具有的能产性决定了它们多数是基本词汇。其次,要考察基本词汇的全民性,看一个词在一个时代的使用是否广泛,共时、平面是基本着眼点,简捷的方法是词频统计。一般说来,一个词的出现频率高,运用便广泛,是全民性的重要表现。自然,在《现汉》出现频率高的语素大多也是基本词汇。最后,从词汇的稳定性特征确定现代汉语的基本词汇,高频语素的大

¹ 斯大林 1971《马克思主义与语言学问题》[M],北京:人民出版社。

² 林涛,《汉语基本词汇中的几个问题》,《中国语文》[J],1954,7,转引自毛远明 2000《左传词汇研究》[M],重庆:西南师范大学出版社,第7页。

多数是古代已有的词，如：兵、天、本、回、告、大、首、身、走等都寿命比较长久。总的来说，本文调查出的高频语素都具备了基本词汇的特征。因为不同历史时段的基本词是不太一致的，先秦两汉汉语的基本词和现代汉语的基本词肯定不完全一样。我们调查出的现代汉语高频语素可以说是现代汉语的基本词汇。

6. 本研究以现代汉语高频语素的基本义（常用义）为判断语素性类的依据。从统计结果来看，现代汉语高频语素名词性最多，占55.46%，动词性次之，占 31.68%，形容词性又次之，占 11.88%，其余，数词性和副词性分别占 0.66% 和 0.33%。可见，名词性语素有很强的或较强的构词能力；动词性语素的构词能力较强，其它性类的语素一般构词能力较弱，甚至不构词。这说明语素的构词能力和语法属性之间具有密切的关系。

7. 从认知的角度分析，随着现代人们对指称事物有了更深刻的认识，语言表达也要求精确性原则。这些概念极易被以各种标准分行分类，以分出不同的下为概念。例如：对事物性状本身进行细化。以色调语素义场为例，在现代汉语高频语素中，出现了白、红、黄、黑、青等五个作为高频语素，在有关色调语素前面加上限制性语素造出成批的新词。这标志着人们对色调的深浅、明暗、浓度、亮度等的细致区分，表明多色调词语的增多、细致，则是明显的发展，这是人们区分色彩能力提高在语言上的反映。

总之，现代汉语高频语素产生的时间短长、语义的宽窄、语法属性、人们的认知心理都是影响语素构词能力的重要因素，研究汉语词汇必须通盘考虑。

参考文献:

(一) 材料

中国社会科学院语言研究所词典编辑室。2007。《现代汉语词典》第 5 版（修订版）[K]。北京：商务印书馆。

(二) 专著

北京语言学院语言教学研究所。1986。《现代汉语频率词典》[K]。北京：北京语言学院出版社。

符淮青。2007。《词典学词汇学语义学文集》[M]。北京：商务印书馆出版社。

毛远明。2000。《左传词汇研究》[M]。重庆：西南师范大学出版社。

苏新春。2002。《汉语词汇计量研究》[M]。厦门：厦门大学出版社。

(三) 论著

胡 华。《以高频语素为核心的现代汉语词汇研究的思路和方法》[J]。长春师范学院学报(人文社会科学版)，2007年第1期。

苑春法、黄昌宁。《基于语素数据库的汉语语素及构词研究》。《世界汉语教学》[J]。1998 年第 2 期。

(บทแปล)

หน่วยคำความถี่สูงในภาษาจีนปัจจุบันและปัจจัยที่ส่งเสริมให้ เกิดมีความถี่สูง

อรอนงค์ อินสอาด¹

คำประสมเป็นส่วนประกอบหลักของระบบคำในภาษาจีนปัจจุบัน หน่วยที่นำมาสร้างคำคือหน่วยคำ(หน่วยที่เล็กกว่าคำหรือรากคำ) ถึงแม้ว่าระบบคำในภาษาจีนปัจจุบัน มีขนาดใหญ่ แต่ หน่วยคำ ยังคงมีจำนวนจำกัด จึงสามารถทำการวิจัยในลักษณะปิดได้ แต่ทว่า หน่วยคำในภาษาจีนปัจจุบันมีจำนวนเท่าไร ปัจจุบันยังไม่มีการคิดคำนวณอย่างเป็นระบบ และในบรรดาหน่วยคำเหล่านี้ ความถี่ในการนำมาใช้สร้างคำประสมของแต่ละหน่วยคำย่อมแตกต่างกัน งานวิจัยนี้จึงเรียกหน่วยคำที่มีความถี่สูงในการนำมาใช้สร้างคำประสมว่า “หน่วยคำความถี่สูง” การให้คำนิยามหน่วยคำความถี่สูง จำนวนหน่วยคำในภาษาจีนปัจจุบัน ค่าความถี่ของหน่วยคำที่นำมาใช้ในการสร้างคำประสม และปัจจัยที่ส่งเสริมให้เกิดความถี่ในการนำหน่วยคำมาใช้สร้างคำประสม เหล่านี้ล้วนเป็นเรื่องที่น่าสนใจอย่างยิ่ง

มีนักวิชาการได้กล่าวถึงแนวคิดและแนวทางพื้นฐานในการศึกษาวิจัยหน่วยคำความถี่สูง ซึ่งเป็นส่วนหลักในระบบคำของภาษาจีนปัจจุบัน แต่ไม่ได้ดำเนินการวิจัยอย่างเป็นรูปธรรม และในปัจจุบันยังไม่มีนักวิชาการท่านใด คิดคำนวณทางสถิติ และวิเคราะห์คุณลักษณะในด้านต่าง ๆ ของหน่วยคำความถี่สูงในภาษาจีนปัจจุบัน

1. ทบทวนวรรณกรรม

นับตั้งแต่ปี ค.ศ.1950 ได้มีการศึกษาวิจัยเพื่อสร้างตารางคำที่ใช้บ่อยตามวัตถุประสงค์ต่าง ๆ ขึ้นอย่างต่อเนื่อง ตามสถิติในเบื้องต้น พบว่ามีตารางคำที่ประกาศใช้และกำลังอยู่ระหว่างการวิจัยจำนวนกว่า 20 ตาราง ในบรรดาตารางคำที่นำมาประกาศใช้มี 2 ตารางที่ได้รับการยอมรับอย่างมาก นั่น คือ พจนานุกรมความถี่ของคำภาษาจีนปัจจุบัน (1986) ซึ่ง

¹ อาจารย์อรอนงค์ อินสอาด สาขาวิชาภาษาจีน คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น และนักศึกษาระดับปริญญาเอก สาขาวิชาภาษาและอักษรจีน สถาบันวิจัยภาษาและอักษรจีน มหาวิทยาลัยเซี่ยงไฮ้ สาธารณรัฐประชาชนจีน

รวบรวมนำคำที่ใช้บ่อยไว้ถึง 8,548 คำ และตารางคำเพื่อการวัดระดับภาษาจีน ที่รวบรวมคำศัพท์ที่ใช้บ่อยไว้ถึง 8,822 คำ

Yin Binyong (1984) ได้นำการวิจัยแบบสถิติมาใช้ในการวิจัยทางด้านคำศัพท์ โดยได้ศึกษาเรื่องการศึกษาหน่วยคำในภาษาจีนโดยวิธีทางสถิติ นับว่าเป็นงานวิจัยที่นำวิธีวิจัยแบบคำนวณค่าสถิติมาใช้ในการวิจัยภาษาศาสตร์เป็นครั้งแรกซึ่งได้เก็บค่าสถิติหน่วยคำพยางค์เดียวในภาษาจีนปัจจุบันจำนวน 4,871 หน่วย อีกทั้งยังวิเคราะห์ลักษณะทางไวยากรณ์ การนำมาใช้เป็นคำโดดและการนำมาใช้ในการประสมคำ ผลการวิจัยพบว่าหน่วยคำนาม คำกริยา และคำวิเศษณ์ ล้วนมีการสร้างคำใกล้เคียงกับอัตราที่คาดการณ์ไว้

Yuan Chunfa และ Huang Changning (1998) ได้นำระบบคอมพิวเตอร์มาใช้ในการทำฐานข้อมูลหน่วยคำ ความสัมพันธ์ระหว่างประเภทหน่วยคำกับประเภทคำ ตำแหน่งของหน่วยคำในการประสมคำ อีกทั้งยังวิเคราะห์กฎเกณฑ์การประสมคำของหน่วยคำนาม คำกริยา คำวิเศษณ์นอกจากนั้นยังมองเห็นถึงหน่วยคำเปลี่ยนความหมายเมื่อมาประสมกัน กลายเป็นคำประสมอีกความหมายหนึ่งได้อีกด้วย

Su Xinchun (2002) ได้สร้างฐานข้อมูลพจนานุกรมจีนสมัยใหม่ โดยรวบรวมและวิเคราะห์คำศัพท์ในพจนานุกรมไว้อย่างครบถ้วน และได้มีผลงานการวิจัยทางด้านคำศัพท์และพจนานุกรมออกเผยแพร่แล้วกว่า 10 เรื่อง

Fu Huaiqing (1983) ยังเขียนบทความเรื่อง การประสมคำของหน่วยคำ “红” (หง) และ บทความเรื่องการวิเคราะห์ความหมายของ “打” (ต้า) (1985) ซึ่งถือว่าเป็นการวิจัยต้นแบบที่ประสบความสำเร็จในการ วิเคราะห์ความหมายของหน่วยคำ และการนำ แต่ละความหมายของหน่วยคำมาใช้ประสมคำ นับจากนั้นเป็นต้นมามีงานวิจัยที่ใช้ระเบียบวิธีวิจัยคล้ายคลึงกันเผยแพร่ออกมาเป็นจำนวนมากโดยมีงานวิจัยข้างต้นเป็นต้นแบบ

ถึงแม้ว่า Hu Hua (2007) ได้กล่าวถึงแนวคิดและแนวทางการวิจัยระบบคำศัพท์ภาษาจีนปัจจุบันโดยนำหน่วยคำความถี่สูงมาเป็นประเด็นหลักในการวิจัยไว้ในบทความ แต่ยังไม่ได้มีการดำเนินการวิจัยอย่างเป็นระบบ

จากงานวิจัยและบทความวิชาการต่างๆ ดังกล่าว แสดงให้เห็นว่า ได้มีผลการวิจัยทางด้านหน่วยคำและการนำมาประสมคำของหน่วยคำมีเป็นจำนวนมาก เนื่องจากนักวิจัยได้ทำการวิจัยจากมุมมองระเบียบวิธีวิจัย กลุ่มเป้าหมายในการวิจัยที่ต่างกัน

2. ขั้นตอนการวิจัย

ขั้นตอนที่หนึ่ง การจัดการข้อมูลเป้าหมาย ป้อนคำประสมในพจนานุกรมจีนปัจจุบัน ฉบับปรับปรุง ค.ศ. 2007 ลงในฐานข้อมูลโดยไม่รวมถึงคำศัพท์ท้องถิ่น คำเติมหน้าและคำเติมท้ายที่มีความถี่ในการประสมคำสูง อย่างเช่น “儿” (เออร์) ซึ่งมี 2 ลักษณะคือ

1. “儿” (เออร์) ที่มีความหมายโดยเนื้อแท้ มีคุณสมบัติในการแยกแยะความหมายของคำคือ หน่วยคำอิสระและหน่วยคำกึ่งอิสระ

2. “儿” (เออร์) ที่ไม่มีความหมายโดยเนื้อแท้ ไม่มีคุณสมบัติในการแยกแยะความหมายของคำคือหน่วยคำไม่อิสระจะไม่ได้ป้อนลงฐานข้อมูล หลังจากนั้นแยกแยะหน่วยคำพ้องรูป และคำพ้องเสียง โดยการเติมตัวเลขเพื่อให้เกิดความแตกต่างเช่น คำพ้องรูป jiao3 และ jue2 นอกจากนั้นยังมี se4 และ shai3 คำพ้องเสียงอย่างเช่น shang¹ คำนาม และ shang² คำกริยา เมื่อจัดการกับข้อมูลเป้าหมายเรียบร้อยแล้ว จะได้ฐานข้อมูลคำประสม ต่อจากนั้นใช้ Excel แยกหน่วยคำ เพื่อให้ได้ฐานข้อมูลหน่วยคำ

ขั้นตอนที่สอง ใช้ Excel สืบค้นหน่วยคำที่มีการนำมาสร้างคำประสมตั้งแต่ 63 คำขึ้นไป จำนวน 300 หน่วยจากฐานข้อมูลหน่วยคำ หน่วยคำเหล่านี้ คือกลุ่มหน่วยคำเป้าหมายทางกริยาวิย ซึ่งงานวิจัยนี้เรียกว่า“หน่วยคำความถี่สูง”

ขั้นตอนที่สาม คำนวณความถี่ในการประสมคำ ลักษณะทางไวยากรณ์ ลักษณะทางความหมายของหน่วยคำความถี่สูง หลังจากนั้นศึกษาวิเคราะห์ปัจจัยต่าง ๆ ที่ส่งเสริมให้เกิดปรากฏการณ์ทางภาษาเหล่านี้

3. การให้ค่านิยามหน่วยคำความถี่สูง

ความถี่ (frequency) ทางสถิติศาสตร์ให้ค่านิยามไว้ว่า“จำนวนครั้งที่เกิดขึ้นในระยะเวลาหนึ่ง ความถี่ของคำมักจะหมายถึงจำนวนครั้งที่ปรากฏในตำรา นิตยสาร หนังสือพิมพ์หรือสิ่งพิมพ์อื่น ๆ”

งานวิจัยนี้ถือจำนวนครั้งที่หน่วยคำหรือรากคำซึ่งเป็นหน่วยประกอบหลักของคำประสมที่ปรากฏในพจนานุกรมจีนปัจจุบัน ฉบับปรับปรุง ค.ศ. 2007 หรือกล่าวอีกนัยหนึ่งจำนวนครั้งของหน่วยคำที่มีการนำมาใช้ในการประสมคำมาเป็นมาตรฐานในการกำหนด “หน่วยคำความถี่สูง”

4. สถิติและการวิเคราะห์ความถี่ของหน่วยคำความถี่สูง

งานวิจัยวิเคราะห์หน่วยคำตามจำนวนครั้งที่ใช้ในการประสมคำ พบว่าหน่วยคำที่มีความถี่สูง-ต่ำต่างกันต่างมีลักษณะเด่นที่แตกต่างกัน ดังนี้

1. ประสมคำมากกว่า 200 คำ มี 16 หน่วยคำ เรียงตามความถี่มากไปหาน้อยได้ดังนี้ ren, da, tou, xin, shui, qi, shou, bu, xiao, feng, sheng, kou, tian, huo, xue, ti

ดูจากความหมายแล้วพบว่า หน่วยคำเหล่านี้ส่วนใหญ่มีความหมายเป็นนามธรรม และมีความหมาย ธรรมดาสามัญที่สุดที่มนุษย์ทุกคนต้องมีและพบเห็น ประกอบด้วยหน่วยคำ ที่มีความหมายเกี่ยวกับส่วนต่าง ๆ ในร่างกายมนุษย์ คำบอกรัตถุของในธรรมชาติ เป็นต้น

Xu Shen (2007) กล่าวไว้ใน “Shuo wen jie zi” ว่า “สิ่งใกล้่นำมาจากตัว สิ่งไกล่นำมาจากวัตถุ”¹ แสดงให้เห็นว่าจุดกำเนิดดั้งเดิมของคำมาจากสิ่งใกล้ตัวก่อน นั่นคือคำนามที่มีความหมายเกี่ยวกับร่างกายมนุษย์ คำกริยาที่มาจากเคลื่อนไหวของร่างกายมนุษย์ คำเหล่านี้ล้วนเป็นคำพื้นฐานในภาษาดั้งเดิมของมนุษยชาติ และเป็นคำหลักในสมัยโบราณ จนกลายเป็นคำพื้นฐานที่ใช้ต่อเนื่องกันมานับพันปี หน่วยคำเหล่านี้จึงเป็นรากคำดั้งเดิมที่เก่าแก่ มีความหมายที่เป็นพื้นฐานโดยไม่เปลี่ยนแปลงด้านความหมาย และเคยใช้เป็นคำอิสระในภาษาจีนโบราณ แต่ในภาษาจีนปัจจุบันบางคำได้กลายเป็นส่วนประกอบหลักของคำ บางคำไม่ได้เป็นคำโดด บางคำยังคงนำมาใช้เป็นคำโดดในบางเงื่อนไข และหน่วยคำเหล่านี้ยังคงเป็นคำพื้นฐานในภาษาจีนดั้งเดิมด้วย

2. ประสมคำตั้งแต่ 100-199 คำ มี 122 หน่วยคำเรียงตามความถี่มากไปหาน้อยได้ดังนี้ che, fa, wen, dian, wai, men, chu, gong, jia, hua, dong, dao, xing, shi, qing, zhong, fa, kai, shu, yan, xian, he, guang, ming, guo, zi, liu, mian, yi, di, da, dian, zhu, se, nian, fen, yin, zheng, shang, shi, xing, jin, yi, ping, tong, gong, chang, wu, yu, lu, ding, hui, ci, you, hai, zhan, re, zhi, jun, bing, bao, jiao, bi, ming, hui, li, ma, ri, shu, jiao, yong, qing, shi, hua, wu, pin, sheng, shen, ji, cheng, nei, hong, hou, xing, zuo, shi, kong, shen, fang, luo, shang, dan, shi, bao, bing, min, zi, zhong, qi, huang, pan, quan, shan,

¹ Xu Shen 2007 Shuo wen jie zi สำนักพิมพ์จงหัว

qi, bai, jie, fang, dui, xue, lao, bian, zhuan, fan, shi, du, pi, bao, ban, guo, yan, lai, shou

หน่วยคำความถี่สูงในกลุ่มนี้เป็นหน่วยคำพื้นฐานที่จำเป็นในการสื่อสารภาษาจีนปัจจุบัน มีความหมายประกอบด้วยตำแหน่งและการเคลื่อนไหว การใช้มือทำอาการต่าง ๆ คำบอกวัตถุของในธรรมชาติ คำเรียกสี เป็นต้น หน่วยคำในคำประสมบางหน่วยมีการประสมคำได้ เป็นจำนวนมากอีกทั้งยังมีความหมายที่ไม่เปลี่ยนแปลง หน่วยคำบางหน่วยมีลักษณะของคำกึ่งคำเติมหน้าหรือกึ่งคำเติมท้าย เช่น ban-, -jia, fan-, -re, wu- เป็นต้น หน่วยคำเหล่านี้นำมาใช้สร้างคำใหม่ได้มากกว่าร้อยคำ เป็นกลุ่มหน่วยคำที่สร้างสมาชิกใหม่ให้แก่ระบบคำของภาษาจีนปัจจุบันได้ เป็นจำนวนมาก

3. ประสมคำน้อยกว่า 100 คำ มี 162 หน่วยคำเรียงตามความถี่มากไปหาน้อยได้ดังนี้

xi, gong, jie, xia, chuan, tai, liang, you, ye, jian, li, si, cai, qin, wei, yuan, yue, pian, jia, hao, ben, qi, biao, huo, mao, gu, tong, shuo, qian, yao, fu, nü, li, hua, ke, fu, hei, lun, jie, jin, shi, jiao, hua, du, qiu, mu, jie, you, biao, xin, gan, yu, li, huo, zhao, jing, shi, li, ru, ya, zui, an, gao, shi, ju, an², mei, ji, chang, yang, ling, shou, ye, zheng, zhi, di, jing, shou, shang, ban, jiu, mu, tie, zhi, jing, lian, xuan, mu, tao, zhan, zhi, guo, ji, cao, qiu, huo, tiao, bu, mi, shang², tu, za, kao, leng, xia², yin, ying, yi, cuo, lun, guan, fang, wu, yan, cun, bu, cai, si, cha, jue, ben², qu, bian, xiao, zou, zhi, chun, an, shi, yin, du, wang, yue, zhang, chen, zhi, quan, cha, tui, gu, xing, guan, xiang, yin, zhang, qu, ban, gai, xiu, yang, gen, bian, fan, piao, guan, jiao, ti, hun, jian, shou, wang

หน่วยคำความถี่สูงในกลุ่มนี้มีความหมายเกี่ยวข้องและใกล้ชิดกับชีวิตประจำวันของมนุษย์ อาทิ เกี่ยวข้องกับการผลิต เศรษฐกิจ วิทยาศาสตร์เทคโนโลยี และวัฒนธรรม เป็นต้น

นอกจากนั้นยังเป็นหน่วยคำที่มีการนำมาใช้ประสมคำอย่างแพร่หลาย หรือกล่าวอีกนัยหนึ่งหน่วยคำความถี่สูงที่มีจำนวนมากกลุ่มนี้ ทำให้มีหน่วยคำบางคำมีความเป็นไปได้สูงที่จะนำมาใช้ประสมคำมากขึ้นตามแนวโน้มสังคมปัจจุบันที่มีสิ่งใหม่เกิดขึ้นตลอดเวลา ดังเช่น เกิดมีหน่วยคำที่มีความหมายเกี่ยวกับการผลิต อาทิ wang, ji, qi, jin, tie, you, shi, shang, zhang, jia เป็นต้น หน่วยคำเหล่านี้ได้สะท้อนให้เห็นถึงลักษณะเด่นของ

สังคมปัจจุบันที่การสื่อสาร อุตสาหกรรม เทคโนโลยี และเศรษฐกิจ ได้กลายเป็นประเด็นที่น่าสนใจของสังคมปัจจุบัน

จากการวิเคราะห์ข้างต้นเห็นได้ว่าหน่วยคำความถี่สูงในช่วงต้นยังคงดำรงไว้ซึ่งลักษณะพิเศษของคำพื้นฐานในภาษาจีนโบราณ หน่วยคำความถี่สูงในช่วงกลาง และช่วงปลายบางหน่วยคำเป็นคำพื้นฐานในสมัยโบราณที่ลดความถี่ในการใช้ลง และหน่วยคำบางคำได้เข้ามาเป็นสมาชิกใหม่ของระบบคำพื้นฐาน ทั้งนี้จะเห็นได้ว่าปัจจัยทางสังคมและวัฒนธรรมทำให้เกิดการเปลี่ยนแปลงของระบบคำพื้นฐาน

5. ปัจจัยที่ส่งเสริมให้เกิดเป็นหน่วยคำความถี่สูง

เมื่อได้บัญชีหน่วยคำความถี่สูง แล้วก็นำมาวิเคราะห์ปัจจัยที่ส่งเสริมให้เกิดเป็นหน่วยคำความถี่สูงในแต่ละด้าน

5.1 ความหลากหลายทางความหมายของหน่วยคำ หน่วยคำที่มีความถี่สูงส่วนใหญ่ มีความหมายหลากหลาย หน่วยคำส่วนมากนอกจากมีความหมายเดิมแล้ว ยังมี ความหมายที่เกิดขึ้นใหม่ และความหมายเปรียบเทียบกับอีกด้วย ตามทฤษฎีอรรถศาสตร์ แต่ละความหมายในหน่วยคำความถี่สูงสามารถทำหน้าที่เป็นส่วนประกอบคำ เพื่อประสมกับหน่วยคำอื่นจนเกิดเป็นคำประสมพร้อมเกิดความหมายใหม่ขึ้นมา โดยทั่วไปแต่ละความหมายมีจำนวนครั้งในการนำมาใช้ประสมคำต่างกัน นั่นหมายถึงการประสมคำของหน่วยคำหนึ่งหน่วย หน่วยคำที่มีความหมายมากจะนำมาสร้างคำประสมใหม่มากขึ้นตามไปด้วย

จากผลการคำนวณจำนวนความหมายของหน่วยคำพบว่า หน่วยคำความถี่สูงในช่วงต้นมีจำนวนความหมายเฉลี่ย 8.56 ความหมาย ในช่วงกลางเฉลี่ย 7.84 ความหมาย และจำนวนความหมายเฉลี่ย 5.90 ความหมาย ในความถี่ช่วงปลาย เห็นได้ว่าหน่วยคำที่มีจำนวนความหมายมาก จำนวนครั้งในการประสมคำก็มากตามไปด้วย

5.2 ความหลากหลายทางด้านไวยากรณ์ของหน่วยคำความถี่สูง หน่วยคำส่วนใหญ่ ล้วนมาจากคำโดดในภาษาจีนโบราณ ต้นกำเนิดที่ยาวนานของหน่วยคำทำให้เกิดมีความหมายใหม่เพิ่มขึ้นตามกาลเวลา ลักษณะทางไวยากรณ์ของแต่ละความหมายก็มีความหลากหลายตามความหมายที่เพิ่มขึ้น หน่วยคำหนึ่งหน่วยอาจเป็นได้ทั้งคำนาม คำกริยา และคำวิเศษณ์ตามแต่ละความหมาย หน่วยคำความถี่สูงที่มีความหลากหลายทางความหมาย และลักษณะทางไวยากรณ์ จะนำมาใช้ในการประสมคำได้เป็นจำนวนมากและเป็นอิสระ

5.3 ขอบข่ายความหมายที่กว้างของหน่วยคำความถี่สูง ทำให้สร้างคำได้อย่างมีอิสระ ความหมายของหน่วยคำแต่ละความหมายประสมกับหน่วยคำอื่น เกิดเป็นคำประสมใหม่ จึงเป็น อีกปัจจัยหนึ่งที่ทำให้หน่วยคำกลายเป็นหน่วยคำที่มีความถี่สูง

5.4 การผลักดันทางจิตวิทยา โดยการนำหน่วยคำความถี่สูงมาเป็นรากคำ เพื่อประสมกับ หน่วยคำอื่น ประกอบเป็นกลุ่มคำประสมที่มีรากศัพท์เดียวกัน ที่เรียกว่าคำรากศัพท์เดียวกัน และเนื่องจากคำที่มาจากรากศัพท์เดียวกันได้เชื่อมโยงความหมายระหว่างกัน อาทิคำประสมที่มี “di” เป็นรากคำนำมาประสมเป็นคำประสมที่มีรากศัพท์เดียวกันคือ dimian, didi, dixia, didong, dijiao, dibao, dipping, dishi, dixing, diwen, dilei, digou เป็นต้น ถึงแม้ว่าคำเหล่านี้จะมีความหมายแตกต่างกัน แต่ในขณะเดียวกันก็มี ความหมายเชื่อมโยงกัน

5.5 หน่วยคำความถี่สูงเมื่อใช้เป็นคำโดดก็คือคำศัพท์พื้นฐาน Stalin (2514) ได้เขียน เกี่ยวกับเรื่องนี้ไว้ในหนังสือเรื่องลัทธิมาร์กกับปัญหาภาษาศาสตร์ว่า “องค์ประกอบหลักใน ด้านศัพทานาสาตร์ทางภาษาศาสตร์ก็คือคำศัพท์พื้นฐาน นั่นก็หมายรวมถึงรากคำทั้งหมด คำศัพท์พื้นฐานมีจำนวนน้อยกว่าคำศัพท์ทั่วไปมาก แต่ทว่ามีความเก่าแก่ กำเนิดขึ้นเมื่อร้อย พันปีก่อน จนกลายเป็นพื้นฐานของการสร้างคำใหม่”¹

Lin Tao นักวิชาการที่ทุ่มเทกับการวิจัยคำศัพท์พื้นฐานเห็นว่าทฤษฎีของ Stalin มีเนื้อหา ที่เป็น “เอกสารสำคัญในการวิจัยคำศัพท์พื้นฐาน” “ปัญหาอยู่ที่ว่าจะตีความ ทำความเข้าใจ และ นำทฤษฎีของเขาไปใช้ให้ถูกต้องได้อย่างไร”²

ต่อมาได้มีการกำหนดมาตรฐานการจำแนกคำศัพท์ออกเป็นคำศัพท์พื้นฐาน และ คำศัพท์ทั่วไป (ไม่พื้นฐาน) ไว้ 3 ประการด้วยกันคือ คำศัพท์พื้นฐานมีลักษณะทาง ความหมายที่ไม่เปลี่ยนแปลง เป็นคำศัพท์ที่ใช้ทั่วไปที่มนุษย์ทุกคนพบเห็น และมี จำนวนครั้งในการนำไปใช้สร้างคำใหม่สูง ลักษณะสำคัญของคำศัพท์พื้นฐานทั้ง 3 ประการนี้มีความสำคัญลดหลั่นกันไป ทั้งนี้ต้องกำหนดความสำคัญตามกลุ่มเป้าหมาย การวิจัย สำหรับการทำการวิจัยในภาษาจีนปัจจุบัน ควรต้องพิจารณาคุณสมบัติที่มี

¹ Stalin 2514 ลัทธิมาร์กกับปัญหาภาษาศาสตร์ สำนักพิมพ์เหรินหมิน

² Lin Tao “ปัญหาบางเรื่องของคำศัพท์พื้นฐานในภาษาจีน” วารสาร Zhongguo Yuwen, ปี 1954 หน้า 7

จำนวนครั้งในการนำมาใช้สร้างคำใหม่มากมาเป็นอันดับแรก ดังนั้นเมื่อ ทำการศึกษา หน่วยค่าความถี่สูงในภาษาจีนปัจจุบัน ประการแรกที่ต้องคำนึงถึง คือคุณสมบัติ ทางด้านจำนวนครั้งที่มีการนำมาใช้สร้างคำใหม่นั้นเอง

ประการต่อมาต้องคำนึงถึงว่า เป็นคำทั่วไปที่มนุษย์ทุกคนพบเห็น ในการวิจัย ภาษาจีนปัจจุบันวิธีที่สำคัญ และกะทัดรัดที่สุด คือการคิดคำนวณสถิติการนำหน่วยค่า มาสร้างคำใหม่ โดยทั่วไป คำศัพท์ที่พบบ่อยคือคำที่มีการนำมาใช้อย่างกว้างขวาง แสดงให้เห็นว่าผู้คนนิยมใช้ดังนั้นหน่วยค่าที่ปรากฏบ่อยในพจนานุกรมจีนปัจจุบันจึง เป็นคำศัพท์พื้นฐานไปโดยปริยาย

ประการสุดท้ายพิจารณาคำศัพท์พื้นฐานจากความหมายที่ไม่เปลี่ยนแปลง หน่วยค่าความถี่สูงส่วนใหญ่เป็นคำที่ใช้มาตั้งแต่สมัยโบราณอย่างเช่น **bing, tian, ben, hui, gao, da, shou, shen, zou**

สรุปได้ว่าหน่วยค่าความถี่สูงของงานวิจัยนี้ มีคุณสมบัติของการเป็นคำศัพท์ พื้นฐานทั้งสิ้น แต่คำศัพท์พื้นฐานแตกต่างกันตามยุคสมัย คำศัพท์พื้นฐานในภาษาจีน สมัย Xianqin และ Lianghan กับคำศัพท์พื้นฐานในภาษาจีนปัจจุบันจึงแตกต่างกัน อย่างแน่นอน ดังนั้นหน่วยค่าความถี่สูงที่เป็นผลการวิจัยของงานวิจัยนี้สามารถกล่าวได้ ว่าเป็นคำศัพท์พื้นฐานของภาษาจีนปัจจุบัน

5.6 งานวิจัยนี้ได้นำความหมายหลัก (ความหมายที่ใช้บ่อย) มาเป็นตัวกำหนด ลักษณะทางไวยากรณ์ของหน่วยค่าความถี่สูงพบว่า หน่วยค่าความถี่สูงในภาษาจีน ปัจจุบันมีลักษณะทางไวยากรณ์ที่เป็นคำนามร้อยละ 55.46 ร้อยละลงมาคือคำกริยาร้อย ละ 31.68 คำวิเศษณ์ร้อยละ 11.88 ที่เหลือคือคำบอกจำนวนร้อยละ 0.33 และคำขยาย กริยาร้อยละ 0.66

5.7 เมื่อวิเคราะห์จากมุมมองปริชาน (Cognitive) พบว่าผู้คนมีความสามารถ ในการบ่งชี้วัตถุมากขึ้นได้ ตามแนวคิดการสื่อภาษาความแม่นยำนี้ทำให้เกิดการสื่อสาร ที่ชัดเจนยิ่งขึ้นด้วยการกำหนดคำหลัก และคำย่อยที่แตกต่างกัน จึงทำให้ภาษาละเอียด และชัดเจนมากขึ้น เช่น ผู้คนรับรู้ถึงคุณสมบัติของสีอย่างละเอียด กล่าวคือบริเวณ ความหมาย (Semantic fields) โทนสี ในผลการวิจัยหน่วยค่าความถี่สูงในภาษาจีน ปัจจุบัน ปรากฏหน่วยค่าเรียกสี อาทิ **bai, hong, huang, hei, qing** โดย

การเติมส่วนขยายบอกลักษณะความละเอียดของโทนสีไว้ข้างหน้าเพื่อสร้างคำใหม่ขึ้นได้เป็นจำนวนมาก แสดงให้เห็นว่าผู้คนแยกแยะโทนสีเข้มอ่อน มีดีสว่างได้อย่างละเอียดมากขึ้นโดยผ่านทางภาษา

เพราะฉะนั้นต้นกำเนิดเก่าแก่ ขอบข่ายกว้างของความหมาย ลักษณะทางไวยากรณ์ ปริมาณทางภาษา ล้วนเป็นปัจจัยที่ทำให้เกิดหน่วยคำความถี่สูงขึ้น ดังนั้นหากศึกษาวิจัยทางด้านคำศัพท์ควรนำปัจจัยต่าง ๆ ดังกล่าวมาพิจารณาให้ครบถ้วน