

基于量化统计的通感式复合词通感模式分析

刘志芳

(河南师范大学 文学院,河南 新乡 453007)

摘要:通感作为一种修辞手法,在语言词汇层面的主要体现是通感式复合词.在选取典型语料的基础上,运用单、双因素方差分析法,深入分析通感式复合词的通感模式.量化统计结果显示,通感式复合词的通感模式主要包括:(1)视觉是通感引申过程中最大的源域;(2)听觉是通感引申过程中最大的目标域;(3)通感引申规律基本遵循“视觉→肤觉→味觉→嗅觉→听觉”的排列顺序,但是个别感觉之间存在双向引申关系,如视觉和肤觉之间存在大量的双向引申的例子;(4)各感觉作为通感引申源域的数量与感觉类形语素数量成正比,即含有形语素数量越多,成为最大源域的可能性就越大.

关键词:量化统计;通感式复合词;通感模式

中图分类号:O212;H03

文献标志码:A

通感在《现代汉语词典》(第6版)中的释义是“人们在审美活动中使各种感官产生的感觉互相沟通,互相转化,产生新奇的效果”^[1].通感作为一种描写手法最早是由钱钟书先生引进文学研究领域的.通感修辞在语言的语音、词汇、语法、语义、文字等子系统中都有体现,通感在语言词汇层面的具体表现主要包括两种,一个是“通感生义”,即通感引起的词义变化,也就是一个词从本义通过通感方式产生了引申义.运用“通感生义”方式产生的词就叫作通感词.另一个是“通感造词”,即运用通感方式组成复合词,通感造词的结果是形成通感式复合词.通感式复合词由两个从本义或基本义来说来自不同感觉域的形语素组合而成,大多为形容词.通感式复合词作为通感修辞在语言词汇层面的典型成员,有其独特的研究价值和研究意义.前人对通感式复合词的研究大多为定性的研究,往往从通感式复合词的结构特点、语义特征和语义演变等角度展开.在前人研究基础之上,力图运用量化统计的方法,以典型语料为数据来源,深入分析通感式复合词的通感模式.

量化统计法,因其客观性、直观性、系统性等特点,在自然科学研究的诸领域中已被广泛使用.随着人们对定性定量相结合的研究法的重视,量化统计法在人文社科研究的各领域中也呈现出极大的发展潜力.语言学相对于其他人文学科来说,是最接近于自然科学的人文学科.语言学下辖的语音、词汇、语法、语义等各领域都非常重视量化统计法.运用量化统计法可以较为直观地分析语言各子系统的属性,真正实现定量与定性的结合.苏新春^[2](2007)指出计量研究是语言研究中的一种重要方法,就是通过对语言的结构、分布、使用、变迁等要素进行数量分析来揭示语言的状态、性质与特点的一种方法.计量方法特别适用于词汇,因为词汇具有以下3个特点:词语数量庞大;词语载体明显,词形清晰;组合关系复杂^[3].前人对词汇的计量研究,为运用量化统计的方法研究通感式复合词的通感模式提供了保障.

1 方差分析法

1.1 方差分析的特点和使用条件

方差分析法(Analysis of Variance,简称ANOVA)是1918年由R. A. Fisher首次引入的,是一种检验两个以上方差的正态总体均值之间是否存在差异的统计方法.它是对多种效应同时作用的测量数据进行分析,

收稿日期:2019-01-10;修回日期:2019-06-16.

基金项目:教育部人文社科项目(17YJC740062);河南师范大学博士科研启动项目.

作者简介(通信作者):刘志芳(1980—),女,河南三门峡人,河南师范大学校聘副教授,博士,研究方向为汉语词汇和词汇教学,E-mail:lzf211@163.com.

判定各种因素对研究对象的特定指标影响大小的方法.方差分析以测量数据为基础,而数据通常由试验获得.在试验中,将要考察的研究对象的特定指标称为试验指标,影响试验指标的条件称为因素,因素所处的状态称为水平^[4].若试验中只有一个因素在变化,而其他因素保持不变,对试验结果所做的这种方差分析称为单因素方差分析.若试验中同时有多个因素在变化,此时进行的方差分析称为多因素方差分析.

设 X 为一随机变量,若 $E[X - E(X)]^2$ 存在,则称这个值为 X 的方差,记为 $D(X)$,即 $D(X) = E[X - E(X)]^2$.又称 $\sqrt{D(X)}$ 为 X 的标准差或均方差^[5].方差分析的基本思想就是将试验指标值的方差分解成条件变差与随机误差,然后将各因素形成的条件变差与随机误差进行比较,评价由某种因素所引起的变异是否具有统计学意义.方差分析的使用条件:被检验的各个总体都服从正态分布;各个总体的方差相等,具有方差同质性;变异是可以分解的和可加的,总变异可以分解为几个不同的、独立的来源.

1.2 单因素方差分析

单因素方差分析的目的在于判断因素 A 对指标 X 是否有影响,若有影响需要进一步指出 A 的哪个水平对 X 最有利.设单因素 A 有 r 个水平 A_1, A_2, \dots, A_r ,在水平 $A_i (i=1, 2, \dots, r)$ 下进行 n_i 次独立试验,样本值为 $x_{i1}, x_{i2}, \dots, x_{in_i}$,它们来自均值分别为 μ_i 且具有相同方差 σ^2 的正态总体 $N(\mu_i, \sigma^2)$.其中 μ_i, σ^2 均未知,在不同水平 A_i 下的样本之间相互独立.单因素方差分析步骤如下.

(1) 计算总偏差平方和及其自由度.计算试验的 $N = \sum_{i=1}^r n_i$ 个数据的总偏差平方和 $S_T = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$, $\bar{x} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} x_{ij}$, $f_T = n - 1$,对 S_T 进行分解可得 $S_T = S_E + S_A$,其中 $S_E = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2$, $\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$, $f_E = n - r$, $S_A = \sum_{i=1}^r \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x})^2$, $f_A = r - 1$,上面两式中, S_E, S_A 分别称为误差平方和及效应平方和, f_E, f_A 是相应的自由度;

(2) 计算均方 $MS_E = \frac{S_E}{f_E}, MS_A = \frac{S_A}{f_A}$;

(3) 计算 F 统计量的值: $F = \frac{MS_A}{MS_E} \sim F(f_A, f_E)$ 对给定的显著性水平 α ,查表得 $F_\alpha(f_A, f_E)$;

(4) 比较 F 与 $F_\alpha(f_A, f_E)$,若 $F > F_\alpha(f_A, f_E)$,则认为组内与组间有明显差异,否则认为组内与组间无明显差异.

实际问题中,如果影响指标 X 的因素不止一个,需要考虑双因素方差分析,分析方法和单因素方差分析步骤类似,这里不再赘述.

1.3 本文要研究的问题

基于方差分析法的统计特性,在自然科学和人文社会科学研究中,经常会用到方差分析法对研究对象进行定量统计.前人主要将方差分析法应用到对环形激光陀螺仪噪声的性能评估^[6]、对重点审计对象的遴选^[7]、对通货膨胀风险下关于累积阶段的固定缴费养老金的研究^[8]、对汉语感官词的语义语法研究^[9]等问题.在前人研究基础上,认为运用方差分析法对现代汉语通感式复合词的通感模式进行量化统计具有可行性.一方面,通感一般发生在两个具体的感觉域之间,通感式复合词内部语素的本义或基本义所属的感觉域是源域,通感式复合词语义所属的感觉域是目标域.通感发生时,总会从源域向目标域发生通感隐喻.通感的发生到底是受源域影响还是受目标域影响,是需要解决的首要问题.方差分析法中的双因素方差分析可以解决这一问题.另一方面,发生通感的五种感觉视、听、肤、味、嗅之间的通感模式是研究的目标,具体来说就是从哪种具体感觉域向另一种具体感觉域发生通感隐喻.在分析通感式复合词的通感模式时,主要使用单因素方差分析法.

2 研究过程

2.1 通感式复合词的界定和提取

以《现代汉语词典》(第6版)为语料来源,运用形式标准和语义标准提取出现代汉语层面的通感式复合

词 114 个,提取结果如下:

暗淡、黯黑、白嫩、白润、苍凉、苍润、长圆、敞亮、沉寂、沉静、沉重、澄碧、粗大、粗浅、脆亮、大方、大红、淡薄、淡青、干瘪、干冷、高寒、高明、高燥、光滑、光润、光鲜、寒苦、寒素、寒酸、寒微、黑暗、红润、厚重、滑腻、灰暗、晦涩、昏沉、昏黑、昏黄、浑厚、尖酸、坚苦、苦寒、苦涩、宽松、冷淡、冷寂、冷静、冷清、明白、明澈、明黄、明锐、明细、嫩红、嫩黄、嫩绿、浓厚、浓重、平白、平淡、平滑、平静、浅白、浅明、青涩、轻薄、轻淡、轻微、清白、清脆、清淡、清高、清寒、清寂、清静、清苦、清冷、清凉、清亮₁、清亮₂、清冽、清明、温馨、热辣、热闹、柔细、潏热、深重、素淡、酸软、甜润、温厚、温柔、温润、温馨、乌亮、细嫩、细腻、细软、细润、纤柔、鲜红、鲜亮、鲜明、鲜嫩、香甜、响亮、圆滑、圆润、燥热、直白、重大

主要通过 114 个词在语言层面和言语层面的语义分析,研究其语义引申规律。所谓语言层面指的是语言系统内部的发展变化,在语言层面分析通感式复合词的语义,主要是依据词典释义静态地分析其语义特点,适当考察其动态的语义变化。所谓言语层面指的是在语言的使用中考察通感式复合词的语义变化,依据现代汉语层面的典型语料,分析其语义变化规律。当然对语言层面和言语层面的语义分析也不是截然分开的,有时也需要针对具体考察对象,将二者有机结合起来。

词义具有概括性,语言层面的词义是从语言实际中总结归纳出来的词的一般用法,具有概括性、客观性和抽象性。言语层面的词义在言语交际过程中会被具体化和主观化。也就是说,概括的语言层面的词义在言语层面会因交际的需要而出现不同的变体。在言语层面分析通感式复合词的语义,就是要从实际语料中发现其语义变体,并分析其产生的原因。

2.2 感觉域之间投射规律的分析

在分析通感式复合词通感引申模式之前要明确统计的基本步骤。首先分析单义词在语言层面和言语层面的通感引申规律,然后分析多义词在语言层面和言语层面的通感引申规律。在统计时,单纯词以本义所属感觉域为源域,以引申义在语言层面和言语层面所属的感觉域为目标域。复合词以其内部两语素本义或基本义所属的感觉域为源域,以复合词词义在语言层面和言语层面所属的感觉域为目标域。如果复合词内的语素义不存在兼属情况,则以语素基本义所属的感觉域为源域,如果复合词内的语素义存在兼属情况,则以语素本义所属的感觉域为源域。

114 个通感式复合词中有 22 个广义通感式复合词的词义在语言层面和言语层面均为两个语素所属语义域的上级感觉域,也就是说,词义所属的感觉域与语素义所属的感觉域之间存在属种关系,无法构成通感引申,因此予以排除。对 92 个发生通感引申的词语的通感引申方向进行统计,统计结果见表 1。

表 1 通感引申方向

Tab.1 Synaesthetic extension direction

通感方向	数量	通感方向	数量	通感方向	数量	通感方向	数量
视→听	23	肤→视	22	味→视	5	嗅→视	0
视→肤	20	肤→听	17	味→听	3	嗅→听	0
视→味	5	肤→味	5	味→肤	3	嗅→肤	1
视→嗅	7	肤→嗅	5	味→嗅	3	嗅→味	1

从表 1 可知,除了听觉外其他 4 种感觉都会向其他感觉域发生通感引申,其中又以视觉和肤觉之间的通感引申,以及视觉向听觉的引申和肤觉向听觉的引申居多。嗅觉作为源域,只向肤觉和味觉发生通感引申,没有向视觉和听觉的通感引申。一方面是因为所选语料中表肤觉和视觉的语素较多,表听觉和嗅觉的语素较少,通感引申的源域数量与语素数量成正比。另一方面可以从心理学对感觉的分类加以解释,心理学按照感觉的重要性将其分为视觉、听觉和其他感觉,其他感觉包括皮肤感觉、嗅觉和味觉,还有内部感觉^[10]。其中视觉是人类最重要的一种感觉,在人类获得的外界信息中,80%来自视觉。肤觉和味觉都是人体相应器官与物体或物质直接接触而产生的感觉。听觉和嗅觉的产生是人体相应器官与物体间接接触而产生的,与肤觉和味觉相比具有不稳定性。

根据表 1 的内容可以得出表 2,表 2 反映了 5 种具体感觉作为源域和目标域的情况。其中视觉是最大的源域,听觉是最大的目标域。

表2 5种感觉作为源域和目标域的分布表

Tab.2 The distributions of five sensations as source and target domains

类型	视觉	听觉	肤觉	味觉	嗅觉
源域	55	0	49	14	2
目标域	27	43	24	11	15

从表2可知,5种具体感觉作为源域的可能性由大到小分别是视觉,肤觉,味觉,嗅觉,听觉,作为目标域的可能性由大到小分别是听觉,视觉,肤觉,嗅觉,味觉.从现有语料来看,视、听、肤、味、嗅5种具体感觉在作为源域和目标域时,没有特定的规律可循.为了更加清晰地分析出通感式复合词的通感模式,有必要对其进行量化统计.运用统计学中的单、双因素方差分析法能够较清晰地反映出通感式复合词的源域和目标域对通感机制的影响,也可以更加直观地反映出5种感觉域在通感引申中的地位和作用.

2.3 双因素方差分析

在通感引申中主要涉及源域和目标域两个因素,因此首先要对通感式复合词做双因素方差分析,从数据中分析出源域和目标域两个因素哪个对通感式复合词的影响较大.为了保证数据分析的准确性,在分析时主要用Excel数据中的数据分析功能进行无重复双因素方差分析和单因素方差分析.

在进行方差分析之前,先要把表1转换为方差分析表,即表3.

表3中以听觉为源域的第2行和以嗅觉为源域的第1列和第2列数据为0,说明在目前搜集到的语料中没有以听觉为源域的例子,也没有以嗅觉为源域、视觉和听觉为目标域的例子.表3是分别以5种感觉域为源域和目标域组成的正方形矩阵,它的对角线分别是5种感觉内部的互通,无法构成通感,因此没有数据存在.为了计算的方便,将该对角线中的数据记为0.通过Excel的数据分析功能得到了无重复双因素方差分析的相关数据(见表4),同时还需要参照无重复试验的双因素方差分析表(见表5).

表3 方差分析表

Tab.3 ANOVA table

源域 <i>s</i>	目标域 <i>t</i>				
	视觉	听觉	肤觉	味觉	嗅觉
视觉	—	23	20	5	7
听觉	0	—	0	0	0
肤觉	22	17	—	5	5
味觉	5	3	3	—	3
嗅觉	0	0	1	1	—

表4 无重复双因素方差分析结果统计表

Tab.4 Statistical table of analysis of variance without duplication of two factors

差异源	SS	df	MS	<i>F</i>	<i>P</i> -value	<i>F</i> -crit
行	549.2	4	137.3	3.482 562	0.031 501	3.006 917
列	124	4	31	0.786 303	0.550 668	3.006 917
误差	630.8	16	39.425			
总计	1 304	24				

表5 无重复试验的双因素方差分析表^[4]

Tab.5 Two-factor ANOVA table without repetitive trials

偏差	平方和	自由度	均方	<i>F</i> 比
因素 A	S_A	$r - 1$	$MS_A = \frac{S_A}{r - 1}$	$F_A = \frac{MS_A}{MS_E}$
因素 B	S_B	$s - 1$	$MS_B = \frac{S_B}{s - 1}$	$F_B = \frac{MS_B}{MS_E}$
误差	S_E	$(r - 1)(s - 1)$	$MS_E = \frac{S_E}{(r - 1)(s - 1)}$	—
总和	S_T	$n - 1$	—	—

从表4和表5的对比可知,表5对应于表4的前5列.其中表4第1列的“差异源”即“偏差”,行是源域即

因素 A 的数据,列是目标域即因素 B 的数据,因素 A 和因素 B 统称为组间差,误差是组内差.第 2 列“平方和”与第 3 列“自由度”均具有加合性,即组间差与组内差之和是总和.第 4 列“均方”是“平方和”与“自由度”的比值.第 5 列 F 比是组间两因素均方与组内均方的比值, F 比是衡量源域和目标域两因素的重要条件.第 6 列 P 值是假定值,是一种在原假设为真的前提下出现观察样本以及更极端情况的概率.一般用 X 表示检验的统计量,当 H_0 为真时,可由样本数据计算出该统计量的值 C ,根据检验统计量 X 的具体分布,可求出 P 值.在分析统计数据时, P 值一般和规定的显著性水平进行比较.第 7 列 F 值是固定的数值,一般是在显著性水平为 0.05 的情况下查表所得,这里 $F_{0.05}(4,16)=3.01$.

从表 4 的数据可知,源域的 F 比 $>F$ 值, P 值 <0.05 ;目标域的 F 比 $<F$ 值, P 值 >0.05 .由此可见,通感式复合词的源域对通感机制有显著影响,目标域对通感机制没有显著影响.

2.4 单因素方差分析

按照表 4 的数据显示源域对通感机制有显著影响,下面只需要从源域入手,分析其内部 5 种感觉哪一种作为源域的可能性最大,哪一种作为源域的可能性最小.通过 Excel 的数据分析功能得到了单因素方差分析的相关数据,见表 6.

表 6 单因素方差分析结果统计表

Tab.6 Statistical table of the results of one-way analysis of variance

差异源	SS	df	MS	F	P -value	F -crit
组间	686.5	4	171.625 00	5.436 906	0.006 561	3.055 568
组内	473.5	15	31.566 67	—	—	—
总计	1 160	19				

从表 6 可知, F 比 $>F$ 值, P 值 <0.05 ,说明源域对通感机制有显著影响.下面先计算出 LSD 值,再来比较源域 5 种感觉之间的 LSD 与 LSD 值之间的差异. $LSD_{0.05} = \sqrt{MS_E \times (\frac{1}{n_1} + \frac{1}{n_2})} \times t_{0.05}$.这里 $MS_E = 31.566 67$, $n_1 = n_2 = 4$, $t_{0.05}$ 是固定值,可以通过查表得出. $t_{0.05} = 2.131 8$.通过计算得出 $LSD_{0.05} = 8.468 4$.

由表 7 可知,标 * 的数据大于 LSD 值,表示以视觉为参照,视觉和肤觉之间无显著差异,听觉、嗅觉和味觉与视觉之间的差异逐渐加大.以肤觉为参照,听觉、嗅觉和味觉与肤觉之间的差异逐渐加大,即听觉,嗅觉,味觉,肤觉/视觉.而味觉、嗅觉、听觉之间并无显著差异.

表 7 词语源域各感觉的 LSD 比较

Tab.7 LSD comparison of sensations in source domains

类型	X_i	$\bar{X}_i - X$ 听	$\bar{X}_i - X$ 嗅	$\bar{X}_i - X$ 味	$\bar{X}_i - X$ 肤
X 视	13.75	* 13.75	* 13.25	* 10.25	1.5
X 肤	12.25	* 12.25	* 11.75	* 8.75	—
X 味	3.5	3.5	3	—	—
X 嗅	0.5	0.5	—	—	—
X 听	0	—	—	—	—

表 4 的数据反映出源域对通感机制有显著影响,由表 7 的数据得出“听觉,嗅觉,味觉,肤觉/视觉”的排序,反映出视觉作为源域的可能性最大,其他感觉作为源域的可能性依次降低.听觉作为目标域的可能性最大,其他感觉作为目标域的可能性依次降低.

3 结 语

本文主要是基于现代汉语通感式复合词的通感引申过程,在提取数据的基础上,运用统计学单、双因素方差分析的方法,得出汉语通感式复合词的通感引申模式.本研究中视觉类形语素相对于其他感觉来说,其类型和数量都较多,因此视觉成为最大的源域.

现代汉语通感式复合词的通感引申规律可以从各感觉重要性的不同和可及性强弱两个方面来解释,其

中又以感觉的重要性为主要决定因素.按照心理学对感觉的分类,感觉可以分为视、听、肤、味、嗅5类,其中视觉是人类最重要的感官,人类对外界信息的获得80%来自视觉,接下来是听觉,然后是其他感觉,其他感觉包括肤觉、味觉和嗅觉.正是因为视觉的重要性,人们为了表义的需要才会对视觉进行下位分类,如分为空间类、颜色类、光线类和形态类等,每一个小类中都有较多的成员,所以视觉才是最大的源域.可及性强弱指的是某种感觉的产生是否需要与物体直接接触,需要与物体直接接触的是可及性强的感觉,不需要与物体直接接触的是可及性弱的感觉.5种感觉中可及性强的的是肤觉和味觉,可及性弱的是嗅觉、听觉和视觉.肤觉和味觉的产生需要感觉器官与物体直接接触,嗅觉、听觉和视觉的产生一般需要借助某种媒介物,如听觉的产生需要借助物体振动产生的声波,视觉的产生需要借助光波等.因此,肤觉是仅次于视觉的第2大源域.

语言是音义结合的符号系统,词汇作为语言系统的子系统也是形式和意义的结合体.前人对词汇的研究大都从单个词或一类词的词形和词义出发,定性地分析词的特点.所谓定性和定量的结合也大多以词频统计为主,其侧重点一般还是定性的分析.运用方差分析法研究通感式复合词能够较为有效地解决其通感模式的问题,对于词汇研究来说,只有运用量化统计的方法,才能更加客观和直观地为我们揭示某类词的词汇特点,从而拓宽词汇研究的思路.

参 考 文 献

- [1] 中国社会科学院语言研究所词典编辑室.现代汉语词典[M].6版.北京:商务印书馆,2013:1302.
- [2] 苏新春.计量方法在词汇研究中的作用及频级统计法[J].长江学术,2007,14(2):118-124.
- [3] 苏新春.词汇计量及实现[M].北京:商务印书馆,2010:22-24.
- [4] 刘文安.概率论与数理统计[M].北京:高等教育出版社,2011.
- [5] 黄清龙,阮宏顺.概率论与数理统计[M].北京:北京大学出版社,2005:81.
- [6] 黎奇,白征东,赵思浩,等. Allan 方差方法分析环形激光陀螺仪噪声的性能评估[J].清华大学学报(自然科学版),2019,59(3):1-8.
- [7] 王盘州,李俊海,刘琳琳,等.基于统计分析的重点审计对象遴选方法研究[J].河南师范大学学报(自然科学版),2014,42(5):18-23.
- [8] 张笑怡,郭军义.通货膨胀风险下关于累积阶段的固定缴费养老金的均值一方差问题[J].中国科学,2019,49(3):1-14.
- [9] 侯博.汉语感官词的语义语法学研究[D].南京:南京师范大学,2008.
- [10] 彭聃龄.普通心理学(修订版)[M].北京:北京师范大学出版社,2007:80.

Synaesthesia model analysis of synaesthetic compound words based on quantitative statistics

Liu Zhifang

(College of Liberal Arts, Henan Normal University, Xinxiang 453007, China)

Abstract: As a rhetorical device, synaesthesia is mainly embodied in synaesthetic compound words at the level of lexicon. On the basis of selecting language material, this article uses single factor and double factor analysis of variance to analyze the semantic extension rules of the synaesthetic compound words. Quantitative statistics show that the synaesthetic patterns of synaesthetic compound words mainly include: (1) vision is the largest source area in the process of synaesthetic extension; (2) hearing is the largest target area in the process of synaesthetic extension; (3) synaesthetic extension rules basically follow the order of "visual sense-dermal sensation- gustatory sense-olfactory sensation-auditory sense", but there is a two-way extension relationship between individual senses, such as visual sense and dermal sensation; (4) The number of sensation as the source domain of synaesthesia is proportional to the number of sensory morphemes, ie., the more morphemes there are, the greater the possibility of being the largest source domain.

Keywords: quantitative Statistics; synaesthetic compounds; synaesthetic model

[责任编辑 陈留院 赵晓华]