

数量词模糊划分的隶属度量表*

袁毓林

北京大学中文系, 北京 100871

yuanyl@pku.edu.cn

提要

本文根据各种数词、量词和数量词的典型成员的语法表现来选定一组分布特征, 按照它们对于有关词类的重要性、根据经验给其中的每个特征设定权值; 再用每类词的非典型成员的语法表现作校验, 做成一套可用以对相关的词进行词类模糊划分和隶属度计算的量表。希望籍此可以使得现代汉语中有关的词不仅能划归到某一个或几个词类之中(即词有定类), 而且能显示出它从属于这一个或几个词类的程度到底有多大(即类中有别)。

关键词: 词类划分 典型成员 分布特征 权值设定 模糊分类 隶属度

1. 引言: 汉语词类模糊划分的基本步骤

汉语词类并不是边界明确的普通集合, 而是从内涵到外延都不甚清晰的模糊集合。对于为数众多的词来说, 它们跟某些词类的隶属关系并不是简单的是(即隶属度为 1)或非(即隶属度为 0)这种两极对立, 而只是在一定程度上属于某一种或几种词类。有鉴于此, 我们尝试运用模糊数学中的模糊集合(fuzzy set)和模糊聚类(fuzzy clustering)的有关观念和方法来处理汉语词类问题。

我们的基本思路是采用动态聚类的方法, 以各词类的典型成员作为初始的聚类中心。具体的工作步骤是: 首先, 根据每类词的典型成员的语法表现, 来选定一组分布特征; 并按照这些不同的分布特征对于相关词类的重要性, 根据经验给其中的每个特征设定权值(weight); 其中有正分(即加分), 也有负分(即扣分)。然后, 计算总分, 典型成员应该得 100 分或接近 100 分, 非典型成员则小于 100 分、但一般大于 50 分。最后, 折合成介于区间[0,1]中的不同的值来描写词类归属模糊的词对于有关词类的隶属度(degree of membership), 从而从量上确定这些词的词类归属。比如, 某个词相对于某种词类的隶属度越趋近 1, 那么越趋近该词类的典型成员。这时, 根据实际的测试, 借鉴利用模糊关系进行聚类的方法中的 λ 值设定法, 通过设定不同的 λ 值, 来

* 本项目的研究得到国家语委语言文字应用研究“十五”科研规划项目基金的资助, 还得到北京市社科基金(百人工程项目)的资助, 在此一并致以诚挚的谢意。另外, 我要特别感谢北京大学、北京师范大学、北京语言大学和延边大学的 30 多位研究生, 他们在选修笔者 2002 年春讲授的《认知语言学》课程时, 用近千个词类属性模糊的词对这套量表的初稿分头进行了仔细的检验, 使得我能够在大量调查材料的基础上作出校验和修订。

确定不同的词从属于某个词类的典型性等级。比如，隶属度 0.8 以上为典型成员，0.6 以上为一般成员，0.6 以下为非典型成员。当一个词从属于不同的词类的最高得分相同或相近时，就说明该词兼属于这些词类（即一词多类）。

为了达到上述目标，我们先后调查和分析了上万个词类归属比较明确的常用词的分布情况，以设定不同词类的有关分布特征的权值。接着，又调查和分析了上千个词类归属不明的常用词的分布情况，以此作为对先前设定的分布权值的校验；然后用校验过的权值来计算它们相对于有关词类的得分情况，确定这些词对于有关词类的隶属度；并对这种校验过的权值的合理性进行评估，不断地作出修正，尽可能调整到令人满意的水平。

限于篇幅，本文只公布这套用以对数量词进行模糊划分的分布特征及其权值设定，希望它们能够成为对现代汉语有关的词进行隶属度计算和模糊聚类的量表(scale)或标尺性的东西(yardstick)，籍此可以对现代汉语的有关词类进行模糊划分。关于对现代汉语其他词类进行模糊划分的隶属度量表，请看袁毓林(2005a、b)。

2. 权值设定和隶属度计算的约定

为了统一规范和便于计算，我们制订如下关于分布特征的权值设定和有关词对于有关词类的隶属度的计算办法的约定：

- (1) 任何一个词，它对于某个词类的隶属度，总是在闭区间[0—1]之间。
- (2) 假定一个词完全满足某个词类的各种主要的分布特征时得 100 分，即它完全属于这一词类；那么，它属于这一词类的隶属度为 1。
- (3) 假定一个词完全不满足某个词类的各种主要的分布特征时得 0 分，即它根本不属于那个词类；那么，它属于这一词类的隶属度为 0。
- (4) 每个分布特征的权值根据经验来设定，符合该特征的得正分，不符合一般特征的得零分，不符合关键特征的得负分。
- (5) 根据一个词的总的得分（积分）来计算（折合）其相对于某个词类的隶属度，并且中最大的隶属度作为判定其词类归属的根据。如果最大的隶属度是相等的两个、或接近相等的两个或多个，那么这个词可能兼属于两种或多种词类。
- (6) 当某个词对应某个词类的积分小于 0 时，按照 0 来计算。

3. 各别词类的分布特征和权值设定

下面依次列出判定数词（分成单纯数词和合成数词两类，单纯数词又分成系数词和位数词两种）、量词（分成简单量词和复合量词两种）和数量词（分成真数量词和准数量词两种）等 3 个词类（6 种小类）的有关分布特征及其权值设定。

3. 1 数词的分布特征和权值设定

数词得分成单纯数词和合成数词两类，其中单纯数词又可根据其分布位置分为系数词和位数词两种，合成数词又可以根据其构成方式分为系位构造和系位组合两种。这些不同种类的数词小类有着不尽相同的分布特征和语法表现，因此，只有先分别讨论了这些不同种类的数词小类之后，才能回过头来讨论一般数词的分布特征和权值设定。

3. 1. 1 系数词的分布特征和权值设定

(1) 系数词包括“一、二、三、四、五、六、七、八、九、十、半、两、几、数”等，它们是粘着词，不能用来单独回答问题。

例如：“来了多少学生？~*八！~八个！~一百二十多！、买了多少斤面粉？~*六！~六斤！~三千六百八十九！、这孩子几岁？~*十！~十岁！~十二！”。

这种分布特征的形式表示是：*#_#；其中，#...# 代表句子前后两头较长的停顿。

其权值设定可以简约地表示如下：DF₁: *#_#; i. √, WV = 20; ii. ×, WV = -20。其中，DF 是分布特征 (distribution feature) 的缩写；WV 是权值 (weight value) 的缩写。上式表示：如果某词满足这种分布 (记作√, 即不能进入“#_#”这种语法环境)，那么它在系数词性方面得 20 分 (记作 WV = 20)；也就是说，光凭该词的这种分布特征，该词对于系数词这个词类的隶属度为 0.2。相反，如果某词不满足这种分布 (记作×, 即能进入“#_#”这种语法环境)，那么该词的分布特征使得它在系数词性方面得 -20 分；也就是说，光凭该词的这种分布特征，该词对于系数词这个词类的隶属度为 0。

(2) 可以用在量词前面，一起构成数量词组；然后作定语、主语、宾语和谓语等句法成分。

例如：“三个工人、五条毛巾、七种乐器、数幅地图、九吨 (化肥) 太多了、两类 (概念) 正合适、三斤 (桃子) 五块 (钱)、买了半斤 (酱油)、坚持了几年 (时间)、弟弟三岁、钢笔六支、橡皮两块、九块 (钱 (买了)) 两本 (旧书)”。这种分布特征的形式表示是：

__L & __L + NP-Head & __L + Pred & Vt + __L & Sub + __L; 其中，L 代表量词。

其权值设定可以简约地表示如下：

DF₂: __L & __L + NP-Head & __L + Pred & Vt + __L & Sub + __L;

i. √, WV = 20; ii. ×, WV = -20。

(3) 可以用在位数词前面组成系位构造，然后用在量词前面一起构成数量词组；再作定语、主语、宾语和谓语等句法成分。

例如：“三十个工人、五百条毛巾、七十种乐器、数千斤粮食、九千吨 (化肥) 太多了、两千种 (概念) 太烦琐了、三十斤 (桃子) 五十块 (钱)、买了三万斤 (小麦)、坚持了几十年 (工夫)、弟弟三十岁、钢笔六百支、橡皮两百块、九百块 (钱 (买了)) 一本 (线装书)”。这种分布特征的形式表示是：

__W + L & (__W + L) + NP-Head & (__W + L) + Pred & Vt + (__W + L) & Sub + (__W + L); 其中，W 代表位数词。

其权值设定可以简约地表示如下：

DF₃: __W + L & (__W + L) + NP-Head & (__W + L) + Pred & Vt + (__W + L) & Sub + (__W + L); i. √, WV = 20; ii. ×, WV = 0。

(4) 可以用在前缀“第”的后面造成序数组合，然后用在量词前面一起构成数量词组，再去修饰名词性成分；或者，这种序数组合直接修饰名词。

例如：“第八层 (楼)、第三位 (选手)、第一个 (冠军)、第四种 (疫苗)、第三只 (眼睛看中国)、第六排 (观众)、第一小提琴手、第二小提琴手、(在) 第一时间 (作出反映)、(本节目的) 第二焦点、第二生产力、第三推动力、第六感觉、第九食堂、(北京大学) 第三医院、第四会议室”。

这种分布特征的形式表示是：di ^ __ & di ^ __ L + NP-Head & di ^ __ + NP-Head;

其中，di ^ 代表“第”。也可以简洁地表示为：di ^ ^ __ [L] + NP-Head。

其权值设定可以简约地表示如下：

DF₄: $di \sim _ \& di \sim _ L + NP-Head \& di \sim _ + NP-Head$;

i. \checkmark , WV = 20; ii. \times , WV = 0.

(5) 不能直接修饰名词, 除非是作为省略了“第”的序数组合; 换句话说, 直接修饰名词的“系数词+名词”组合前一定可以加上前缀“第”。

例如: “第九食堂~九食堂、第二组~二组、第五区~五区、第三单元~三单元、第六分队~六分队、第八中队~八中队、第九床~九床”。这种分布特征的形式表示是:

$*(_ + NP-Head) \vee \{ (_ + NP-Head) \& (di \sim _ + NP-Head) \}$;

这个表达式的意思是: 要么不能直接修饰名词性成分, 要么既可以前加“第”然后修饰名词性成分、又可以省略“第”而直接修饰名词性成分。因此, 更直接的表达方式是:

$*(_ + NP-Head) \vee \{ (_ + NP-Head) \rightarrow (di \sim _ + NP-Head) \}$;

这个表达式的意思是: 一般不能直接修饰名词性成分, 如果可以直接修饰名词性成分、那么其前面一定可以加“第”。其权值设定可以简约地表示如下:

DF₅: $*(_ + NP-Head) \vee \{ (_ + NP-Head) \& (di \sim _ + NP-Head) \}$;

i. \checkmark , WV = 20; ii. \times , WV = 0.

3. 1. 2 位数词的分布特征和权值设定

(1) 位数词包括“十、百、千、万、亿”等, 它们是粘着词, 不能单独回答问题。

例如: “招了多少学生? ~*千! ~一千!、买了多少斤面粉? ~*万! ~三万!、老爷爷今年多大? ~*百! ~一百!”。这种分布特征的形式表示是: *#_#;

其权值设定可以表示如下: DF₁: *#_#; i. \checkmark , WV = 20; ii. \times , WV = -20.

(2) 一般不能单独用在量词前面, 一起构成数量词组。

例如: “*千个~一千个、*万斤~三万斤、*百岁~一百岁、*亿元~八亿元、*万颗~一万颗”。⁽¹⁾

这种分布特征的形式表示是: *_L; 其权值设定可以简约地表示如下:

DF₂: *_L; i. \checkmark , WV = 10; ii. \times , WV = 0.

(3) 可以用在系数词后面组成系位构造, 然后一起用在量词前面构成数量词组; 然后, 再作定语、主语、宾语和谓语等句法成分。

例如: “四十个农场、五百顶帐篷、七百种民间戏曲、九千吨(粮食)太多了、两千种(树苗)已有销路了、三十斤(桃子)五十块(钱)、买了三万斤(小麦)、坚持了几十年(工夫)、弟弟三十岁、钢笔六百支、橡皮两百块、九百块(钱)一本(线装书)、数万名战士奋战在抗洪抢险第一线”。

这种分布特征的形式表示是: Xi_ + L & (Xi_ + L) + NP-Head & (Xi_ + L) + Pred & Vt + (Xi_ + L) & Sub + (Xi_ + L); 其中, Xi 代表系数词。

其权值设定可以简约地表示如下: DF₃: Xi_ + L & (Xi_ + L) + NP-Head &

(Xi_ + L) + Pred & Vt + (Xi_ + L) & Sub + (Xi_ + L);

i. \checkmark , WV = 20; ii. \times , WV = -20.

(4) 不能用在“第”的后面造成序数组合, 只有跟系数词组成系位构造后才能前加“第”。

例如: “*第百层(楼)~第一百层(楼)、*第千位(选手)~第一千位(选手)、*第万名(志愿着)~第三万名(志愿着)、*第亿种(生物)~第六亿种(生物)”。

这种分布特征的形式表示是： $*di \sim _ \& di \sim Xi_L + NP\text{-Head}$;

其权值设定可以简约地表示如下：

DF₄: $*di \sim _ \& di \sim Xi_L + NP\text{-Head}$; i. \surd , WV = 20; ii. \times , WV = -20。

(5) 不能直接修饰名词。

例如：“*百企业~四百个企业、*千帐篷~两千顶帐篷、*万矿工~八万名矿工、*亿行星~几亿颗行星”。

这种分布特征的形式表示是： $*_NP\text{-Head} \& Xi_L + NP\text{-Head}$;

其权值设定可以简约地表示如下：

DF₅: $*_NP\text{-Head} \& Xi_L + NP\text{-Head}$; i. \surd , WV = 20; ii. \times , WV = 0。

(6) 可用在助词“来、把”之前构成数词组合，然后用在量词前面构成数量词组。

例如：“十来个人、百把条枪、千把张瓦、万把块钱、?亿把颗星星”。

这种分布特征的形式表示是： $_la \triangle i/ba' \& _la \triangle i/ba' + L + NP\text{-Head}$;

其中， $la \triangle i/ba'$ 代表助词“来、把”。其权值设定可以简约地表示如下：

DF₆: $_la \triangle i/ba' \& _la \triangle i/ba' + L + NP\text{-Head}$; i. \surd , WV = 10; ii. \times , WV = 0。

3. 1. 3 由系数词和位数词构成的合成数词

根据朱德熙(1982: 46-47)，由系数词和位数词作为直接成分(immediate constituents)可以造成复杂的数词，具体的规则如下：

(1) 由系数词和位数词两部分组成的复合数词是一种系位构造，系位构造的两部分之间是相乘的关系。例如：

二十 = 2×10 五万 = 5×10000

(2) 几个系位构造按照位数由大到小顺序排列造成的数词结构叫做系位组合。构成系位组合的几个系位构造之间是相加关系。例如：

三千五百 = $(3 \times 1000) + (5 \times 100)$

五万一千九百七十 = $(5 \times 10000) + (1 \times 1000) + (9 \times 100) + (7 \times 10)$

(3) 系位构造是可以递归地生成的，有如下四种表现形式：

(3. 1) 系数部分本身是一个系位构造，例如：二十万、五千万、一百亿；

(3. 2) 系数部分本身是一个并列的数词，例如：两三千、七八万；

(3. 3) 系数部分本身是一个系位组合，例如：二十五万、九千五百零三万；

(3. 4) 位数部分本身是一个系位构造(万万)，例如：四万万(同胞)。

(4) 另外，根据马庆株(1990: 19)，系位组合可以有一种省略形式：由各系位构造的系数组成，即省去各系位构造中的位数。例如：

完全式	简略式
一千八百四十	一八四 0
三百二十七	三二七
四百零八	四 0 八
二十一，二十二……	二一，二二

完全式既可以表示基数，又可以表示序数；比如：“一千八百四十年”也可以是“第一千八百四十年”的意思。一百以内的数词主要用以数数儿，如：“二一、二二、二

三……九八、九九”。数数儿时用的数词也可以看作是一种序数。一百以上的简略式主要用在定语位置上表示序数。例如：

三五九旅 八三四一部队 二六七号牢房 一〇一中学 一九四九年

为了方便，我们把由系数词和位数词组成的复杂数词叫作合成数词；而不仔细区分其中哪些形式可能是词，哪些则可能是词组。⁽²⁾ 合成数词是可以按照规则递归地生成的，它基本上是一种临时性的组合；因此，一般的词典是不会列作词条的。由于合成数词的分布是有很强的规律的，因而下面我们还是要讨论一下它的分布特征和权值设定；尽管人们识别合成数词是一件极为容易的事，并且一般不会产生分歧性的意见。为了简单，我们只讨论完全式的合成数词，而不讨论简略式的合成数词。

3. 1. 4 合成数词的分布特征和权值设定

(1) 合成数词包括由系数词和位数词组成的系位构造和系位组合等复合形式。它们不是完全的粘着词、但也不是完全的自由词；即它们有时具有一定的自由活动的的能力，可以用来单独回答问题。

例如：“来了多少学生？~?一百二十六!、买了多少斤面粉？~?三千六百八十九!、这孩子几岁？~十二!、您老今年多大？~六十八!、中国有多少人口？~十二亿!”。这种分布特征的形式表示是：? #_#；其中，?表示该格式的合格性不明确。

其权值设定可以表示如下：DF₁: ? #_#；i. √, WV = 10; ii. ×, WV = 0。

(2) 可以用在量词前面，一起构成数量词组；然后作定语、主语、宾语和谓语等句法成分。

例如：“十三个工人、数十年时间、五百条毛巾、二十七种乐器、九百六十五吨（化肥）太多了、一二十类（概念）正合适、三十斤（桃子）五十块（钱）、买了一百八十六斤（南瓜）、坚持了几十年（时间）、弟弟二十三岁、钢笔六十六支、橡皮两百零八块、九千块（钱）二十本（旧书）”。这种分布特征的形式表示是：

_L & _L + NP-Head & _L + Pred & Vt + _L & Sub + _L;

其权值设定可以简约地表示如下：

DF₂: _L & _L + NP-Head & _L + Pred & Vt + _L & Sub + _L;

i. √, WV = 20; ii. ×, WV = -20。

(3) 可以用在前缀“第”的后面造成序数组组合，然后用在量词前面一起构成数量词组，再去修饰名词性成分；或者，这种序数组组合直接修饰名词。

例如：“第十八层（楼）、第三十位（选手）、第一百零一个（冠军）、第五十四种（疫苗）、第二十三类（建筑风格）、第三十六排（观众），（康熙的）第十四子、（《水浒》中的）第一百零八将（是地狗星金毛犬段景住）、第二十六楼、第十二小组、第二十九食堂、第十三医院、第二十六病床、第十四会议室”。这种分布特征的形式表示是： $di^{\sim} _ & di^{\sim} _ L + NP-Head & di^{\sim} _ + NP-Head$;

也可以简洁地表示为： $di^{\sim} _ [L] + NP-Head$ 。其权值设定可以简约地表示如下：

DF₃: $di^{\sim} _ & di^{\sim} _ L + NP-Head & di^{\sim} _ + NP-Head$;

i. √, WV = 20; ii. ×, WV = -20。

(4) 不能直接修饰名词，除非是作为省略了“第”的序数组组合。换句话说，直接修饰名词的“合成数词+名词”组合前一定可以加上前缀“第”。

例如：“第十二食堂~十二食堂、第二十组~二十组、第五十六区~五十六区、第十六分队~十六分队、第十五宿舍~十五宿舍、第十八赛区~十八赛区”。这种分布特

征的形式表示是:

$*(_ + \text{NP-Head}) \vee \{ (_ + \text{NP-Head}) \& (di \wedge _ + \text{NP-Head}) \}$; 或者:
 $*(_ + \text{NP-Head}) \vee \{ (_ + \text{NP-Head}) \rightarrow (di \wedge _ + \text{NP-Head}) \}$;

其权值设定可以简约地表示如下:

DF₄: $*(_ + \text{NP-Head}) \vee \{ (_ + \text{NP-Head}) \& (di \wedge _ + \text{NP-Head}) \}$;
 i. \checkmark , WV = 20; ii. \times , WV = 0。

(5) 合成数词的末位, 要么是位数词“十、百、千、万、亿”等; 这时可以加上助词“来、多、余”, 从而造成一个表示概数的词组; 要么是系数词“一~九、几”等, 这时不能加上助词“来、多、余”。⁽³⁾

例如: “十二(个)~*十二来(个)、三十(个)~三十来(个)、四十二(张)~*四十二多(张)、四十(张)~四十多(张)、五百六十三(件)~*五百六十三余(件)、五百(件)~五百余(件)”。此外, 末位是位数词的合成数词不能后附助词“把”。

例如: “*十把个人~*二十把个人、百把块钱~*三百把块钱、千把斤大米~*两千把斤大米、万把条鱼苗~*几万把条鱼苗”。末位是系数词的合成数词后面可以加上位数词“万”、但是不能加上位数词“千”; 而系数词后面既可以加上位数词“千”, 又可以加上位数词“万”。例如: “三千~三万~*十三千~十三万、六千~六万~*八十六千~八十六万、九千~九万~*四千零九千~四千零九万”。因此, 可这样来概括合成数词的这种分布特点: 或者可以出现在助词“来、多、余”之前, 并且不能出现在助词“把”之前; 或者可出现在位数词“万”之前, 但不能出现在位数词“千”之前。

这种分布特征的形式表示是: $(_ -la \triangleq i/duo \backslash yu \triangleq \& * _ -ba \cdot) \vee (_ -wa \wedge n \& * _ -qia \wedge n)$; 其中, $yu \triangleq$ 代表“余”, $qia \wedge n$ 和 $wa \wedge n$ 分别代表位数词“千”和“万”。

其权值设定可以简约地表示如下:

DF₅: $(_ -la \triangleq i/duo \backslash yu \triangleq \& * _ -ba \cdot) \vee (_ -wa \wedge n \& * _ -qia \wedge n)$;
 i. \checkmark , WV = 30; ii. \times , WV = 0。

3. 1. 5 数词的分布特征和权值设定

(1) 可以单独用在量词前面或跟其他数词组成合成数词用在量词前面, 一起构成量词组。这种分布特征的形式表示是: $_ L \vee \dots _ \dots + L$; 其中, “ $\dots _ \dots$ ”代表合成数词。其权值设定可以简约地表示如下:

DF₁: $_ L \vee \dots _ \dots + L$; i. \checkmark , WV = 30; ii. \times , WV = -30。

(2) 不能直接修饰名词, 除非是作为省略了“第”的序数组合; 换句话说, 直接修饰名词的“数词+名词”组合前一定可以加上前缀“第”。这种分布特征的形式表示是:

$*(_ + \text{NP-Head}) \vee \{ (_ + \text{NP-Head}) \& (di \wedge _ + \text{NP-Head}) \}$; 或者:

$*(_ + \text{NP-Head}) \vee \{ (_ + \text{NP-Head}) \rightarrow (di \wedge _ + \text{NP-Head}) \}$;

其权值设定可以简约地表示如下:

DF₂: $*(_ + \text{NP-Head}) \vee \{ (_ + \text{NP-Head}) \& (di \wedge _ + \text{NP-Head}) \}$;
 i. \checkmark , WV = 30; ii. \times , WV = 0。

(3) 在书面语中, 可以抛开量词直接用在名词后面表示序数意义。例如: “图三~第三(个)图、表五~第五(个)表、例六~第六(个)例(子)、册二~第二(个分)册、卷三~第三(个分)卷, 步骤一~第一个步骤、参考题二~第二道参考题、附录

三~第三个附录、公理一~第一个公理、注十一~第十一条注释；封三~封之三/封面的第三面、页十五~页之十五/书页的第十五面、参宿四~参宿之四/参宿的第四颗(星)、织女二~织女之二/织女(座)的第二颗(星)”。⁽⁴⁾ 这种分布特征的形式表示是： $N_ \& \{(di \wedge _ (L)N) \vee (N-zhi \ ` + _) \vee (N-de + di \wedge _ L)\}$;

其中，*zhi* 代表文言虚词“之”。其权值设定可以简约地表示如下：

DF₃: $N_ \& \{(di \wedge _ (L)N) \vee (N-zhi \ ` _) \vee (N-de + di \wedge _ L)\}$;

i. \checkmark , WV = 20; ii. \times , WV = 0。

(4) 一般不能用来单独回答问题。这种分布特征的形式表示是： $*\# _ \#$;

其权值设定可以简约地表示如下：DF₄: $*\# _ \#$; i. \checkmark , WV = 10; ii. \times , WV = 0。

(5) 一般不能作主语、宾语和谓语，除非是算术表达式⁽⁵⁾或流水帐式的列举等特殊格式中。

例如：“一加一等于二、十个十是一百；昨天五百，今天八百，两天下来赚了一千三百”。这种分布特征的形式表示是： $*_ Pred \& *Vt _ \& *Sub _$;

其权值设定可以简约地表示如下：

DF₅: $*_ Pred \& *Vt _ \& *Sub _$; i. \checkmark , WV = 10; ii. \times , WV = 0。

3. 2 量词的分布特征和权值设定

量词先得分成简单量词和复合量词两类来分别进行讨论，因为它们的分布特征有相当的差别；然后，才能抽取两者共同的分布特征，来描写整个量词的分布特征，并设定其权值。简单量词指“次、架、辆、起、场、秒、小时、千瓦、立方米”等能单独用在数词后面构成数量词组的单位词，之所以不叫它们为单纯量词，是因为像“小时、千瓦、立方米”等是由多个语素构成的合成词。复合量词指由简单量词跟简单量词、或简单量词跟量词性语素组合构成的复杂量词，例如：“人次、架次、人起、场次、机时、车公里、千米小时、焦耳每秒、牛顿每平方米、弧度每2次方秒、千克每立方米”等等。⁽⁶⁾

3. 2. 1 简单量词的分布特征和权值设定

(1) 简单量词是粘着词，不能用来单独回答问题。例如：“来了多少学生？~*个!、买了多少面粉？~*斤!、这孩子多大？~*岁!、你参加了多少足球赛？~*场!”。

这种分布特征的形式表示是： $*\# _ \#$ ；其权值设定可以简约地表示如下：

DF₁: $*\# _ \#$; i. \checkmark , WV = 20; ii. \times , WV = -20。

(2) 可以直接用在数词(包括简单数词和合成数词)之后构成数量词组，然后作定语修饰名词性成分或作动词性成分的准宾语。

例如：“三十枝(枪)、两根(雪糕)、四十一个(客人)、十六双(筷子)、几千亩(良田)、八万公斤(大豆)、一点儿(麻烦)、一些(问题)、(打)一下、(去)两次、(走)一趟、(赛)一场、(读)几遍”。

这种分布特征的形式表示是： $S _ + NP \vee VP + S _$ ；其中，S代表数词。

其权值设定可以简约地表示如下：

(3) 可以直接用在由位数词和数量助词(包括前助词“上”和后助词“来、把”)构成的数词短语之后，一起构成数量词组。例如：“上百个、上万个、十来场、百来斤、万把条、万来吨、千把次”。这种分布特征的形式表示是：

$sha \hat{ng}-W _ \vee W-la \hat{a}/ba \hat{a}$; 其中, $sha \hat{ng}$ 代表前助词“上”。

其权值设定可以简约地表示如下:

DF₃: $sha \hat{ng}-W _ \vee W-la \hat{a}/ba \hat{a}$; i. \checkmark , WV = 20; ii. \times , WV = -20。

(4) 可以前加“每”来表示全称 (universal, 或称“周遍性”), 再去修饰名词性成分、或独立指称事物; 并且, 在“每”和量词之间可以自由地插入或删除数词“一”。例如: “每(一) 个人、每(一) 条船、每(一) 户人家、每(一) 批钢材、每(一) 尺棉布、每(一) 点儿成绩、? 每(一) 些问题、每(一) 次呼吸、每(一) 下撞击、每(一) 趟考试”。这种分布特征的形式表示是: $me \hat{i}-(yi \hat{a}) _$; 其中, $me \hat{i}$ 代表“每”, $yi \hat{a}$ 代表数词“一”。其权值设定可以简约地表示如下: DF₄: $me \hat{i}-(yi \hat{a}) _$; i. \checkmark , WV = 10; ii. \times , WV = 0。

(5) 可以通过重叠来表示全称, 再去修饰名词性成分、或独立指称事物; 有时, 还可以在这种重叠式之前加上数词“一”。

例如: “(一) 个个赛手、(一) 张张笑脸、(一) 朵朵白云、(一) 双双大手、(一) 群群绵羊、(一) 亩亩水稻、(一) 次次表决、(一) 场场竞赛、(一) 阵阵欢呼”。这种分布特征的形式表示是: $(yi \hat{a}) _ _$; 其权值设定可以简约地表示如下: DF₅: $(yi \hat{a}) _ _$; i. \checkmark , WV = 10; ii. \times , WV = 0。

(6) 可以用在代词“这、那”和名词性成分之间, 并且“这、那”和量词之间可以自由地插入或删除数词“一”。

例如: “这(一) 张纸、那(一) 个家伙、这(一) 套设备、那(一) 斗稻子、这(一) 点儿报酬、那(一) 些司机、这(一) 次旅行、那(一) 场电影”。这种分布特征的形式表示是: $zhe \hat{a}/na \hat{a} (yi \hat{a}) _ + NP-Head$; 其中, $zhe \hat{a}/na \hat{a}$ 代表代词“这”和“那”。其权值设定可以简约地表示如下:

DF₆: $zhe \hat{a}/na \hat{a} (yi \hat{a}) _ + NP-Head$; i. \checkmark , WV = 10; ii. \times , WV = 0。

(7) 可以后附名词性语素“数”构成临时性的复合名词。例如: “(西瓜的) 个数、(海报的) 张数、(耕牛的) 头数、(子弹的) 颗数、(药品的) 批数、(预赛的) 组数、(舞伴的) 对数、(鲜鱼的) 斤数、(化肥的) 吨数、*(报酬的) 点儿数、*(司机的) 些儿数、(出勤的) 次数、(出诊的) 趟数、(演出的) 场数”。或者用在动词“论”的后面作宾语, 构成粘着性的述宾短语。例如: “这些书论本卖、这些药丸论颗卖、救灾物质论批统计、黄沙论吨出售、*困难论点统计、*纸张论些计算、出勤论次统计、比赛论场计算”。

这种分布特征的形式表示是: $_ -shu \hat{a} \vee lu \hat{n} _$; 其中, $shu \hat{a}$ 代表语素“数”, $lu \hat{n}$ 代表动词“论”。

其权值设定可以简约地表示如下:

DF₇: $_ -shu \hat{a} \vee lu \hat{n} _$; i. \checkmark , WV = 10; ii. \times , WV = 0。

3. 2. 2 复合量词的分布特征和权值设定

(1) 复合量词也是粘着词, 不能用来单独回答问题。

例如: “举办了多少训练班? ~ *班次! ~60 班次!、查询了多少档案? ~*卷次! ~2600 卷次!、出动了多少飞机? ~*架次! ~38 架次!、你们参加了多少足球赛? ~*

场次! ~80 场次!”。

这种分布特征的形式表示是: *#_#; 其权值设定可以简约地表示如下:

DF₁: *#_#; i. √, WV = 20; ii. ×, WV = -20。

(2) 可以直接用在数词(包括简单数词和合成数词)之后构成数量词组, 然后主要作动词性成分(包括述宾短语)的准宾语, 也可以作定语修饰名词性成分。

例如: “出动的车辆的总数达 682 辆次、作业面积达 13 亿亩次、做好事达 1000 多件次、我国器官移植总数已超过 5 千例次, 输出长片 105 部次、抽查检验药品 60 多万批次、查出倒卖服装 12 人起、取送影片 300 场次; 发射了 10 架次航天飞机、出动 300 多辆次汽车、有 20 部次电影获奖、先后为 85 艘次舰艇进行修理”。⁽⁷⁾ 这种分布特征的形式表示是:

VP + S__ √ S__ + NP; 其权值设定可以简约地表示如下:

DF₂: VP + S__ √ S__ + NP; i. √, WV = 20; ii. ×, WV = -20。

(3) 可以直接用在由位数词和数量助词(包括前助词“上”和后助词“来、把”)构成的数词短语之后, 一起构成数量词组。

例如: “上百人次、上万卷次、十来场次、百来部次、万把件次、千把架次”。这种分布特征的形式表示是:

sha^ng-W__ √ W-la≡i/ba^__; 其权值设定可以简约地表示如下:

DF₃: sha^ng-W__ √ W-la≡i/ba^__; i. √, WV = 20; ii. ×, WV = 0。

(4) 一般不能前加“每(一)”来表示全称。

例如: “*每(一)人次、*每(一)艘次、*每(一)户次、*每(一)批次、*每(一)场次、*每(一)人起、*每(一)机时、*每(一)架次、*每(一)例次”。这种分布特征的形式表示是: *me^i-(yi^)^__;

其权值设定可以简约地表示如下: DF₄: *me^i-(yi^)^__; i. √, WV = 10; ii. ×, WV = 0。

(5) 不能通过重叠来表示全称, 再去修饰名词性成分、或独立指称事物。例如:

“*(一)人次人次、*(一)架次架次、*(一)卷次卷次、*(一)批次批次、*(一)场次场次、*(一)人起人起、*(一)例次例次”。这种分布特征的形式表示是: *(yi^)^__ __; 其权值设定可以简约地表示如下: DF₅: *(yi^)^__ __; i. √, WV = 10; ii. ×, WV = -10。

(6) 不能用在代词“这(一)、那(一)”和名词性成分之间。例如: “*这(一)人次、*那(一)架次、*这(一)卷次、*那(一)批次、*这(一)场次、*那(一)人起、*这(一)机时、*那(一)例次”。这种分布特征的形式表示是: *zhe^/na^^(yi^)^__ + NP-Head;

其权值设定可以简约地表示如下:

DF₆: *zhe^/na^^(yi^)^__ + NP-Head; i. √, WV = 10; ii. ×, WV = 0。

(7) 不能后附名词性语素“数”构成临时性的复合名词。

例如: “*(出席的)人次数、*(出动飞机的)架次数、*(查阅档案的)卷次数、*(运送救灾物资的)批次数、*(手术的)例次数、*(预赛的)场次数、*(查处的)人起数、*(实验用的)机时数、*(获奖电影的)部次数、*(举办学习班的)班次数”。一般不能用在动词“论”的后面作宾语, 构成粘着性的述宾短语。例如: “*学

习班论班次统计、*这类手术论人例计算、*救灾物质论批次统计、*获奖电影论部次统计、*观摩比赛论场次计算、*出动的飞机论架次统计”。这种分布特征的形式表示是： $*_shu^{\wedge} \& *lu^{\wedge}n_;$ 其权值设定可以简约地表示如下：

DF₇: $*_shu^{\wedge} \& *lu^{\wedge}n_;$ i. \surd , WV = 10; ii. \times , WV = 0。

3. 2. 3 量词的分布特征和权值设定

(1) 量词是粘着词，不能用来单独回答问题。这种分布特征的形式表示是： $*\#_#;$ 其权值设定可以简约地表示如下：DF₁: $*\#_#;$ i. \surd , WV = 20; ii. \times , WV = -20。

(2) 可以直接用在数词（包括简单数词和合成数词）之后构成数量词组，然后作定语修饰名词性成分或作动词性成分的准宾语。这种分布特征的形式表示是：

$S_ + NP \surd VP + S_;$ 其权值设定可以简约地表示如下：

DF₂: $S_ + NP \surd VP + S_;$ i. \surd , WV = 20; ii. \times , WV = -20。

(3) 可以直接用在由位数词和数量助词（包括前助词“上”和后助词“来、把”）构成的数词短语之后，一起构成数量词组。这种分布特征的形式表示是：

$sha^{\wedge}ng-W_ \surd W-la^{\triangle}i/ba^{\cdot}_;$ 其权值设定可以简约地表示如下：

DF₃: $sha^{\wedge}ng-W_ \surd W-la^{\triangle}i/ba^{\cdot}_;$ i. \surd , WV = 20; ii. \times , WV = 0。

(4) 可以直接用在由合成数词和数量助词“多、余”构成的数词短语之后，一起构成数量词组。

例如：“二十多架~二十多次~二十多架次、三百六十余卷~三百六十余次~三百六十余卷次、八千九百五十多人起、三千0五十余场次”。这种分布特征的形式表示是：

$S-duo^{\wedge}yu^{\triangle}_;$ 其权值设定可以简约地表示如下：

DF₄: $S-duo^{\wedge}yu^{\triangle}_;$ i. \surd , WV = 20; ii. \times , WV = -20。

(5) 不能作主语、谓语等句法成分。这种分布特征的形式表示是： $*_Pred \& *Sub_;$ 其权值设定可以简约地表示如下：

DF₅: $*_Pred \& *Sub_;$ i. \surd , WV = 20; ii. \times , WV = -20。

3. 3 准数量词的分布特征和权值设定

3. 3. 1 真数量词和准数量词

根据袁毓林（1998b: 262-5）真正由数词和量词构成的词组叫真数量词，简称数量词。真数量词从构造上可以分为下列三种：

(一) 组合式数量词。例如：

一张、两件、三十个、四十八条、一百六十九架次……

一斤、两钱、三十吨、四十八里、一百六十九牛顿米……

一批、两番、三十趟、四十八次、一千六百五十九场次……

这种组合式数量词是不可列举的，它们兼具体词和谓词的某些主要的分布特征；可以作主语、宾语和定语，还可以作谓语。⁽⁸⁾ 例如：

一斤卖两块 买了三张(纸) 一批旅客 黄瓜两斤，白菜六斤

(二) 粘合式数量词。例如：

半壁(江山) 半天(时间) 一阵(骚动) 半截(烂了)
半拉(糊了) 一下(也行) (忙了)半晌 (去)一会儿

这种粘合式数量词是可以列举的,它们具有体词性;主要作定语,也可以作主语和宾语,一般不作谓语。它们跟名词的主要差别是,一般不受数量词的修饰。

(三)融合式数量词。例如:

俩(人) (我买)俩 (咱们)俩
仨(人) (你拿)仨 (哥儿)仨

融合式数量词只有“俩、仨”两个。它们具有体词性,通常只作定语、宾语和同位语。

准数量词指那些不是由“数词+量词”构成的、但是功能跟真数量词(特别是其中的粘合式数量词)相似的合成词。这种准数量词,常用的有:“许多、好多、很多、?不少、大量、一切、种种、所有、全部、整个、部分、少数、有的、许久、片刻、一半”等。⁽⁹⁾这种准数量词的主要的语法功能是作定语、也可以作主语和宾语(所以不是区别词),但是不受“很”等程度副词的修饰(所以不是形容词)、也不受形容词和数量词等形式类的修饰(所以不是名词)。

3.3.2 准数量词的分布特征和权值设定

(1)可以作定语修饰名词性成分,但是不受形容词和副词“很”修饰。

例如:“许多学生(走了)、(来了)好多老师、(种了)很多蔬菜、不少农田(被淹了)、(进口了)大量洋货、一切权力(归农会)、(克服了)种种困难、所有住户(都得交卫生费)、(变卖了)全部家产、(影响到)整个战局、部分群众(不理解这一政策)、少数用户(拖欠月租费)、有的干部(收受贿赂)、?等待了许久时间、等待了片刻工夫、(有)一半时间(在国外)”。

这种分布特征的形式表示是: __NP-Head & *A-Det / he'n __;

其权值设定可以简约地表示如下:

DF₁: __NP-Head & *A-Det / he'n __; i. √, WV = 20; ii. ×, WV = -20。

(2)可以通过转指来代替整个偏正结构,从而作主语或宾语;或者直接作动词(包括及物动词和不及物动词)或形容词的准宾语(表时量)。⁽¹⁰⁾

例如:“许多(学生)走了、来了好多(老师)、种了很多(蔬菜)、不少(农田)被淹了、?进口了大量(洋货)、一切(问题)好商量、?克服了种种(困难)、?所有(住户)都得交卫生费、全部(家产)都在这儿、整个(工厂)都盘出去了、只留下部分(日记)、少数(人)服从多数(人)、有的(干部)收受贿赂、?等待了许久(时间)~走红了许久、?等待了片刻(工夫)~稍候片刻、(一年有)一半(时间)在国外”。这种分布特征的形式表示是:

__Pred √ Vt__ √ V/A + __; 其中,“V/A + __”表示可以作动词或形容词的准宾语。其权值设定可以简约地表示如下:

DF₂: __Pred √ Vt__ √ V/A + __; i. √, WV = 20; ii. ×, WV = 0。

(3)作定语时可以出现在领属定语之后、带“的”的描写性定语之前。⁽¹¹⁾

例如:“我国许多优秀的人才都出国了、郊区不少肥沃的农田被工厂占用、日本大量高档的家电产品倾销到中国、揭露发生在这儿的种种欺上瞒下的行为、变卖了我家全部值钱一点儿的東西、我们少数意志薄弱的干部被腐蚀了、美国有的著名的上市公司

编造假账”。

这种分布特征的形式表示是: Poss-Det + __ + Det-*de* + NP-Head;

其中, Poss-Det 代表领属定语, Det-*de* 代表带“的”的描写性定语;

其权值设定可以简约地表示如下:

DF₃: Poss-Det + __ + Det-*de* + NP-Head; i. √, WV = 10; ii. ×, WV = 0。

(4) 不能作谓语或谓语核心, 不能受状语和补语修饰。这种分布特征的形式表示是:

*Sub__ (+ Obj) & *Ad__ & *__Comp; 其中 Ad 代表状语, Comp 代表补语。

其权值设定可以简约地表示如下:

DF₄: *Sub__ (+ Obj) & *Ad__ & *__Comp; i. √, WV = 10; ii. ×, WV = -10。

(5) 不能受真数量词修饰, 也不能修饰真数量词。这种分布特征的形式表示是:

*SL__ & *__SL; 其权值设定可以简约地表示如下:

DF₅: *SL__ & *__SL; i. √, WV = 10; ii. ×, WV = -10。

(6) 不能用在系数词之后或位数词之前, 一起构成合成数词。

这种分布特征的形式表示是: *Xi__ & *__W; 其权值设定可以简约地表示如下:

DF₆: *Xi__ & *__W; i. √, WV = 10; ii. ×, WV = -10。

(7) 不能用在前置“第”之后, 一起构成序数组合。这种分布特征的形式表示是:

*Di-__ ; 其权值设定可以简约地表示如下:

DF₇: *Di-__ ; i. √, WV = 10; ii. ×, WV = -10。

(8) 不能用在“上、几、数”之后或“来、把、多、余”之前, 一起构成表示概数的复合形式。这种分布特征的形式表示是:

*sha[^]ng/ji[^]/shu[^]__ & *__la[△]i/ba[^]/duo[^]/yu[△];

其权值设定可以简约地表示如下:

DF₈: *sha[^]ng/ji[^]/shu[^]__ & *__la[△]i/ba[^]/duo[^]/yu[△]; i. √, WV = 10; ii. ×, WV = -10。

4. 结语: 怎样运用这套量表来进行模糊分类

根据上述量表进行计算, 如果某词相对于甲种词类的隶属度高于其他词类, 那么可以把它归入甲种词类; 如果某词相对于甲乙两种词类的隶属度相同或相近、并且高于其他词类, 那么可以把它处理为兼属甲乙两种词类。如此推进, 基本上可用以对现代汉语有关的词进行词类判定。其中, 最理想的结果是: 根据这套分布特征及其权值设定, 使得现代汉语中有关的词不仅能划归到某一个词类之中 (即确定它到底属于什么词类), 而且能显示出它从属于这一词类的程度到底有多大 (即它跟相关的典型成员的差别有多大)。

附注:

- (1) 像“百岁老人、长命百岁、万元户、亿元村”等是固定组合, 其中的“百岁、万元、亿元”等不具有可类推的性质。
- (2) 郭锐 (2002) § 8.5.8 把系位构造、系位组合、前加“第”的序数组合、分数和小数表达式等, 都看作是数词词组 (第 219—20 页)。
- (3) (5) 参考郭锐 (2002) § 8.5.8, 第 219—22 页。
- (4) 详见马庆株 (1990) § 1.3.2, 第 21 页。
- (6) 详见马庆株 (1990) § 2.3, 第 25—26 页; 张万起 (1992) § 1, 第 145 页。至于, 像张万起 (1992: 154) 所举的下列例子:

- a) (我厂) 年产服装 60 万件套。(光明日报, 1990, 5, 13/4)
- b) 从 1970 年到 1985 年, 荣获各种锦旗、奖状 150 面个。(解放军报, 1988, 7, 20/2)
- c) 8 万平方米的建筑, 动力和工艺设备约 7 万台、套、件, 工艺管道 101 公里。(文艺报, 1990, 7, 12/3)

我们认为, 其中的“件套, 面个, 台、套、件”可以看作是由量词并列构成的联合词组。因为, “件套”的意思是“件或套”。

(7) 其中的例子主要来自张万起(1992), 并作了一定的精简处理。

(8) 详见朱德熙(1982), § 4.13, 第 51—52 页。

(9) 参考郭锐(2002) § 8.5.9.1, 第 222 页。注意, “不少”可以直接修饰名词, 而一般的形容词受“不”修饰后不能直接修饰名词。例如:

大房子 ~ 不大的房子 ~ *不大房子 *少房间 ~ *不少的房间 ~ 不少房间

据此, 可以说可以直接修饰名词的“不少(=许多)”是词; 但是, 当“不少(=多)”是词组时, 可以作谓语、可以受副词修饰。例如:

部门不大, 问题不少 ~ 这儿的问题真不少 ~ 这儿的问题还确实不少

(10) 详见朱德熙(1982), § 5.1.3, 第 56 页和 § 4.13.2, 第 51 页; 郭锐(2002) § 8.5.9.2, 第 223—4 页。

(11) 参考朱德熙(1982), § 10.9, 第 150—151 页; 郭锐(2002) § 8.5.9.1, 第 223 页。

参考文献

- 郭锐 2002 《现代汉语词类研究》, 北京: 商务印书馆。
- 马庆株 1990 《数词、量词的语义成分和数量结构的语法功能》, 《中国语文》第 3 期; 收入马庆株(1998)《汉语语义语法范畴问题》, 北京: 北京语言文化大学出版社, 本文据此。
- 袁毓林 1995 《词类范畴的家族相似性》, 《中国社会科学》第 1 期。收入袁毓林(1998a), 第 221—246 页。
- 袁毓林 1998a 《语言的认知研究和计算分析》, 北京: 北京大学出版社。
- 袁毓林 1998b 《基于原型的汉语词类分析》, 收入袁毓林(1998a), 第 247—269 页。
- 袁毓林 2000 《一个汉语词类的准公理系统》, 《语言研究》第 4 期。
- 袁毓林 2005a 《基于隶属度的汉语词类的模糊划分》, 《中国社会科学》第 1 期。
- 袁毓林 2005b 《现代汉语虚词模糊划分的隶属度量表》, 《汉语学报》第 4 期。
- 张万起 1992 《论现代汉语复合量词》, 《语法研究和探索》(六), 北京: 语文出版社。
- 朱德熙 1982 《语法讲义》, 北京: 商务印书馆。
- Bloomfield, L., 1926, A Set of Postulates for the Science of Language, *Language* 2, pp. 153-64; in Joos (1958) (ed), pp. 26-31.
- Chao Yuen Ren, 1968, A Grammar of Spoken Chinese, University of California Press. 据台湾版, 敦煌书局, 1981 年。丁邦新 全译本《中国话的文法》, 香港中

- 文大学出版社，1980年，据刘梦溪主编《中国现代学术经典·赵元任卷》，胡明扬 王启龙 编校，石家庄：河北教育出版社，1996年。
- 吕叔湘节译本《汉语口语语法》，北京：商务印书馆，1979年。
- Harris, Z. S., 1946, From Morpheme to Utterance, *Language* 22, pp. 161-83; in Joos (1958) (ed), pp. 142-53.
- Joos, M., 1958, (ed.) *Readings in Linguistics: the Development of Descriptive Linguistics in America since 1925*, American Council of Learned Societies.
- Wells, R. S., 1947, Immediate Constituents, *Language* 23, pp. 81-117; in Joos (1958) (ed), pp. 186-207.

Scales for the Degree of Membership of Numeral and Classifier

Yuan Yu-lin

Dept. of Chinese, Peking University, Beijing 100871

Abstract

This paper selects a set of distribution features of the typical members of numeral and classifier, and assigns different weight value to them according their importance to the related word-class. Then, it checks and corrects these weight values on the ground of the grammatical behaviour of the non-typical members of numeral and classifier, and constructs a set of distribution scales for measuring the degree of membership of these word-classes in the fuzzy classification of Mandarin words. Finally, it demonstrates how to use these scales to determine the word-class of words with complicated grammatical function.

Keywords

word-class classification, typical members, distribution features, weight value setting, fuzzy classification, degree of membership