

概数助词“来”的形式句法语义界面研究



宁娜*

广东外语外贸大学英语语言文化学院, 广东广州 510420

摘要: 作为人类自然语言的子系统, 数词不仅是窥探一个民族思维和文化的重要窗口, 同时, 因其特殊的形态句法语义特征, 更是备受语言学家们的关注。概数助词作为数词系统的有机组成部分, 在汉语中使用频率高, 用法纷繁复杂, 与之相关的研究成果丰硕, 但也争议不断。本文在形式语言学的理论框架下, 主要探讨了现代汉语中概数助词“来”的句法分布和语义诠释。句法上, “来”作为黏着词素, 不管其是黏附在数词后, 还是黏附在数量词或量词后, 都是作用于前面的整个短语, 但语义上却只与其临近的位数词或者量词发生关联。语义上, 本文确定了“来”前为数词的语义取值为 $\lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)]$, 不仅通过了各种数值的验证, 并且给出了充分的解释, 同时, 以类推的方式证明了该语义模型也符合“来”后为数量词或量词的情况。本文通过对“来”的句法语义界面分析, 可深化人们对整个数词系统的认识, 从而促进人类自然语言本体的研究, 同时对汉外教学和人工智能的发展也有一定的助推作用。

关键词: 概数助词; “来”; 形式句法; 形式语义; 类推

DOI: [10.57237/j.cll.2023.02.003](https://doi.org/10.57237/j.cll.2023.02.003)

A Study on the Formal Syntax-Semantics Interface of the Particle for Numerical Approximation *Lai*

Ning Na*

Faculty of English Language and Culture, Guangdong University of Foreign Language Studies, Guangzhou 510420, China

Abstract: As the subsystem of human natural language, numeral is not only an important window to pry into the mindset and culture of a nation, meanwhile, due to its specialty in morphological, syntax and semantics, it has always been drawing the high attention of linguists. As the organic part of numeral system, particles for numerical approximation are frequently used in Mandarin Chinese, and their usage is rather complicated. There have been fruitful achievements about the research on such linguistic phenomenon, but controversies have also never ceased. Under the theoretical framework of formal linguistics, this paper mainly explores the syntactic distribution and semantic interpretation of the particle for numeral approximation *lai* in modern Mandarin Chinese. It is demonstrated that syntactically, whether suffixed behind numerals, numeral-classifier, or just classifier, *lai*, as a bound morpheme, always functions upon the whole phrase ahead, but it just semantically associates with its immediately adjacent digit word or classifier. Semantically, this paper ascertains the value

基金项目: 广东外语外贸大学研究生科研创新项目《汉语概数助词的句法语义界面研究》(22GWCXXM-004);

广东外语外贸大学研究生科研创新项目《语言哲学视角下的时间问题及其当代语言学意蕴》(22GWCXXM-010).

*通信作者: 宁娜, 1402575544@qq.com

收稿日期: 2022-09-28; 接受日期: 2023-03-08; 在线出版日期: 2023-05-25

<http://www.chlanglit.com>

of *lai* as $\lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)]$ when it follows a numeral, which is evidenced by various kinds of numerals and adequate explanation, and in the meantime, it verifies that the aforementioned semantic model is also in accordance with the case when *lai* suffixes a numeral-classifier or classifier by analogy. Through the syntactic-semantic interface analysis on the particle for numeral approximation *lai*, this research is likely to deepen people's understanding about the whole numeral system, thereby facilitating the study on the ontology of human natural language, and at the meantime, it may play a positive role in the enhancement of Chinese-foreign language teaching and artificial intelligence.

Keywords: Particle for Numerical Approximation; *lai*; Formal Syntax; Formal Semantics; Analogy

1 引言

古希腊著名数学家毕达哥拉斯曾说：“数支配着宇宙”。不同的计数方式不仅反映一种语言的数学思维，更是透视一个民族心理和认知的重要窗口。自然语言数词系统由于其本身形态的丰富性、句法的独特性和语义的完备性，一直是语言本体研究的核心议题，也是众多跨学科如计算机、人类学、数学史、哲学和心理学的热点话题。概数助词作为数词系统的重要组成部分，使用频率高，用法错综复杂，对其句法分布和语义诠释进行客观地描写和充分地解释，不仅有助于深化人们对数词系统的认识，有益于推动语言本体的研究，对其它跨学科研究也具有相当重要的启示和借鉴意义。

“来”，作为汉语中最常用的概数助词之一，主要黏附在数词、数量词或量词之后。回顾以往文献可以发现，很多学者曾试图从不同方面对“来”进行过剖析。吕淑湘（1957）[1]指出，“来”应该是由“以来”经虚化演变而来，江蓝生（1984）[2]、王庆（2017）[3]分别从历时角度对“来”的语法化演化动因和机制进行了深入探析。句法上，吕叔湘（1999a: 209-216）[4]在《数量词后的来、多、半》一文中，比较了汉语中三个概数助词“来、多、半”的句法位置，并总结了如下的公式：

- (1) a...0, ...00, ...000, ...0000+来+类别词/度量单位+名词
 b₁. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+度量单位+来+名
 b₂...1, ...2, ...3, ...4, ...5, ...6, ...7, ...8, ...9+度量单位+来+名
 c₁...0, ...00, ...000, ...0000+来+类别词/度量单位+‘重, 长, 大, 厚’
 c₂.(...)1, (...)2,, (...)9+度量单位+来+‘重, 长, 大, 厚’

从上可知，位于“来”前面的要么是整数，要么是数量词或量词。在此基础上，杨德峰（1993）[5]、张谊

生（2001）[6]、杜晓艺（2006）[7]、李柔冰（2011）[8]、应学凤和王晓辉（2014）[9]等分别就“来”和其它概数助词如“多”和“把”的分布差异进行横向对比，进一步明确了“来”的分布情况。

语义上，对于“来”的在数轴上的辖域，学界主要存在以下三种分歧：一、“来”只能左向取值，表示略少于20%；二、“来”只能右向取值，表示略多于20%；三、“来”的取值介于左右两个区间，但不等于其黏附的基数。对此，（吕叔湘 1999b: 347）[10]持第三种观点，他将“来”语义具体表述为，“表示大概的数目。一般指不到那个数，有时也表示稍大或稍小”，邢福义（2011）[11]也赞同该说法。王明仁（1997）[12]和王改改（2001）[13]也就“来”的语义问题，进行了不同侧重地描写和阐释。

总体来说，学界关注的焦点主要集中于“来”的句法分布和语义诠释。句法方面，主要停留在对语言事实的描写、归纳和总结，缺乏理据论证；语义方面，主要是定义模糊，辖域不明，缺乏明确的形式化分析。本文将采纳吕淑湘的观点，赞同“来”表示上下略微浮动，但与吕稍相左的是，笔者认为，即使所指刚好等于基数，也可使用“来”表示模糊义，同时，本文将重点论证作为概数助词的“来”在与其它语言成分组合时正确的句法结构，并在此基础上对其语义内涵进行形式化地刻画和描述。

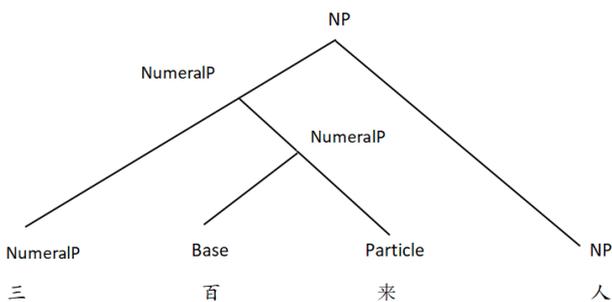
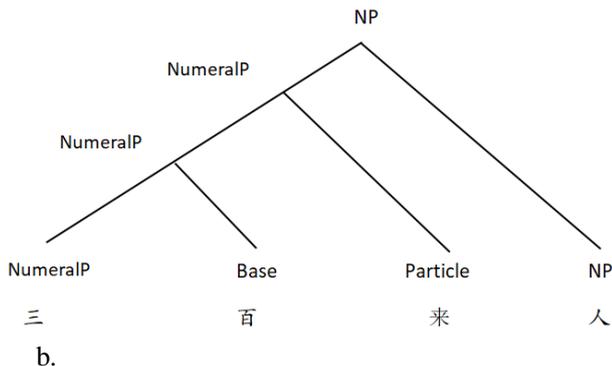
2 “来”的形式句法

“来”作为黏着词素，不管是黏附在数词还是量词后面，其后必须要加上一定的句法成分，如位数词、量词或名词，即汉语中“来”是不能悬空的，前后必须有依附和被依附的成分，这进一步说明了作为概数助词的“来”具有黏着词缀的特性。关于含有“来”的概数短语的

句法结构，可以参照张谊生(2001) [6]、邢福义(2003: 101-102) [14]和贺川生等(2020) [15]对于概数助词“多”的分析，虽然“多”本身是助词，但“数词+多”的组合却可以看成是一个数词短语。我们不妨假设同为概数助词的“来”也具有相似的特征。

以“三百来人”为例，助词“来”黏附在数词“三百”之后，构成一个表示大概数目的数词“三百来”，该数词再与名词短语“人”组合，形成新的名词短语。凭直觉，我们对于“三百来人”的理解应该是“二百七八十人”到“三百一二十人”之间。这个结果是怎么得出来的？“三百来人”从句法上应该如何分析呢？对此，存在以下两种解读，如图所示：

(2) a.



(2)a 表明概数助词“来”是黏附在整个数词短语“三百”后面，然后，概数词“[三百]来”再与 NP“人”组合，表示“三百上下或左右的人”。这种分析貌似合情合理，我们再看看其它语言的情况是否也是如此。根据前文的分析，作为概数助词的“来”，表示“±位数词×20%”，在外语中，语义上并没有与之完全吻合的助词来表达这样一个约量，所以，我们姑且将“来”翻译成“大概”或者“大约”。如：

- (3) a. 英语: around three hundred people
 b. 德语: rund drei Hundert Personen
 c. 日语: およそ 参 ひゃ 人
 d. 韩语: 대략 삼 백 명

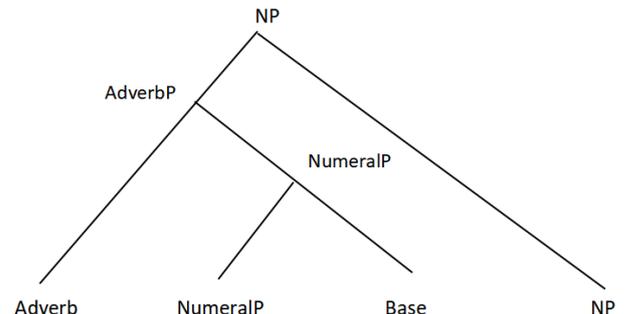
e. 法语: environ trois cents personnes

f. 汉语: 大约 三 百 人

‘三百来人’

从以上不同种类的语言中可以看出，表示“大约”的词都位于数词“三百”之前，这与汉语的“来”附于数词之后的现象刚好相反，但这只是语序的区别，并不影响“大约”与“数词“三百”的关系，其句法结构如(4)所示：

(4)

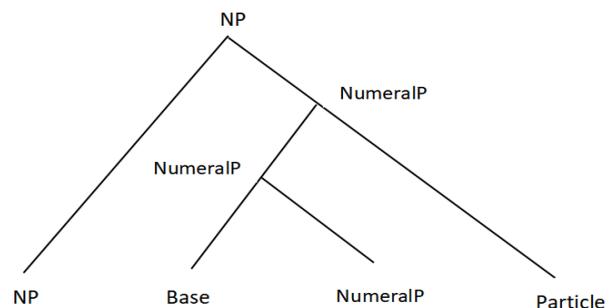


- a. 英语: around three hundred people
 b. 德语: rund drei Hundert Personen
 c. 日语: およそ 参 ひゃ 人
 d. 韩语: 대략 삼 백 명
 e. 法语: environ trois cents personnes
 f. 汉语 大约 三 百 人
 ‘三百来人’

很明显，这五种语言包括汉语中的“大约”都是作用于整个数词短语“三百”，而不是其中的“三”或“百”，表示以“三百”为基准上下浮动，同样，在语义取值上，“大约”也仅依赖于位数词“百”，这与汉语“来”的情况极为相似。我们有理由以类推的方式相信“来”也是修饰其附着的整个数词短语，而后再与之后的成分组合成词。可见，以上分析可以证明 2(a)的解读是合理的。

其实这一点，也可以在我们的少数民族语言——僜语(贺川生等 2020) [16]和格曼语(孙宏开等 1980: 194, 255) [17]中得到验证，以“多”为例，如下图(5)所示：

(5)



- a. 僂语: ŋejin law k'un maluɔdɔi
 社员 万 一 多
 ‘一万多社员’
- b. 格曼语: tʂ'utsu waje mu kuɣlat
 尺 百 一 多
 ‘一百多尺’

从图(5)可以看出,在类似于僂语和格曼语的少数名族语言中,数字是在位数词之后,这同样是语序的问题,在语义上,并不影响人们对于“万—”就是“一万”的理解。在这些语言中,表示概数的助词“多”也是黏着在整个数词短语之后,但语义上只与位数词相关联。

基于以上分析,我们有理由相信,同表概数的“来”在这类语言当中,也应该是位于整个数词短语之后的,如果这种推理能够站得住脚的话,那么它将再次证明(2)a的合理性。

通过以上跨语言的比较和分析可以得出,不管是印欧语系还是汉藏语系,不管是汉语还是少数民族语言,也不管系数词在位数词之前还是之后,亦或是概数词在数词之前还是之后,概数助词都是作用于整个数词短语的。

接下来我们看看(2) b的分析有什么不妥之处。(2) b中,概数助词“来”先黏附在位数词“百”之后,再与“三”组合成概数词“[三[百来]]”,然后再与名词短语“人”结合。这种分析似乎能很好地说明为什么“来”的取值仅取决于其前的位数词,这貌似与我们的语义分析很匹配。但是,在汉语的同一数字中,系数词和位数词是相乘的关系,不同等级的系数词之间才是相加的关系,如九万八千七百五十六应该是 $9 \times 10000 + 8 \times 1000 + 7 \times 100 + 5 \times 10 + 6 \times 1$,也就是九个万、八个千、七个百与六个一的和。这样,“[三[百来]]”的意思就是“ $3 \times (100 + \text{来})$ ”。很明显“百”和“来”的位数词属于不同的等级,依据前文的分析,“来”的语义辖域是“ \pm 位数词 $\times 20\%$ ”,“[三[百来]]”中的“来”表示在“百”的基础上“ $\pm 20\%$ ”,整个数值表示“ $300 \pm 100 \times 20\% = 280 \sim 320$ ”,也就是说,在这个短语中,“来”的位数词是“十”,比“百”低一级,所以“百”和“来”是相加的关系,显示在句法树上,就是互为姐妹节点,如(2) b所示。而且,“来”似乎与其黏附的位数词“百”前面的数字“三”并没有关联,因为当“来”与“百”发生完语义关系后,“来”就完成了自己的使命,接着便将自已隐身起来,不再参与之后的句法运算,这与(2) b的句法结构似乎也很一致。

那么问题究竟出在哪儿呢?按照刚才的分析,“百”与“来”先组合成概数词“百来”,再与“三”构成“三百来”。

接下来,我们看看“三”与“百来”的语义组合。既然汉语中,系数词与位数词是相乘的关系,那么按照(2) b的理解,“[三[百来]]”就是“三”个“百来”,即 $3 \times [(300-20) \sim (300+20)] = 3 \times (280 \sim 320) = 840 \sim 960$,这显然不符合其 $280 \sim 320$ 的语义指向。这就是为什么(2) b所投射的结构一定是不可取的。

3 “来”的形式语义

确定了概数词短语“三百来人”的句法结构,接下来我们面对的是如何刻画“来”的语义取值。按照(2)a的理解,在概数词“三百来”中,“来”是附着在整个数词短语“三百”后面,但其语义辖域却只取决于其黏附的位数词“百”。需要明确一点的是,像“十来人”“百来人”“千来人”以及“万来人”,位数词前面的系数词“一”都被省略掉了。这是汉语数字的一种特殊用法,英语是不被允许的,比如表示数字“一百”,只能是“one hundred”,“一千”,只能是“one thousand”,但是同为印欧语系的德语,百位和千位却允许系数“一”的省略,就像汉语一样,如“一百”是“(ein)hundert”,“一千”是“(ein)tausend”。

现在我们来看看怎样用数学公式描绘“来”在不同的位数词后面的取值情况,如(6)所示:

- (6) a. \parallel 十来人 $\parallel = (10 - 10 \times 20\%) \sim (10 + 10 \times 20\%)$
 $= 8 \sim 12$ 人
- b. \parallel 二十来人 $\parallel = (20 - 10 \times 20\%) \sim (20 + 10 \times 20\%)$
 $= 18 \sim 22$ 人
- c. \parallel 一百来人 $\parallel = (100 - 100 \times 20\%) \sim (100 + 100 \times 20\%)$
 $= 80 \sim 120$ 人
- d. \parallel 四千来人 \parallel
 $= (4000 - 1000 \times 20\%) \sim (4000 + 1000 \times 20\%)$
 $= 3800 \sim 4200$ 人
- e. \parallel 五万来人 \parallel
 $= (50000 - 10000 \times 20\%) \sim (50000 + 10000 \times 20\%)$
 $= 48000 \sim 52000$ 人

由于“来”的本义是倾向于“较少”的意思,而人们对于“百万”“千万”“亿”这样的单位,直觉上会觉得它们多得难以计数,所以,像“七百万来人”、“八千万来人”和“九亿来人”这样的表达,可能会产生内部的语义冲突,因此,其可接受度并不高。这就是为什么 BCC 和 CCL 语料库鲜有相关语料。吕叔湘(1999b) [10]的公式(1) a也表明,“来”所附着的位数词只包括“十、百、千、万”。但其实,在特殊情况下,为了强调某一数字虽然其本身表示比较大的数值,而实际上说话人想要强调它还

4.1 简单数词

不管后面是接名词、类别词还是度量单位，“来”都不能黏附在简单数词上，如“五来个人”、“八来只鸡”、“六来斤重”等。根据公式(8)，我们来计算一下“五来个人”的取值范围。其中 x 为 5，因为 5 相对于 10 的幂为 0，所以 y 的取值为 0，演算过程如下：

$$\begin{aligned} (11) \parallel \text{五来个人} \parallel \\ = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \\ = [(5 - 10^0 \times 20\%) \sim (5 + 10^0 \times 20\%)] \\ = 4.8 \sim 5.2 \text{ 个人} \end{aligned}$$

“五来个人”得出的是小数。乍一看，这个结果似乎与“来”的语义辖域刚好相符，即表示上下浮动 20%，但是，在汉语中，4.8 个人是没有意义的，人只能按整个算，不能分割。同样，套用公式(8)，“八来只鸡”为“7.8~8.2 只鸡”，也不符常理。

再看看“六来斤重”，省略推算过程，它的结果是“5.8~6.2 斤重”，这貌似是合理的。但很遗憾，这个说法在汉语中也是不可接受的。为什么？比较合理的解释是，“来”表示的是概数或者说比较模糊的数字，而小数表示的是比较精确或者说比较确切的数字，语义上，这两者存在矛盾，即人们不能用一个确数去表达一个概数，所以，“简单数词+来”的结构不符合人们的心理预期，因而在日常表达中是不被接受的。同理，“末尾为简单数词的多位数+来”，其推算结果末尾也将出现小数，也不合法。

4.2 分数和小数

学者们常将概数词“来”与“多”放在一起比较。汉语中，有“二分之一多”、“十分之一多”的说法，“多”前可以容纳分数词（贺川生等 2020）[16]。但“来”之前却不能出现分数，如我们不说“三分之一来米长”、“五分之三来米宽”、“七分之五来斤重”等。

首先，需要明确一点，在短语“五分之三来米宽”中，“来”并不是附着在“三”上，而是附着在整个分数“五分之三”即 0.6 上。这一点很容易就能够得到证明。如果“来”黏附在“三”后，“五分之三来米宽”就是“五分之（2.8~3.2）米宽”，即“1.4~1.6 米宽”，这与原义相差甚远。所以比较合理的方法是，先把“五分之三”换算成 0.6，“五分之三来米宽”的意思就是“0.6 来米宽”，“来”附着在数字 6 上，其位数词，按照汉语的计算系统，是“分”，即 10^{-1} 。套用一下公式(8)，可得出：

$$\begin{aligned} (12) \parallel \text{五分之三来米宽} \parallel = \parallel 0.6 \text{ 来米宽} \parallel \\ = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \end{aligned}$$

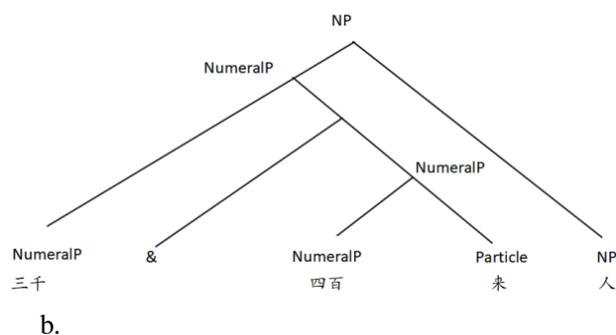
$$\begin{aligned} = [(0.6 - 10^{-1} \times 20\%) \sim (0.6 + 10^{-1} \times 20\%)] \\ = 0.58 \sim 0.62 \text{ 米宽} \end{aligned}$$

同样，得到的也是一个小数，且比原来的分数更加精确的小数。上文已经讨论过，为了避免语义上的冲突，表概数词的“来”前面不接表示精确数目的小数，因为分数可以换算成小数，所以分数也也不与“来”相结合。人们使用助词“来”表示概数，为的是快速地估算某个数值的大小，而小数或分数表示的都是很具体的数字，这在语言交流地过程中，需要人们具有相当的心算功底，这无疑会增加语言交际的负荷，显然也背离了人们使用语言的省力原则。所以，“分数或小数+来”是不可取的。

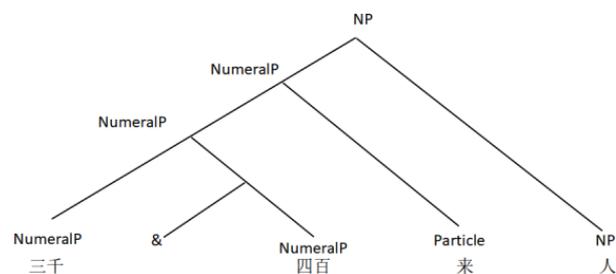
4.3 多项式系位结构

个位数为 0 的多项式系位结构如“三千四百”、“五万六千三百二十”等后面也可以跟“来”表概数，只不过其内部各个位数词之间是并列的关系。这时，只能认为“来”是作为后缀黏附在最后一个位数词短语上，而不是整个数词之后。以“三千四百来人”为例，根据 He（2015）[16]，在句法结构中，两个相邻的位数词短语之间需加入一个隐形的连词“&”，表示算术加，否则该类短语在语义上无法解读。其句法结构如(13)a 所示，而非(13)b：

(13) a.



b.



根据常识，“三千四百来人”的意思是大约 3380~3420 人。根据公式(9)， $x=3400$ ， y 表示“ x 相对于 10 的幂”，即 $y=3$ 其推算过程如下：

$$\begin{aligned}
 (14) \text{ a.} & \parallel \text{三千四百来人} \parallel \\
 & = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \\
 & = [(3400 - 10^3 \times 20\%) \sim (3400 + 10^3 \times 20\%)] \\
 & = 3200 \sim 3600 \text{ 人}
 \end{aligned}$$

(13)a 的错误在于简单系位结构和多项式系位结构是有区别的。简单系位结构中, x 为“来”所黏附的整个数字, y 表示相对于 10 的指数, 因为 x 末尾的位数词只有一个, 紧临助词“来”。而在多项式系位结构中, 位数词不止一个, 而与“来”发生语义关联的只有整个数词末尾的位数词。显然, 按照公式(8), 并不能得出正确答案。

根据上文分析, “来”的数值仅取决于其黏附的位数词, 与整个数词并没有关系, 所以, y 的取值并不是依赖于 x, 而是取决于位于 x 末尾的位数词, 只是当 x 碰巧是简单系位结构时, x 相对于 10 的指数刚好和其末尾的位数词相对于 10 的指数重合了。因此, 在“三千四百来人”中, “来”紧临位数词“百”, 所以, y 等于 2, 而非 3。其正确演算过程应如下:

$$\begin{aligned}
 (14) \text{ b.} & \parallel \text{三千四百来人} \parallel \\
 & = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \\
 & = [(3400 - 10^2 \times 20\%) \sim (3400 + 10^2 \times 20\%)] \\
 & = 3380 \sim 3420 \text{ 人}
 \end{aligned}$$

4.4 小数式复杂系位结构

在含有概数助词“来”短语中, 可以出现小数式复杂系位结构, 如“二点五万人来”。同样, “来”也是黏附在整个数词短语之后, 其取值范围如下:

$$\begin{aligned}
 (15) \text{ a.} & \parallel \text{二点五万人来} \parallel \\
 & = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \\
 & = [(25000 - 10^4 \times 20\%) \sim (25000 + 10^4 \times 20\%)] \\
 & = 23000 \sim 27000 \text{ 人}
 \end{aligned}$$

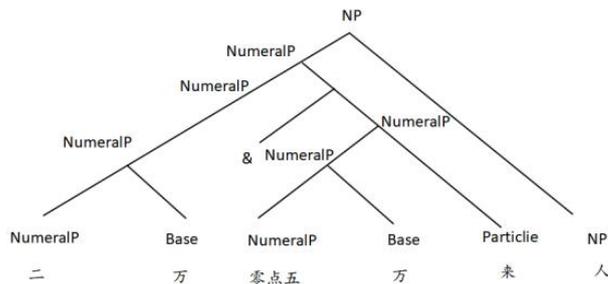
很明显, 这个结果是错误的。“二点五万人来”的实际取值应该是“两万五千人左右”, 即在“两万五千”的基础上上下浮动“两百人”, 而不是“两千人”。这是怎么得出来的呢? 这要从小数式复杂系位结构的特殊性说起。“二点五万”的意思等同于“二万五千”, 那么在“二点五万人来”中, “来”其实是黏附在“千”, 而不是“万”上。所以, 其正确的推算过程应当如下:

$$\begin{aligned}
 (15) \text{ b.} & \parallel \text{二点五万人来} \parallel \\
 & = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \\
 & = [(25000 - 10^3 \times 20\%) \sim (25000 + 10^3 \times 20\%)] \\
 & = 24800 \sim 25200 \text{ 人}
 \end{aligned}$$

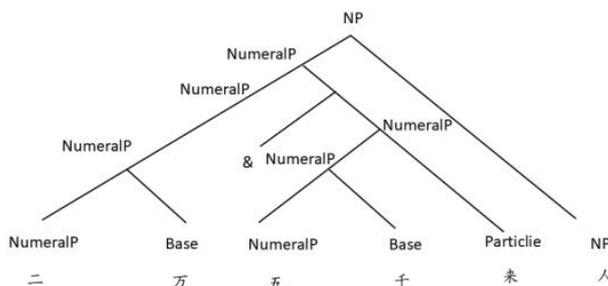
以上分析再次证明, 概数助词“来”并不是直接附着在位数词上, 而是作为整个数词短语的后缀。而对于

小数式复杂系位结构的概数短语, 因为其中含有位数词省略的情况, 所以必须先将其换算成常规的多项式系位结构, 才能确定“来”所真正黏附的位数词, 其句法结构如(16b) 所示:

(16) a.



b.



5 类推

“来”可以黏附在数词之后, 也可以黏附在数量词之后。前文主要分析了“数词+来”的结构, 接下来本文将主要探讨“数量词+来”的情况。根据吕叔湘(1957)[1], “来”的位置主要遵循以下三条规律:

- (17) a. 量词是类别词, 如个、只、块等, 只有末尾为零的多位数才能带“来”, 且“来”不能放在量词之后。
- b. 量词是度量单位, 如果数词是末尾为零的多位数, “来”则放在数词之后; 如果数词是 1~9, “来”则放在量词之后。
- c. 只有当数词是“十”, 且量词是度量单位时, 两种顺序都可以, 如“十来斤米”, 或者“十斤来米”。

根据吕叔湘的分析可得出, 能够出现在概数词“来”的短语中的量词分为类别词和度量单位。当“来”前的数词是末尾为零的多位数时, 其后可跟类别词也可跟度量单位, 即“数词+来+类别词/度量单位”, 如:

- (18) a. 三百来只鸡/四千来匹马/二十来尾鱼
类别词

- b. 五十来米长/六百来斤重/七千来英尺来高
度量单位

其实, 这种情况下, “来”也可放在度量单位之后, 如:

- c. 五十米来长/六百斤来重/七千英尺来高

当“来”前的数词是 1~9, 其后只能跟度量单位, 不能跟类别词, 如:

- (29) a.* 三只来鸡/四匹来马/二尾来鱼

类别词

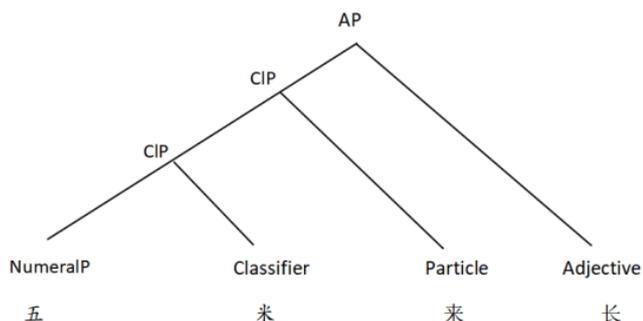
- b. 五米来长/六斤来重/七英尺来高

度量单位

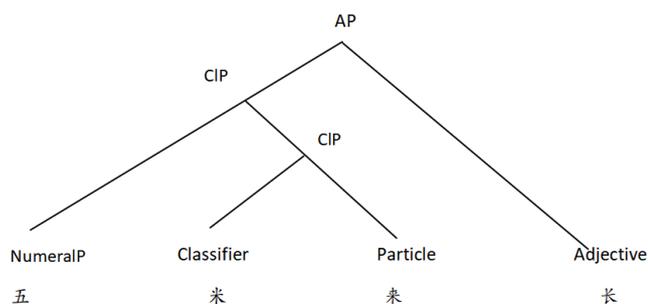
数词“十”比较特殊, 其后不仅可以接类别词, 当量词是度量单位时, 量词可前可后, 只不过, 意思上有些许差异, 这可能跟“十”本身是位数词, 但又可以做简单数词有关。

接下来, 我们来探讨一下“来”后跟量词的句法结构。依照类推分析, 我们有理由相信, “来”是黏附在整个数量词短语后面, 而不是单个的量词后面, 即“五米来长”的正确句法结构是(20)a, 而不是(20)b。

(20) a.



b.



按照(20)b的句法结构---“五[米来]长”, 首先“米来”本身就不能构成一个具有独立意义的语义单元, 即便它能, “米来”表示在“(一)米”的基础上上下浮动 20%, 即“0.8~1.2 米”, “五[米来]长”表示“五个 0.8~1.2 米”, 大约是“4~6”米, 这与我们常识中所认为的“五米来长”相去甚远, 它表示的应该是大概 4.8~5.2 米。所以, 与

“数词+来”一样, 在“量词+来”的结构中, “来”也是黏附在整个数量词之后, 但语义上, 也只与其前最近的量词发生关联。

现在, 我们面临的问题是如何确定该结构的语义解读。对于“数词+来”, 前文已经确定了可以应用公式(8)来求得“来”的取值范围, 那么, 对于“量词+来”情况, 该公式是不是也适用呢? 答案是肯定的, 只不过需要在原公式的基础上对变量 x 进行一些限制, 如下所示:

(21) $\|来\| = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)]$ (x 为数词或者数量词)

为什么当 x 为量词时, 仍然适用于该公式呢? 下面我们将就此进行深入探讨。根据(16)的规律可以看出, 类别词不能出现在“来”前, 因为与类别词搭配的名词一般都只能看成是一个整体, 不能分割, 如“一个人/三只鸡/四条狗”, 而“来”表示的是相较于其黏附的事物“略多或略少一点点”, 如“三只来鸡”, 指的是“三只鸡上下”, 即“不到三只鸡, 或者比三只鸡多点点”, 这样, 只能把一只鸡剁成鸡块才能进行计算, 显然, 这有悖常理。因此, 在“数量词+来”中, 量词只能是度量词, 邢福义(2011) [11]也对此进行过探讨--“只有一类数量词可以出现在‘来’字的前边, 这就是由度量衡单位构成的数量词, 即‘数词+度量衡单位’”。以“米”为例, “米”是计量单位不是计数单位, 且总是以“1 米”为单位, “五米”表示 5 个 1 米, “五十米”表示 50 个 1 米, “五千米”表示 5000 个 1 米, 依此类推。那么“五米来长”, 表示 5 个 1 米, 再在“1 米”的基础上少一点或多一点。这里, 我们可以确定基数 x 为 5 米, 可是 y 的取值, 如何界定呢? 根据前文分析, y 指的是与“来”发生语义关联的词相对于 10 的幂。“五米来长”中与“来”发生语义关系的则是居于其前的“米”, 即“1 米”, 而“米”相对于 10 的幂是 0, 即 $10^0 = 1$ 米, 所以 $y = 0$ 米, 其推算过程如下:

$$\begin{aligned} (22) \quad \|五米来长\| \\ &= \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)] \\ &= [(5 \text{ 米} - 10^0 \text{ 米} \times 20\%) \sim (5 \text{ 米} + 10^0 \text{ 米} \times 20\%)] \\ &= 4.8 \text{ 米} \sim 5.2 \text{ 米} \end{aligned}$$

所以, 公式(21)可进一步修改为:

(23) $\|来\| = \lambda x \lambda y [(x - 10^y \times 20\%) \sim (x + 10^y \times 20\%)]$ (x 为数词时, y 为位数词相对于 10 的幂; 当 x 数量词时, y 的值固定取 0)

这意味着我们可以将包含概数助词“来”的所有数词短语结构进行统一刻画和描写。这将大大提高计算机对于概数助词语义的计算和识别性能, 从而提升自然语言分析和处理的速度, 有助于推动人工智能的发展。

6 结语

本文以当代语言学的主流学派--生成句法学和形式语义学的最新研究成果为切入点，基于汉语和少数民族以及其它语系语言的真实语料，论证了概数助词“来”在构成数词短语时的句法分布，并构建了精准诠释其语义的模型。本文属于一种界面研究，为汉语概数助词的研究提供了新的视角，对语言本体的跨学科研究也具有一定的启示和借鉴意义。该研究成果可应用于计算机对于自然语言的识别和破译，也可直接应用于汉外教学。

参考文献

- [1] 吕叔湘. 试说表概数的“来” [J]. 中国语文, 1957 (04).
- [2] 江蓝生. 概数词“来”的历史考察 [J]. 中国语文, 1984 (02).
- [3] 王庆. 概数助词“来”的语法化演变动因与机制探析 [J]. 现代语文, 2017 (01): 59-61.
- [4] 吕叔湘. 数量词后的来、多、半. 汉语语法论文集 [C]. 北京: 商务印书馆, 1999a: 209-216.
- [5] 杨德峰. 表示概数德“多”和“来”德全方位考察 [J]. 汉语学习, 1993 (3): 10-16.
- [6] 张谊生. 概数助词“来”和“多” [J]. 江苏师范大学学报(哲学社会科学版), 2011 (03): 80-84.
- [7] 杜晓艺. 概数助词“来”的句法分布 [J]. 高等函授学报(哲学社会科学版), 2006 (06): 31-32.
- [8] 李柔冰. 现代汉语概数助词“把”、“来”、“多”的用法考察 [D]. 上海: 上海师范大学, 2011: 1-68.
- [9] 应学风、王晓辉. 数量结构中概数词“来”和“多”的分布 [J]. 汉语学习, 2014 (04): 38-45.
- [10] 吕叔湘. 现代汉语八百词(增订本) [M]. 北京: 商务印书馆, 1999b: 347.
- [11] 邢福义. 事实终判：“来”字概数结构形义辩证 [J]. 语言研究, 2011 (01): 1-10.
- [12] 王明仁. 表概数“来”的语义和由来 [J]. 香港词库建设通讯, 1997 (12).
- [13] 王改改. 概数词的语义调查和研究 [J]. 汉语学习, 2001 (06): 46-49.
- [14] 邢福义. 词类辨难(修订本) [M]. 北京: 商务印书馆, 2003: 101-102.
- [15] He Chuansheng. Complex numerals in Mandarin Chinese are constituents [J]. *Lingua*, 2015 (164): 189-214.
- [16] 贺川生、邓丽芳、谢丽丽. 概数助词“多”的形式句法和形式语义 [J]. 当代语言学, 2020 (02): 237-253.
- [17] 孙宏开、陆绍尊、张济川、欧阳觉亚. 门巴、珞巴、僜人的语言 [M]. 北京: 中国社会科学出版社, 2008.
- [18] 朱德熙. 数词和数词结构 [J]. 中国语文, 1958 (04): 185-186.