

A Research on the Construction of the Common Collocation Corpus and Hierarchical Corpus of Degree Complement

Li CAI

College of Chinese Language and Culture, Jinan University, Guangzhou, Guangdong, China

Keywords: Degree complement, Predicate, The common collocation corpus, Hierarchical corpus.

Abstract. This research is to explore possible combination of error analysis with formulated collocation patterns and resources which could support Chinese teaching and research methodologies. By forming the degree complement and the collocation of predicate in differences levels sample data, it will support in the writing of Chinese grammar books and dictionaries, as well as, Chinese learning.

程度补语与述语常用搭配库以及分级例句库的构建

蔡丽

暨南大学华文学院，广州，广东，中国

关键词: 程度补语；述语；搭配库；分级例句库

摘要: 本文基于留学生程度补语使用情况和偏误分析结果，结合述语及其程度补语搭配的特征与规律，探索将偏误分析与搭配规律、搭配资源结合起来辅助汉语教学的研究方法，构建程度补语与述语搭配库以及按难易程度分级的搭配的例句库，辅助汉语教材语法点的编写、汉语学习词典编撰和辅助汉语学习。

1. 引言

在现代汉语中，程度补语同程度副词一样，是一个相对封闭的类，程度补语存在一系列有别于其他类型补语的句法语义特征。Li Cai (2011) 全面分析了程度补语的立类依据，并依此标准确定现代汉语中程度补语的43个成员，并进行了多角度的分类。Li Cai (2012) 采用计算语言学的自动分词技术对北京大学现代汉语语料库中检索到的全部相关语料进行计算与分析，在大规模语料统计结果的基础上，考察、分析43个程度补语与充当述语的词语的组合特征，总结其中的规律。Li Cai (2017) 分别以43个程度补语为检索项检索北京语言大学HSK作文语料库和暨南大学华文学院东南亚留学生作文语料库，考察外国留学生对汉语程度补语的掌握与使用情况。基于上述留学生程度补语使用情况和偏误分析结果，结合述语及其程度补语搭配的特征与规律，我们可探索将偏误分析与搭配规律、搭配资源结合起来辅助汉语教学的研究方法，进一步构建程度补语与述语搭配库以及按难易程度分级的搭配的例句库，辅助汉语教材语法点的编写、汉语学习词典编撰和辅助汉语学习。

2. 面向汉语教学的程度补语与述语常用搭配库构建

2.1 常用搭配库的入库对象

现代汉语中共有43个程度补语，但这些程度补语在语言使用中的频率各不相同，我们根据对北京大学现代汉语语料库的统计结果，结合程度补语在北京语言大学HSK作文语料库和暨南大学华文学院东南亚留学生作文语料库两个中介语语料库和海外汉语教材语料库中出现的情况，确定入库标准如下：第一，凡在两个中介语语料库和海外汉语教材语料库中出现过

的程度补语，均确定为搭配库入库对象；第二，未在中介语及教材语料库中出现、但在母语语料库中出现次数高于100的。据此，分别确定如下程度补语为入库对象，共计32个：（1）在两个中介语语料库和海外汉语教材语料库中均出现了的程度补语，共11个，包括“极、多、坏、很、厉害、死、万分、无比、透、不得了、不过”。（2）在两个中介语语料库中均出现了的程度补语，除上述11个外，还有以下10个，包括“不堪、要命、慌、可怜、之至、要死、透顶、不行、疯、过分”。（3）仅在HSK作文语料库里出现的，有4个，分别是“异常、够呛、绝顶、绝伦”。（4）仅在东南亚留学生作文语料库中出现的，1个，即“了不得”。（5）仅在海外汉语教材语料库中出现的，1个，即“惊人”。（6）除上面27个程度补语外，在母语语料库中出现次数高于100的，还有以下5个：出奇、吓人、非常、傻、可以。

2.2 搭配库的入库内容

我们将上述32个程度补语在各个语料库中出现的组合词语分别进行了整理，由于北京大学现代汉语语料库中出现的搭配词种数量较多，为体现常用性，对该语料库中的词种取出现频率排在每个程度补语前40位的左搭配词进入搭配库。如果第40位序的包含若干出现同样频次的左搭配词，则同时收入。对于各语料库中左搭配词使用不当或个性化、带有活用特点的用例，我们进行人工干预，去除不正确及不具代表性的用例，保留适用于教学用例。每个词语之后用括号标明其出现的频次，均为干预后的数字，与前面的原始统计数字略有出入。以东南亚留学生作文语料库中程度补语“要命”为例：

要命（总词种数：24，总词语次数：47）热（10）、累（9）、痛（3）、哭（2）、高兴（2）、怕（2）、难（2）、笑、漂亮、担心、撑、滑、气、饿、辛苦、冷、瘦、任性、心疼、头疼、挤、害怕、思念、痒

2.3 搭配库入库内容的异同分析

2.3.1 搭配库自动检索平台的设计

为方便教学，我们将上述32个程度补语在各个语料库中出现的组合词语分别进行异同对比分析，并设计了相应的检索平台。在检索平台中输入进入搭配库中的某个程度补语，即可自动呈现这个程度补语的常用搭配词语，以及这些词语在每个语料库中出现的频次。这个检索平台可为汉语作为第二语言教材及词典相关语法点的编写、教师教学用例的设计提供参考。

2.3.2 搭配库自动检索结果的相关分析

下面选取在中介语语料库、海外汉语教材语料库以及母语语料库三种不同类型的四个语料库中均出现了的程度补语，即“极、多、坏、很、厉害、死、万分、无比、透、不得了、不过”，统计分析在三种不同类型语料库中与这些程度补语搭配的词语及这些词语出现的频次，并在此基础上分析每个程度补语在用例选择方面的原则与方式。

需要说明的是，由于数据量大，我们在选择入库词语时，母语语料库入选的是排序在前40位的组合词语，40位之后的词语我们未收搭配库，而中介语语料库和海外汉语教材语料库中的搭配词语是全部收入。

3. 按难易程度分级的程度补语与述语搭配的例句库构建

在对本体研究、留学生中介语偏误分析及汉语教材相关语法点编排的基础上，讨论构建辅助对外汉语教学的程度补语与述语组合的例句资源库的设想，探索将搭配规律研究与搭配资源构建结合起来辅助汉语教学的方法。

3.1 不同级别的对外汉语语法教学的基本要求

初、中、高级三个阶段语法教学内容与方法各有侧重，在进行例句分级处理时，首先要明确各个不同阶段教学的基本特点。

3.1.1 初级阶段语法教学

初级阶段的语法教学应注意以下几方面：(1) 只需教最基本的语法形式，使习得者具备区分正误的能力。讲解语法点时讲究句法结构，引导学生掌握汉语的句型、词序，常见虚词的用法等，不要求做过多的语义辨析。(2) 初级阶段的语法教学不仅要教语法知识本身，更要教语法知识怎样在交际中正确、有效地使用。学习语言的终极目的不应该在于习得语言本身，而应该在于知道如何应用语言来满足各种实际的需要。所以，在初级阶段综合课的语法教学中，教师要更加注重引导学生能在真实的生活交际中有效而正确地应用语法知识而不是单纯地讲解语法本身。(3) 在具体语法点讲解时，侧重讲在课文中出现的那种用法。对于一些较复杂的语法点或语法规则，要分化难点。对初级阶段的学生来说，如果同时教两种用法，他们会容易混淆，可根据课文中出现的句子确定教哪一种用法。(4) 尽量少用语法术语，用形象生动的方式来解释，尽量多用典型例句，引导学生发现、归纳其中的规律。

3.1.2 中级阶段语法教学

中级阶段讲语法侧重语义分析，注意句中成分的语义关系及语义搭配，使习得者具备区别语言形式异同的能力。在中级阶段，我们应该将习得者在初级阶段学过的相近语法点加以总结，把它们归在一个总的形式之下，然后辨别它们细微的语义差别。

3.1.3 高级阶段语法教学

高级阶段所讲语法侧重语用功能，着重语用的选择和词语的应用，目的在于表达得体。高级阶段侧重语法的语用功能教学，使习得者具备区别语言形式之高下的能力。有时单看一句话，就连它的确切含义都很难判断，如：“有了他的帮助，事情就好办了。”这句话可以理解为“因为……，所以……”，也可以理解为“如果……”，要靠上下文来决定，这就是语法现象使用的环境和语言背景的作用。

3.2 对外汉语语法教学例句设计的原则

3.2.1 可接受性

(1) 例句应简单易懂，除所教语法点外，不应有学生未掌握的生词或语言点。在设计例句时，教师应注意，例句中不宜出现学生之前未学过的词语，也不宜出现与所教语法点无关的其他未掌握的语言点，以免分散学生的学习注意力。

(2) 例句所表达的内容应该是学生所熟悉或熟知的。展示语法点的例句所表达的内容应该尽可能是贴近学生生活的常识性内容，不宜选用过于生僻或少用的内容。

(3) 设计例句时可选用学生刚刚学过的课文内容。设计例句时可选用学生刚刚学过的课文内容，一则可以起到温故而知新的效果，二则刚学过的内容学生印象深刻，有亲切感。

3.2.2 针对性

(1) 针对所教的语法现象的特点及其交际功能来设计例句。每个语法点都有其自身的使用特点，有的用孤立的例句，无需上下文语境就可以让学生明白，但有的语法点因其语义特点，需有一定的语境才能让学生掌握其用法。

(2) 针对学生常犯的语法错误来设计例句。

(3) 针对学生的语言水平来设计例句。针对不同学习阶段学生的语言水平的不同，语法教学设计的例句应有所区别，初级阶段的例句应尽量简短，不宜出现过长或过于复杂的句子结构，而中级阶段例句长度可适当增加，到高级阶段则可使用较为复杂的句子结构。

3.2.3 层次性

语法教学要注意分化难点，合理安排教学层次。同一个虚词或同一个句式往往会有不同的用法。例如：“除了……以外，……”这一句式，“除了”可以放在主语前，也可以放在主语后。在教学时，要区分两种不同位置的用法，根据课文中出现的相关句子来确定先教哪一种用法，后教哪一种用法。同时，也得根据学生的水平来确定是否需要教几种用法。

此外，句子层面的表达比较复杂，同一种句型或者句式往往有多个具体的句子表达形式，分别适应不同的句法环境和交际场合，使用的频率往往也不相同，如“把”字句等。因此，

句子的教学需要分化句式、句型的难点，区分不同用法的教学顺序，在不同的教学阶段确定合适的教学内容，而不能不分步骤地全部教给学生，那样只能适得其反。

3.2.4 典型性

汉语的语法规则并非整齐划一的，各种规则都会有例外的情况，在选择例句时应尽量选取最能体现该语法点特点的典型例句。

3.3 例句按难易程度分级的方法

例句的难易度需要从两个方面考虑：其一，例句中所使用词语的常用度。例句中的词语越常用，其已习得的可能性就越高，学生就更有可能将注意力集中在要学习的语法点上。其二，整句的难易度。影响句子难易程度的因素包括句长、句子结构形式、句中所有词语在文本中出现频率的高低、词语语义类别的多少等等。其中，句子结构形式的复杂与否较难通过机器自动识别，需要用人工干预方式进行。

根据上述思路与方法，我们以海外汉语教材中出现的程度补语例句为对象，首先分别对其进行难易度自动分级处理，然后再进行适当的人工干预，形成每个程度补语的分级例句资源。选择海外汉语教材中的出现的程度补语例句，是因为教材中的例句都是教材编写者根据学生学习汉语的基本特点和规律编写的，例句本身经过了一定加工，且具有一定的示范性、代表性。

下面以程度补语“极”为例展示具体例句分级过程。

步骤一：确定原始素材。在海外汉语教材语料库中带有程度补语“极”的例句共62句。

步骤二：利用自动分级程序对例句进行难易度分级。采用自行设计的例句难易度自动分级程序对上述例句进行难易度自动计算处理。限于篇幅，以下展示部分例句（前10句和后10句）难易度计算排序结果：

(2) 我高兴极了。63.6810745015205

(4) 文波高兴极了。85.2361088166333

(3) 味道好极了！91.5876454451224

(1) 好看极了。95.3408212784676

(7) 同学们高兴极了。97.5093560705038

(5) 小狗高兴极了。101.936159164322

(15) 我高兴极了，也想给他拍一张。121.699180530878

(9) 小猴子心里高兴极了。147.353854052887

(11) 家里笑声不断，热闹极了！161.661082950816

(23) 狼说：“我饿极了，让我把你吃了吧！”185.546245071262

……

(52) 我的话刚说完，眼前电光一闪，雷声又响起，声音大极了，妹妹尖叫起来，我也吓了一跳。461.893347151367

(28) 老渔翁高兴极了，招呼海豚说：“上来吧，海豚老弟！”466.396074586451

(58) 狐狸又饿又渴，抬头看见葡萄架上挂满了一串串葡萄，高兴极了，连口水都流出来了。479.011640501257

(37) 在新年之前，大家都忙着大扫除、办年货、做糕饼，热闹极了。488.375858909378

(54) 今天，老师带我们走出课室，到海边游玩，大家在洁白的沙滩上玩得开心极了。494.344099416756

(39) 五光十色的彩球一闪一闪，圣诞树最高处挂上一颗大星星，好看极了！497.721811759491

(53) 财主气极了，快步往老二的房间走去，一推开房门，只见满地都是撕破了的书。540.287735435265

(60) 他们高兴极了，忽东忽西，忽高忽低，不停地飞着，配合着清风流水，唱出悦耳动听的歌曲。568.411613321642

(62) 朝天开的, 像一支支扬声吹奏的小喇叭; 倒过来的, 好像小姑娘的裙子, 美丽极了! 596.513988030281

(49) 极乐寺依山而建, 我们拾级而上, 阵阵微风迎面吹来, 凉爽极了! 652.249585253539
每个例句之后的数字均为根据难易度计算公式计算出的例句的难易度系数Ls, Ls系数越小, 代表难度越低, Ls系数越大, 则代表难度越高。

步骤三: 对自动分级结果进行人工干预。我们认为, 难易度自动计算结果总体而言与教学和语感经验基本吻合。从句式上来看, 单句的结构较复句简单, 应先出单句, 而单句中使用的词语越简单越常用, 句中出现的词语数量越少, 句子则越简单。复句则先依据其层次来判定整句难度, 层次越多的复句越难理解, 然后再根据各分句的句式特点、句中出现词语的常用度和语义复杂性判定综合难度。上述排序中, 有以下可调整之处: 第一, 关于单句的位序调整。例(9)应排在例(15)之前, 因为例(9)是一个单句, 而例(15)是一个复句, 例(9)除了比排在第一位的例(2)多一个状语外, 其他内容的难度基本一致。根据我们前面讨论的例句设计的原则, 有的例句选取的内容是课文中讲过的相关内容, 句子结构本身较为简单, 那么排除了词汇的难度之后, 句子的整体难度就会大大降低。少数句子句中除有一词常用度略低外, 句式本身简单的, 位序需适当调前。第二, 属于复句的例句, 根据构成复句的分句的结构层次特点以及各分句所使用的句式的难易程度来适当调整位序。例如, 例(23)中狼说的话是一个复句, 其中第二个分句是兼语兼把字句, 是较难习得的句式, 位序应调后。例(15)的第二个分句有两个状语, 较一个状语的要难一些。依据上述相应原则调整位序共计10句, 调整率为16.13%, 自动分级结果的认可率为83.87%。

依据上述方法和流程, 将程度补语的原始例句资源一一进行自动分级和人工校验, 就形成了按难易程度排序的程度补语与述语搭配的例句库。这个按难易度排序的例句库可作为教学用例的参考, 排序在前的适合初级阶段使用, 排序居中的适合中级阶段使用, 排序居后的则建议在高级阶段使用。该例句库对于汉语教学、教材编写和词典编撰都有一定的参考作用。

致谢

本文为国家社科基金一般项目《海外华语语言生活状况调查及多媒体语言资源库建设》(13BYY090)和2016年广东省应用型科技研发专项资金项目《基于大数据和游戏化学习的教育云平台研发及规模化应用》(2016B010124008)的阶段性成果之一。

References

- [1] The Language Class of Grade 1955-1957 of Chinese Department of Peking University . Explanation of Function Words in Modern Chinese, The Commercial Press, 1982.
- [2] Jingmin Shao. On the principle of semantic mutual option in Chinese Grammar. Journal of Chinese Linguistics (the 8th issue), The Commercial Press, 1997.
- [3] Yisheng Zhang. Exploration of modern Chinese adverbs, Academia Press. 2004.
- [4] Li Cai, A study on the degree complement based on computational linguistics, Lecture Notes in Electrical Engineering, vol. 137, 2011, pp. 733-740.
- [5] Li Cai, A study on the "hen shi" structure and other issues based on calculation of large-scale corpus, Advances in Education Research (AER), vol. 48, 2013, pp. 8-13.
- [6] Li Cai, Range and category of degree complements in modern Chinese, Journal of Ningxia University (Humanities and Social Sciences Edition), vol. 33, 2011, pp. 9-14/32.
- [7] Li Cai, A Study on the Use of Chinese Degree Complement of Foreign Students, The 2017 2nd International Conference on Humanities Science and Society Development, 2017.