

硕士研究生学位论文

**题目：构式视角下的现代汉语四字表达的能产性研究**

**姓　名**：　　　田　骏

**学　号**：　　1301211280

**院　系**：　中国语言文学系

**专　业**：　 汉语言文字学

**研究方向**：　 中文信息处理

**指导教师**：　 詹卫东　教授

二〇一六 年 六 月

**版权声明**

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

# 摘要

A Productive Research on Modern Chinese Four-Word Expression: from a Constructional Perspective

Tian Jun (Chinese Linguistics)

Directed by Professor Zhan Weidong

# ABSTRACT

KEY WORDS:

# 目　录

摘要 I

ABSTRACT II

目　录 III

第一章　引言 1

1.1 问题的提出 1

1.2 选题背景及意义 1

1.3 文献综述 1

1.3.1 不同的术语体系 2

1.3.2 前人研究的阶段性特点 4

1.3.3 前人研究的不足 5

1.4研究方法 7

1.5论文结构安排 7

第二章　现代汉语四字表达概述 8

2.1 四字表达的定义 8

2.2 四字格式的定义、性质及规模，兼谈半凝固构式 8

2.3 能产性？从词库扩展说开去 8

第三章　现代汉语四字表达的几种能产性 9

3.1不经意的错误：误用能产性 9

3.2 巧妙的引用：仿造能产性 9

3.3 言之不足，故复言之：对举能产性 9

3.4 韵律之美：紧缩能产性 9

第四章　构式视角下的能产性解读 10

第五章　结语 11

参考文献 12

致谢 13

北京大学学位论文原创性声明和使用授权说明 14

# 第一章　引言

## 1.1 问题的提出

四字表达在汉语中占据着重要地位。粗看下，四字表达可分为两种：成语和常规短语。“人穷志短”是成语，需要收进词典；“中国制造”则用一个偏正式短语规则就能很好地处理。但仔细考察，则会遇到一些困难。例如：“人怂志短”[[1]](#footnote-1)是成语词，还是自由短语呢？

如果看作自由短语，组成这个短语的成分却并不都是词，如“志”一般不能单说；我们也会感觉到这些成分的组合并不是完全自由的，成分间结合的紧密程度也不一样。而如果看作词，这样的词“临时性”很强，不是那种会进入我们的“语言素材库”中、被我们频繁运用的词。

类似的表达数量非常之多，因为它们的表达效果非常强，甚至盖过了它们作为信息载体、传递信息的作用。随着个性的解放和互联网的发展，这种表达非常符合“网络虚拟社区”的交际需求：为了避免常规表述的乏味，引起听话者的兴趣，人们乐意创造并使用它们。

《现代汉语词典》，及一些前人的研究，已经通过 “格式”的概念，如“千…万…”，来规整这些表达，并启发人们创造。但我们认为，“套格式”[[2]](#footnote-2)并不能完全诠释四字表达能产的本质，例如对“悲极而泣”的解释力就极其有限。前人也有意无意提到了其他的一些能产机制，但都是作为补充，而未进行充分考察和量化。这就给我们带来的下述值得研究的问题：

四字表达在今天呈现出繁荣的态势，它到底有哪些能产机制？这些机制的能产程度又是怎样的？

## 1.2 选题背景及意义

未登录词识别是中文信息处理的重要一环，它能够提升自动分词的准确率，从而改善基于分词的任务的运行结果。对于“中国制造”类四字表达，可以很好地被分词程序处理为“中国/n 制造/v”；但对于“人怂志短”而言，很可能得到“人/n 怂/a 志/n 短/a”甚至更离谱的结果[[3]](#footnote-3)。我们的工作能够帮助机器识别一定数目的四字表达，因而能提升相关中文信息处理任务的性能。

（构式语法理论近年来受到越来越多的关注。作为认知学派的理论，它更加鼓励人们去关注语言使用上的细节，比如哪些话更常说，哪些话不常说。……）我们的工作也是一次构式语法理论在汉语中的实践，我们希望借助四字表达来阐述我们的构式观，以起到抛砖引玉的效果。

## 1.3 文献综述

有关四字表达的研究很丰富，光术语体系就分好几套：以（陆志韦，1956）为开端的“四字格”表述，以（文炼，1988）为开端的“类固定格式”表述，以（张国宪，1993）为开端的“对举格式”表述[[4]](#footnote-4)，以（周荐，1999）为开端的“待嵌格式”表述等等。此外，还有一些学者的相关术语也为后人借鉴，如（张拱贵，1985）提出的“语汇格式”，（李宇明，1999）提出的“词语模”，（邵敬敏，2008）提出的“框式结构”等。

从发表时间、研究角度和研究方法上进行比较，这些研究呈现出从定性到定量、从单一的结构主义理论到多元理论的阶段性特点，显示出学者对问题的认识逐渐深入、具体化。在1.3.1小节中，我们首先按术语体系对前人研究进行综述；在1.3.2小节中，我们再依阶段对前人研究进行总结；在1.3.3小节中，我们最后指出前人研究的不足之处。

1.3.1 不同的术语体系

“四字格”表述

（陆志韦，1956）首称所有完整的有意义的四字组合为“四字格”，并把“并立四字格”定义为内部紧凑、形式上为“2+2”并立的四字格。文章主要说明“并立四字格是词”的问题，理由是：并立四字格形式上不可拆分单说，内部结构的句法性减弱，不宜再在词性或句法结构上较真。对于一些半自由格式，作者则认为应特别小心，因为这些格式一旦能产，产出的实例的并立意味难免打了折扣，因为它们并非出于并立的构词方式被构造，而仅仅是一种套格式的行为。换言之，作者认为并立和套格式是两种能产方式。另外，作者还注意到一些并立的四字格的内部汉字顺序可以变换，譬如“开天辟地”能变换为“开辟天地”，从动宾式并立格变为动宾短语，但作者认为后者并非并立格，并未深究下去。

（姚殿芳，潘兆明，1985）采纳（吕叔湘，1963）的定义，把四言短语也纳入四字格范围。作者主要从汉语教学目的出发，分“认识四字格-创造四字格-恰当运用四字格”三步进行介绍，指出四字格生命力强大。作者主要提到两种创造四字格的方式：衬垫虚字和骈立，骈立又分两种，一种是生造，多半得利用古汉语单音节词；一种是套用现成的格式。作者认为，套用格式这种派生方式“体现了能产型骈立格的灵活性”。

（马国凡，1987）也提到了某些四字格的形成途径，如词的AXAB式扩展、词组的紧缩与节缩等，其中最具启发意义的是他称之为的成语的衍化格式。他认为成语的衍化有非固定的和框架格式衍化，前者如从“人云亦云”衍化出“‘帮’云亦云”（指帮派、派性）、“人云不云”（不随便附和）、“人倒亦倒”的说法，后者主要通过向一些“A＿B＿”式的框架中填入临时需要的成分构成。

“类固定短语”表述

（文炼，1988）把类固定短语定义为格式和功能上像成语、却是临时创造的、且含义可根据字面推断的汉语语言形式。作者指出，类固定短语不像固定短语那样能够作为语言材料供遣词造句，但其“组词成句”的规则又是很特别的，掌握这种格式对理解汉语是有帮助的。他具体举出四个包含单一固定汉字的类固定四字格式，并区分了形成该格式的成语和由该格式产生的类固定短语，并没有追求格式本身的意义，较好地诠释了“模仿”能产性的含义。作者还观察到一类特殊的现象，如“水落油出”[[5]](#footnote-5)，他称之为“仿造”，并认为这种仿造的成语必须依附被仿造的成语，才能显示出它的修辞效果。

（齐沪扬，2001）进一步强调类固定短语“结构模式固定”和“意义单一”两个特点，其中结构模式固定是指“类固定短语是是替换了成语中的一些组成成分后构成的”。作者根据替换成分的多少分成了“替换两个成分”和“替换三个成分”两类，前一类是（文炼，1988）没有涉及的部分，其中又进一步分出“保留数词”、“保留相同语素”和“保留同义或反义语素”三类。这种命名方式也体现出与（文炼，1988）一致的“成语-格式-类固定短语”能产模式。

（李传军，2008）同样出于辅助教学的目的，更系统地考察了类固定短语。他沿用齐沪扬的按替换成分分类方式，把类固定短语分为偶位内嵌型（A…C…）、奇位内嵌型（…B…D）、首尾内嵌型（…BC…）、一字支撑型（……C…，…B……）和组块内嵌型（其他型式）。他采用定量方法，“例不十，不立格”，提炼出300个格式，并统计了每个格式的不同实例[[6]](#footnote-6)数目，形成了“型式-格式-实例”三层计量体系，然而，他并未区分出成语，而把成语也看作一种类固定短语。此外，他还注意到同一格式下的类固定短语存在变换现象，例如“越生越穷”和“越穷越生”，并根据二者频率的差异进一步分出自由变换和有限变换。

“待嵌格式”表述

（周荐，1999）观察到《现代汉语词典》中存在这样一类特殊的词条，像“不…不…”、“千…万…”，他在该文的附注中首次采用“待嵌格式”来命名这一类词条，而《现代汉语词典》本身并未对这一类词条做出更多的解释。

（周荐，2001）认为，之所以设定此类条目，是因为其在口语中出现频繁，却不足以设定条目，以“待嵌格式”以纲带目，目的是引导人们用某种架构去类推创造出新的固定语。他认为，四字格是待嵌格式最适宜选择的格式，而且《现代汉语词典》中绝大多数格式都是“一三字显，二四字隐”。他认为，“一三字隐，二四字显”的格式同样存在，《现代汉语词典》中没有收录，应该是无法为之在音序序列中找到一个合适的位置的缘故，属于技术处理上的困难，他将这两类待嵌格式统称为交替显现[[7]](#footnote-7)待嵌格式。此外，他还举出《现代汉语词典》中收录的“自相……”一例，说明这类邻接显现的格式远不如交替显现的格式的能产性强。

（范喜梅，2009）重点分析了由成对方位词构成的待嵌格式。她认为，这些待嵌格式内部结构虽大多并列，却要遵循一种约定俗成的语序，如“上”在“下”前，“左”在“右”前。作者还指出，在并列关系之外，还存在转折、因果、目的等其他关系。此外，作者还为方位待嵌格式拟出了6个释义模板。

（孟祥英，2010）从汉语教学和词典编纂的角度出发，对待嵌格式进行了系统研究。她整理出240个固定字与填充字交替出现的待嵌格式，并根据固定字的词性组合、意义异同分为多个小类。作者指出，固定部分多属于基本词汇范畴，且多具丰厚的文化蕴涵（如“三”表多）。嵌入部分则与固定部分的词性、词义相关，有时候可以嵌入多种词性。作者把格式产生的语言机制归为“对称”、“类推”、“经济”三方面，为我们叙述能产性提供了很好的参考作用。

其他术语体系下的研究

（张拱贵，1985）中首次提出了语汇格式的概念。这篇文章是作者为朱林清《现代汉语“格式”初探》作的序，他指出，朱作中分析的格式都是语法格式，这些格式包含固定词语，一般充作关联词组构成短语，且使用上带有某种习惯性，不能用一般语法规则解释。他认为，朱林清对格式的研究是启发性的，但“‘格式’不仅语法里有，语汇里（成语、惯用语、谚语等）更多”。他举出自己于1963年发表的《成语中的成套格式》一文，说明此类格式在成语中的大量存在。

（冯胜利，1997）从韵律角度讨论了四字格式的成因，他提出了格套化理论，认为某些四字格式的形成源于双音节词的格套化，即将双音节词拆开，中间填入某种具有音缀性质的填充词而成。他认为，双音节词以[[＿A][＿B]]形式拆分居多，而填充词用久用熟以后，便能形成[X＿Y＿]式的格套，从而产生更多的词。

（李宇明，1999）提出了词语模的概念：具有批量产生新词语的框架，其中“模标”指不变的成分，而“模槽”指空位，例如“X盲”中，“盲”是模标，“X”是模槽。他还强调，词语模不仅仅是一种创造词语的模子，也是一种思维模式，模标具有一种“类”功能，例如人们习惯于把有目的、有计划、影响面广、意义重大的工作称为“X工程”，“X工程”便成了一个思维的模式。而他认为，词语模主要分前空型和后空型，中空型的词语模很少，如“吃X饭”。

（黄燕璇，2011）沿用（邵敬敏，2008）提出的框式结构概念分析了方位对举四字结构的特点，以便帮助人们成类地掌握这种结构，“减轻逐个记忆的负担”。她认为，这类四字结构中的方位词大多已虚化，虚化后才能形成“量多”这个结构义。此外，作者还提到成语的凝固性分两种，一种是来自典故的、一开始就凝固的结构，另一种则是经过繁衍、筛选、淘汰后形成的优势结构。因此诸如“东张西望”这样的凝固成语，从历史的角度也属于框式结构。

（韩立秋，2015）用构式语法分析了古汉语的凝固结构，提出了许多前人不曾涉及到的凝固结构，如“不亦…乎”、“何其……”等，作者把这些称为半实体性凝固结构。此外，作者引入“动态浮现语法”，强调语言一直处于一个动态过程中，“一旦临时结构所反映的认知经验得到了认同，并能够引起人们再次使用的兴趣时，凝固结构就很有可能在重复使用中发生变化”。他还提及“重新分析”机制，指出，“随着使用频率的提高，各个成分的本源义就会被遗忘，人们对其内部形式就可能做出重新分析”，我们认为这是非常重要的能产机制。

1.3.2 前人研究的阶段性特点

从定性到定量

“定量分析”是近年来的大趋势，也是青年学者的优势所在。2000年之后，相关的研究基本都带有“语料库”特色，且无一例外地选用了北京大学中国语言学研究中心（CCL）在线语料库，这可能与该语料库提供丰富的查询手段、能很好地查找结构不连续的语法和语汇格式有关。

在定量分析上，李传军的研究最为立体。他提出了“型式的能产性与格式的能产性并不是简单的正相关关系”观点，从“型式-格式-实例”三层对涉及的类固定短语进行了统计，清晰地描述了这类短语的总体分布情况。由于语料的大量性，学者们往往能更仔细、更全面地考察格式特点，也使得文中论点更具有说服力，使读者能更好地掌握格式用法。

从单一的结构主义理论到多元理论

语言理论的发展是另一个不可忽视的因素。较早的时候，由于对结构主义语言学还处于学习、探索的阶段，这一时期的文章多从句法层面立意，如陆志韦主要讨论了并立四字格是不是构词格，而文炼则旨在说明四字表达具有特别的短语性质。

而随着国外先进理论的引入，更多的青年学者开始从寻求更多途径来解释能产性。如孟祥英解释“类推”机制时提到了青年语法学派，韩立秋则引用Langacker定义的重新分析解释句法结构的演变等。多元理论的引入使得观察角度更加多样，使研究的对象变得更加立体，同样增加了结论的说服力。

总而言之，随着研究的深入，人们对四字表达逐渐建立起一些共识，如语法性质更偏向词、语义透明而表现力丰富等，人们开始用更精细的手段去研究它的特点。

1.3.3 前人研究的不足

以上阶段性特点，既是优点，也有不足。

在定量分析上，学者们只把精力放在“套格式”上，而仍用“仿造”、“类推”等术语定性地看待“人怂志短”这种能产模式。另一方面，所有的青年学者都把一个格式的不同实例数量等价于该格式的能产性，而忽略了实例在语料中的频率分布，而后者才是真正的能产性标准。可以说，前人的定量分析结论是很有限的，有很大的改进空间。

另一方面，前人对构式语法理论存在一定的认识误区。如韩立秋谈及构式时说，“因此，我们认为对于这样一种既非词亦非短语更非句子的特殊结构，最好的解释方法莫过于‘构式’”。孟祥英对构式同样存在类似理解：“无法用常规语法规则和从构成成分的语义来推知其意义”。把构式等价于“特殊的语法规则”，仍然是结构主义的分析模式，而没有体现出构式的认知基础。另外，许多学者都跟孟祥英一样，为了给构式一个独立的构式义，常常把词素甚至语境带来的意义“捆绑销售”，造成不必要的复杂和混乱，这也是没用很好地体会构式概念而产生的盲从做法。

我们认为，前人的不足一方面在于对能产性的认识不足，只考虑了一种能产方式。另一方面，在构式语法的运用上，他们大多只把构式看作一种“带有意义的框架”，本质上还是结构主义的思维，无法很好地说明能产性的本质。

## 1.4研究方法

## 1.5论文结构安排

# 第二章　现代汉语四字表达概述

## 2.1 四字表达的定义

（我们认为这种表达偏向词的性质，但不应该收入词典中[[8]](#footnote-8)。因为这种表达足够临时、足够低频，而意义又足够清晰，并且足够能产；收进词典，既起不到太多“正视听”的作用，更重要的是“收之不尽”，没有办法启发人们造出更多的表达。

（但我们认为这样的努力作用是很有限的。一方面，这些表达的意义是很直白的，非要提炼出一个格式义，只会把现象复杂化，我们也不相信人们在理解“千言万语”、“千刀万剐”的基础上会提炼出一个“千…万…”的上位概念；更重要的是，在缺乏语感积累的前提下，有多少人能在遇到一个急切的表达需求时，准确地想到并使用这些格式呢？）

## 2.2 四字格式的定义、性质及规模，兼谈半凝固构式

## 2.3 能产性？从词库扩展说开去

# 第三章　现代汉语四字表达的几种能产性

## 3.1不经意的错误：误用能产性

## 3.2 巧妙的引用：仿造能产性

## 3.3 言之不足，故复言之：对举能产性

## 3.4 韵律之美：紧缩能产性

# 第四章　构式视角下的能产性解读

# 第五章　结语

# 参考文献

陆志韦．汉语的并立四字格［Ａ］．//陆志韦．陆志韦集［Ｃ］．北京：中国社会科学出版社，2003: 400-457

吕叔湘．现代汉语单双音节问题初探［Ｊ］．北京：中国语文，1963（1）

郭绍虞．汉语语法修辞新探［Ｍ］．北京：商务印书馆，1979

姚殿芳，潘兆明．说“四字格”［Ｊ］．新疆：新疆大学学报（哲学社会科学版），1985（4）

文　炼．固定短语和类固定短语［Ｊ］．北京：世界汉语教学，1988（2）

齐沪扬．有关类固定短语的问题［Ｊ］．上海：修辞学习，2001（1）

范喜梅．由成对方位词构成的待嵌格式研究［Ｊ］．山东：现代语文，2009（11）

黄燕旋．方位对举四字框式结构研究［Ｄ］．广州：暨南大学，2011

王　娇．现代汉语对举嵌置式四字格习得研究［Ｄ］．南京：南京师范大学，2013

韩立秋．构式语法视野下古汉语凝固结构研究［Ｄ］．吉林：东北师范大学，2015

李传军．类固定短语相关问题研究［Ｄ］．上海：上海师范大学，2008

李宇明．词语模［Ａ］．//邢福义．汉语语法特点面面观［Ｃ］．北京：北京语言文化出版社，1999: 146-157

周　荐．双字组合与词典收条［Ｊ］．北京：中国语文，1999（4）.

周　荐．《现代汉语词典》中的待嵌格式［Ｊ］．北京：中国语文，2001（6）

马国凡．四字格论［Ｊ］．内蒙古师大学报，1987（3，4）

张拱贵．语法格式与语汇格式——《现代汉语“格式”初探》序［Ｊ］．吉林：汉语学习，1985（5）

孟祥英．汉语待嵌格式研究［Ｄ］．山东：山东师范大学，2010

# 致谢

# 北京大学学位论文原创性声明和使用授权说明

**原创性声明**

本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 日期： 年 月 日

**学位论文使用授权说明**

（必须装订在提交学校图书馆的印刷本）

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

* 按照学校要求提交学位论文的印刷本和电子版本；
* 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
* 学校可以采用影印、缩印、数字化或其它复制手段保存论文；
* 因某种特殊原因需要延迟发布学位论文电子版，授权学校□一年/□两年/□三年以后，在校园网上全文发布。

（保密论文在解密后遵守此规定）

论文作者签名： 导师签名：

日期： 年 月 日

1. 语出作者某同学微信朋友圈状态：“兴致冲冲跑去听大神报告，结果一瞄会议室里坐满了人还欢快地聊着，突然无端地脑袋一抽，拉开隔壁实验室的门搬砖去了……真是人怂志短，一念之间。” [↑](#footnote-ref-1)
2. 格式有两种，一种是具体的格式，即包含固定词素的格式，如“千…万…”；另一种是抽象的格式，例如“对举”格式，可以概括为“X1Y1X2Y2”，其中X1Y1、X2Y2的语法结构相同。这里说的“套格式”主要是针对前一种格式而言。 [↑](#footnote-ref-2)
3. 张华平博士提供的NLPIR的分词结果为“人/n 怂/vg 志/n 短/a”，其中vg代表“动词词素”，应该与“怂恿”相关，说明分词程序不理解“怂”作为形容词的含义。 [↑](#footnote-ref-3)
4. 张国宪这篇文章并不针对四字表达进行讨论，但后人研究多以其文为参考，沿用“对举”概念。另外，“对举格式”只是四字表达中的一部分，不能概括四字表达全貌，这与其他派系不同。 [↑](#footnote-ref-4)
5. 例句如下：“龙二井有油和水的矛盾，这是它的特殊性。周队长说，要促使矛盾转化，就要捞水，把水捞干。我们想一不做，二不休，搞它个水落油出。” [↑](#footnote-ref-5)
6. 原文采用的是“语例”的说法。 [↑](#footnote-ref-6)
7. 原文中称“两字交替显现，两字交替隐含”，我们简称为交替显现；同样地，后面的邻接显现，原文为“两字邻接显现，两字邻接隐含”。 [↑](#footnote-ref-7)
8. 当然，作为信息处理提高匹配效率和匹配准确率的手段，对于这种表达“见一个收一个”也无可非议。 [↑](#footnote-ref-8)