Modern Linguistics 现代语言学, 2025, 13(7), 212-220
Published Online July 2025 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2025.137701

Hans汉斯

述宾短语的扩展及其向量化计算

谢佳芸

西南交通大学人文学院，四川 成都

收稿日期：2025年5月29日；录用日期：2025年6月30日；发布日期：2025年7月15日

摘　要

本文以述宾短语的组配规律为研究对象，利用国家语委现代汉语语料库收集的述宾短语语料，使用已有的词向量模型，通过对比加入扩展成分前后的述语部分与宾语部分的相似度变化，证明了补语与定语在述宾短语组配中的重要作用。

关键词

述宾短语，定语，补语，词向量

The Extension of Predicate-Object Phrase and Its Vectorization Calculation

Jiayun Xie

School of Humanities, Southwest Jiaotong University, Chengdu Sichuan

Received: May 29th, 2025; accepted: Jun. 30th, 2025; published: Jul. 15th, 2025

Abstract

This paper takes the grouping rules of predicate object phrases as the research object, uses the predicate object phrase corpus collected from the Modern Chinese Corpus of the National Language Commission, and uses existing word vector models to compare the similarity changes between the predicate and object parts before and after adding extended components, proving the important role of complements and attributives in predicate object phrase grouping.

Keywords

Predicate-Object Phrase, Attribute, Complement, Word Vector

文章引用: 谢佳芸. 述宾短语的扩展及其向量化计算[J]. 现代语言学, 2025, 13(7): 212-220.
DOI: 10.12677/ml.2025.137701

1. 引言

自然语言处理(Natural Language Processing, NLP)作为计算机科学、人工智能与语言学的交叉领域，致力于使计算机能够理解、处理和生成人类自然语言。Word2vec 是 Mikolov 等人提出的一种神经网络概率语言模型[1]，可以用于计算词的词向量(Word Embedding)。词向量展示了语言学中的“价值”和“分布”等概念[2]，通过基于神经网络的方法，将词映射到一个高维空间中，使用上下文信息来学习每个单词的语义含义，把每个词表示为一个向量，使得具有相似语义的词在向量空间中距离较近。这种表征方式不仅广泛应用于文本分类、情感分析等任务，也为句法结构的量化研究提供了新的可能。

在汉语中，述语动词和宾语名词在词义和句法功能上都具有对立性，而词向量模型能够捕捉这种对立性。在理想情况下，述语动词和宾语名词的相似度应当趋近于 0，表明其语义特征分属不同维度，且在区别中又有联系。相似度过高可能暗示潜在的歧义，而过低则可能违反组配限制[3]。而定语与补语作为扩展成分，能够进一步优化述语与宾语的向量表征。

本文基于国家语委现代汉语语料库的述宾短语语料，利用预训练词向量模型，通过计算扩展成分加入前后述语与宾语的余弦相似度变化，考察其对述宾组配的调节作用。

2. 基于大型词向量模型的句法计算验证

本项目的语料来源为国家语委语料库，并使用 Python 语言和人工方法对语料进行提取与筛选，使用的词向量模型是一个大型的已训练好的开源词向量模型。

由于词向量通过语料中的上下文信息来学习每个词的含义，所以具有相似语义和语法功能的词在向量空间中距离较近，语义相似度较高(表 1)。

Table 1. Words with high semantic similarity
表 1. 高语义相似度词语

词	高语义相似度词语	语义相似度
火锅	涮锅	0.94207865
	酸辣粉	0.938524067
	凉面	0.937887967
	炸酱面	0.936975181
	米线	0.936187565
摧毁	毁掉	0.936736166
	击溃	0.935804784
	击垮	0.929788649
	击毁	0.924495697
	击破	0.916792333

大部分语义相似度高的词在词语意义和语法功能类型上相近或者相关，即与名词相似度高的也基本是名词，与动词相似度高的也基本是动词。因此，从动词和名词的语法组配来看，动词与名词的语法类

型不同，两者的语义相似度越小越好。当动词和名词的词义相似度偏高时，动词的“名词性”可能较强，或者名词的“动词性”可能较强，导致两词的组配不成立或有歧义。

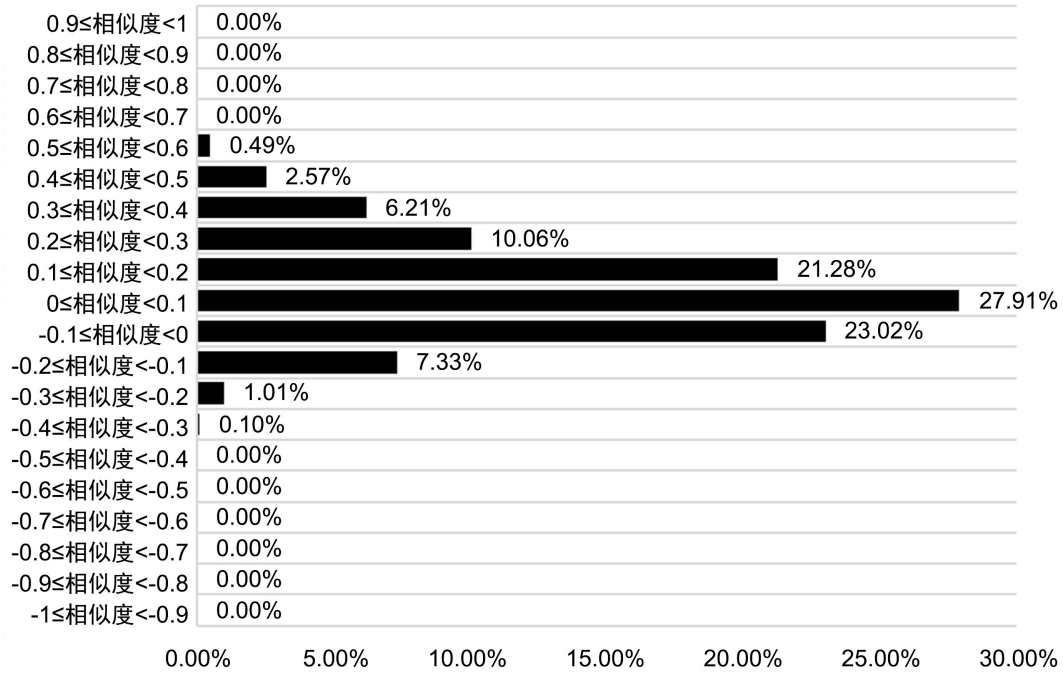

Figure 1. Similarity distribution between predicate verbs and object nouns
图 1. 述语动词与宾语名词的相似度分布

从图 1 可以看出，绝大部分的述语动词与宾语名词的语义相似度都在−0.3 到 0.5 这个区间内，最多集中于−0.1 到 0.2 这一区间。说明了述语动词与宾语名词有一定的相似性，但由于词义和句法功能的不同而又有明显区分。

如果在述宾短语的扩展中，插入的定语和补语能使述语部分和宾语部分的语义相似度进一步减小，就说明扩展部分的加入能够使述语动词与宾语名词的对立进一步扩大，就能够在词向量计算层面验证定语与补语对述宾组配过程中的影响作用。

以动词“吃”与名词“饭”的语义相似度计算为例：

假设“吃”的词向量为 $a=(a_1,a_2,\cdots,a_{128})$，“饭”的词向量为 $b=(b_1,b_2,\cdots,b_{128})$，其中第 i 个元素表示在第 i 维的取值，如 a_2 表示词“吃”的向量在第 2 维的取值。

那么它们的余弦相似度可以用以下公式计算：

$$\text{sim}(\text{吃},\text{饭})=\text{sim}(a,b)=\frac{\sum_{i=1}^{128}(a_i\cdot b_i)}{\sqrt{\sum_{i=1}^{128}(a_i)^2}\sqrt{\sum_{i=1}^{128}(b_i)^2}} \tag{1}$$

计算结果为动词“吃”与名词“饭”的词义相似度为 0.431009590625762，没有超过阈值 0.5。

下面通过对比述宾短语加入补语、定语前后的余弦相似度变化，来验证定语与补语在词向量层面对述宾组配的影响力。

2.1. 定语的向量计算

从名词来看，有些宾语名词在语义上比较抽象，是概括性较强、限定性和个体性较弱、表示种类范

畴的上位名词，常在与述语动词的组配中需要定语的扩展支持。在述宾短语中加入定语进行扩展，能够增加宾语名词的语义具体度，或者赋予宾语名词更多的、与动词词义相匹配的感情色彩，促使述宾短语组配的成立或消歧。同时，“名词前是否必带定语是由组配的两个成分共同决定”[4]，定语作为述宾短语的扩展成分虽然直接作用于名词，但其出现与否是由动词与名词的组配共同决定的。

在述宾短语间加入定语，常常能够减小述语动词与宾语部分之间的语义相似度。在 4624 条不重复的单定语的述宾短语语料中，有 2208 条述宾组配在加入定语后语义相似度减小，占比 47.75%，示例见表 2。表中原述宾相似度值是述语动词比宾语名词的相似度值，扩展后述宾相似度值是述语动词比定语加上宾语名词的宾语部分的相似度值[1]。

Table 2. Changes in similarity between predicate-object phrases after adding attributives
表 2. 述宾短语间加入定语后的相似度变化

述语动词	定语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
摆脱	花瓶	形象	−0.0500	−0.1596	−0.1096
驳斥	那些	不实之词	0.2664	0.076	−0.1904
采访	普通	大学生	0.1084	0.0099	−0.0985
藏有	一个	胶袋	0.1496	−0.0521	−0.2017
成为	打虎	英雄	0.1008	0.0526	−0.0482
吃	大	馍	0.4508	0.1167	−0.3341

这些例子都体现出定语对减小动词与名词语义相似度有作用，即对增强述宾组配能力有显著影响。下面以述宾短语组配能力的不同情况分别讨论定语对于述宾向量计算的作用。

2.1.1. 述宾可以直接组配但常带定语

在述宾短语的组配过程中，常常会有定语加入，对述宾短语进行内部扩展。有些述宾短语的直接组配能够成立，并且语料数量多，扩展成分主要起到丰富语义的作用，在句法上并不强制出现。另一些述宾短语，虽然有能够直接组配的例子，但更多时候常以内部扩展的形式出现。词向量计算可以验证定语对那些可以直接组配但常携带定语的述宾短语的扩展补充作用，示例见表 3。

Table 3. The effect of attributives on predicate-object phrases that can be directly combined but often carry attributives
表 3. 定语对可以直接组配但常携带定语的述宾短语的作用

述语动词	定语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
撤销	保险	合同	0.2074	0.0891	−0.1183
给予	他	折扣	0.0940	−0.0103	−0.1043
给予	定额	补助	0.2548	0.2179	−0.0369
核定	资金	定额	0.5080	0.2817	−0.2263
提高	泥浆	比重	0.1528	−0.0634	−0.2162

如“撤销保险合同”这一述宾组配，虽然述语动词与宾语名词可以直接组配为“撤销合同”这一述宾短语，但常常以携带定语的方式出现，加入定语能够使宾语名词更加具体。在词向量相似度计算中，加入定语名词“保险”后，述语动词“撤销”与宾语部分“保险合同”的语义相似度值降低为 0.0891，减小了 0.1183。

[1]相似度值均保留小数点后四位。

2.1.2. 述宾无法直接组配

有些述语动词与宾语名词是无法直接组配的，只在有定语或补语的扩展补充下才能够组配成为短语。林杏光认为词语组配的性质是“词汇·语法范畴”[5]，词语组配既受词性的制约，也受词义的制约，不是所有的动词都能和所有的名词组配。沈家煊认为，在空间上名词有“有界”和“无界”之分，而定语的加入能够将事物从“无界”转为“有界”[6]。定语可以使通常无法直接组配的动词和名词能够间接组配，而词向量计算可以验证定语对那些无法直接组配的述宾短语的作用，示例见表 4。

Table 4. The effect of attributives on predicate-object phrases that cannot be directly combined
表 4. 定语对无法直接组配的述宾短语的作用

述语动词	定语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
摆脱	文盲	状态	0.0523	0.0045	−0.0478
摆脱	病人	角色	0.0474	−0.0083	−0.0558
驳斥	这些	论调	0.3096	0.0742	−0.2354
阐释	马列	文论	0.3028	0.2860	−0.0168
成为	采血	对象	0.0604	0.0152	−0.0452

如“驳斥这些论调”这一述宾组配，述语动词“驳斥”是无法与宾语名词“论调”直接组配的，需要定语“这些”的扩展补充，依靠定语使宾语名词有定指。在词向量相似度计算中，述语动词“驳斥”与宾语名词“论调”的语义相似度为 0.3096，而加入定语 “这些”后，述语动词“驳斥”与宾语部分“这些论调”的语义相似度值为 0.0742，减小了 0.2354。

2.1.3. 述宾直接组配有歧义

有许多“动词 + 名词”构成的短语能够成立，但有述宾结构和定中结构两种歧义理解。如“学习教材”，既可以理解为“学习某个教材”，即述宾结构，也可以理解为“用于学习的教材”，即定中结构。李晋霞(2008)曾从名词的典型性、具体度和定指度，以及动词的典型性、常规用法等多种角度入手，讨论“双音节动词 + 双音节名词”的歧义的优先理解模式的制约因素。并基于袁毓林关于名词配价的研究，进一步从配价角度分析名词前带定语的现象，认为“二价名词充当宾语通常要求有限定成分的共现”，否则无标记的动宾组配容易优先激活定中关系[7]。定语的加入可以消解这类短语的歧义。词向量计算可以验证定语对那些可以直接组配但有歧义理解的述宾短语的消歧作用，示例见表 5。

Table 5. The effect of attributives on directly composing ambiguous predicate-object phrases
表 5. 定语对直接组配有歧义的述宾短语的作用

述语动词	定语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
测量	目标	数据	0.5322	0.5013	−0.0309
供给	日军	物资	0.5073	0.3141	−0.1932
规定	这些	标准	0.3939	0.1845	−0.2094
规定	检查	制度	0.5909	0.4248	−0.1661
规定	逃避	条款	0.5309	0.4133	−0.1177

如动名短语“测量数据”，既可以理解为定中短语，即“测量得到的数据”，如“我们得到了测量数据”；也可以理解为述宾短语，即“对某物的数据进行测量”，如“师傅正在测量数据”。从动词来看，动词“测量”是持续性动词，没有明显的动作界限，既可以作述语又可以作定语，作述语时需要组配意

义具体的宾语。从名词来看，名词“数据”是抽象且不具体的抽象名词，概括性较强，充当宾语时的语义不自足，不容易直接与动词“测量”组配为述宾短语，反而会激活动词作定语，使短语产生歧义。在词向量相似度计算中，述语动词“测量”与宾语名词“数据”的语义相似度为 0.5322，而加入定语名词“目标”后，述语动词“测量”与宾语部分“目标数据”的语义相似度值为 0.5013，减小了 0.0309。语义相似度值的减小说明述语部分与宾语部分的距离增加，增大了谓词和体词的功能距离，实现了加入定语消解歧义的作用。

2.2. 补语的向量计算

从动词来看，有些述语动词的动作界限较为模糊，特征性弱，缺少在时态等方面的意义，常在与宾语名词的组配中需要补语的扩展支持。在述宾短语中加入补语进行扩展，能够使动词的动作界限更加明确，从无界转为有界，从而使述宾短语的组配成立或消歧。同时，与定语相似，虽然补语作为扩展成分直接作用于述语动词，但“从语义上看，补语不一定都是指向述语的”[8]，补语既可能作用于述语动词(或者说动作的发出者)，也可能作用于宾语名词。

从词向量角度看，补语的加入通常能够减小述语动词与宾语部分之间的语义相似度。在 1164 条不重复的单补语的述宾短语语料中，有 578 条述宾组配在加入补语后语义相似度减小，占比 49.66%。

表 6 列举了述宾短语在加入补语前后相似度值的变化情况。下面仍然以述宾组配能力的不同情况讨论补语对于述宾短语向量计算的作用。

Table 6. Changes in similarity between predicate-object phrases after adding complements
表 6. 述宾短语间加入补语后的相似度变化

述语动词	补语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
挨	过	处分	0.2206	0.1310	−0.0896
吃	着	饭	0.4310	0.2104	−0.2206
吹	着	口哨	0.3416	0.1106	−0.2310
歌颂	着	英雄	0.2598	0.0798	−0.1800
驻扎	过	军队	0.4308	0.2911	−0.1398

2.2.1. 述宾可以直接组配但常带补语

可以与宾语名词直接组配但常带补语的述语动词，大都为动作的界限较为模糊、时间动程不够清晰的动词，需要补语的辅助。词向量计算可以验证补语对那些可以直接组配但常带补语的述宾短语的扩展补充作用，示例见表 7。

Table 7. The effect of complements on predicate-object phrases that can be directly combined but often carry attributes
表 7. 补语对可以直接组配但常携带定语的述宾短语的作用

述语动词	补语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
爆发	了	战争	0.1496	0.1178	−0.0318
闭	着	嘴巴	0.0342	0.0313	−0.0029
闭	着	眼	0.4489	0.4433	−0.0056
筹集	不到	钱	0.2795	0.1324	−0.1471
出席	了	盛典	0.5268	0.4450	−0.0817

如“爆发了战争”这一述宾组配，虽然述语动词与宾语名词可以直接组配为“爆发战争”这一述宾短语，但由于述语动词“爆发”是非持续性的，所以这一述宾短语常常以携带补语的方式出现。在词向量相似度计算中，述语动词“爆发”与宾语名词“战争”的语义相似度为0.1496，而加入补语“了”后，述语部分“爆发了”与宾语名词“战争”的语义相似度值为0.1178，减小了0.0318。

2.2.2. 述宾无法直接组配

与宾语名词相对，也有些动词经常与名词无法直接组配，或者组配结果倾向于被理解为定中短语，需要补语进行连接。词向量计算也可以验证补语对那些无法直接组配的述宾短语的连接作用，示例见表8。

Table 8. The effect of complements on predicate-object phrases that cannot be directly combined
表 8. 补语对无法直接组配的述宾短语的作用

述语动词	补语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
吃	出	名堂	−0.0289	−0.0734	−0.0445
撑	着	脑袋	−0.0430	−0.0527	−0.0097
荡	着	双腿	0.0999	0.0942	−0.0057
画	出	立体感	0.4824	0.4024	−0.0800
融化	成	水洼	0.4621	0.3923	−0.0697

如“吃出名堂”这一述宾组配，由于宾语名词“名堂”是述语动词“吃”的抽象结果，无法直接组配，需要结果补语“出”的扩展补充。在词向量相似度计算中，述语动词“吃”与宾语名词“名堂”的语义相似度为−0.0289，而加入补语“出”后，述语部分“吃出”与宾语名词“名堂”的语义相似度值为−0.0734，减小了0.0445。

2.2.3. 述宾直接组配有歧义

从动词来看，歧义产生主要与动词的动作性、持续性和使用频率相关。当动词既可以作述语，又可以作定语，且动作性较弱、为持续性动词，或日常生活使用频率较高时，就容易可能充当定语。而补语的加入能够丰富语义信息，使述宾短语的组配消歧。词向量计算还可以验证补语对那些可以直接组配但有歧义的述宾短语的消歧作用，示例见表9。

Table 9. The role of complement in directly forming ambiguous predicate-object phrases
表 9. 补语对直接组配有歧义的述宾短语的作用

述语动词	补语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
测量	着	数据	0.5322	0.4965	−0.0358
学习	着	笔记	0.4406	0.4139	−0.0267
研究	着	计划	0.5469	0.5309	−0.0160
注册	过	商标	0.6073	0.5386	−0.0687

如动名短语“学习笔记”，既可以理解为定中短语，即“学习过程中记录的笔记”，如“她在看学习笔记”；也可以理解为述宾短语，即“对已有的笔记进行学习”，如“老师让大家学习笔记”。从动词来看，动词“学习”是持续性动词，没有明显的动作界限，既可以作述语又可以作定语，作述语时需要组配意义具体的宾语。从名词来看，名词“笔记”是抽象且不具体的抽象名词，概括性较强，充当宾语时的语

义不自足，不容易直接与动词“学习”组配为述宾短语，反而使短语产生歧义。在词向量相似度计算中，述语动词“学习”与宾语名词“笔记”的语义相似度为 0.4406，而加入补语“着”后，述语部分“学习着”与宾语名词“笔记”的语义相似度值为 0.4139，减小了 0.0267。语义相似度值的减小说明述语部分与宾语部分的距离增加，即加入补语具有消解歧义的作用。

2.3. 定语和补语的综合向量计算

总体来说，定语与补语在述宾短语的扩展结构中，不仅能够使述宾短语所表达的语义更加丰富，同时还起到重要的使之成立或消歧作用，是许多述宾短语组配中不可缺少的一部分。在述宾短语间同时加入定语和补语，能够同时减小述语动词与宾语部分之间的语义相似度。在 1142 条不重复的单补语且单定语的述宾短语语料中，有 563 条述宾组配在加入补语和定语后相似度减小，占比 49.30%，示例见表 10。

Table 10. The function of adding both attributive and complement between predicate object phrases
表 10. 述宾短语间同时加入定语和补语的作用

述语动词	补语	定语	宾语名词	述宾相似度值	扩展后述宾相似度值	相似度变化
吃	了	正宗	火锅	0.4941	0.3669	−0.1272
采访	了	有关	专家	0.2639	0.1183	−0.1457
拨	出	这个	号码	0.3709	0.0676	−0.3033
给予	了	肯定	回答	0.2086	−0.0812	−0.2899
拨	出	一点	空儿	0.1516	−0.1163	−0.2679

如“采访了有关专家”这一述宾组配，在直接组配为“采访专家”时是有歧义的，但加入补语和定语后就消解了歧义，只能被理解为述宾短语。在词向量相似度的计算中，述语动词“采访”与宾语名词“专家”的语义相似度为 0.2639，而加入补语“了”以及定语“正宗”后，述语部分“采访了”与宾语部分“有关专家”的相似度值为 0.1183，减小了 0.1457，体现出补语和定语同时出现对减小动词、名词相似度的作用，即对识别述宾组配有作用。

3. 总结

本文使用已训练好的大型词向量模型，尝试对述宾短语扩展组配的句法规律进行向量化表征，使用余弦相似度的计算方法考察述宾短语中扩展成分的作用。研究将述宾短语语料划分为单定语、单补语两大类，并在每类中进一步区分可直接组配但常带扩展成分、无法直接组配以及直接组配有歧义三种情况，最后对同时包含定语和补语的复杂结构进行了综合考察。

从相似度的数值来看，词汇意义与句法功能越接近的词，相似度越高。而述语动词与宾语名词是句法功能对立的两类词，相似度在一定范围内应当越低越好。统计分析表明，绝大多数述宾短语的述宾相似度值集中在−0.3 至 0.5 区间内。在加入定语或补语作为扩展成分后，有近一半的述宾短语的相似度下降，更接近理想的述语宾语相似度分布区间。

通过词向量与余弦相似度计算，本文对述宾短语扩展组配规律进行了考察，证明了定语与补语作为扩展部分的加入能够使述语动词与宾语名词的对立进一步扩大，在词向量计算层面定语与补语作为扩展成分对述宾组配的重要作用。

参考文献

[1] Mikolov, T., Chen, K., Corrado, G., *et al*. (2014) Efficient Estimation of Word Representations in Vector Space.

http://arxiv.org/abs/1301.3781v3

[2] 冯志伟. 词向量及其在自然语言处理中的应用[J]. 外语电化教学, 2019(1): 3-11.

[3] 陆旭, Aleksandr Mitkov, 冉启斌. 从词向量计算看汉语名词和动词的关系[J]. 语言教学与研究, 2024(3): 57-67.

[4] 辛平. 基于语料库的动宾组配中定语受限问题研究[J]. 汉语学习, 2013(4): 87-91.

[5] 林杏光. 词语搭配的性质与研究[J]. 汉语学习, 1990(1): 7-13.

[6] 沈家煊. “有界”与“无界” [J]. 中国语文, 1995(5): 367-380.

[7] 李晋霞. 现代汉语动词直接做定语研究[M]. 北京: 商务印书馆, 2008: 39-43+121-142.

[8] 马真, 陆俭明. 形容词作结果补语情况考察(一) [J]. 汉语学习, 1997(1): 3-7.