

汉语合成复合词的词库和句法界面特征

沈阳、顾阳

北京大学中文系 / 香港中文大学现代语言及文化系
syshen@pku.edu.cn / yanggu@cuhk.edu.hk

摘要: 本文根据“词汇的完整性理论”证明合成复合词从内部组合特征和外部扩展特征两方面都不同于词组结构形式, 即与词汇不同层次的句法操作不能进入词语内部, 合成复合词的结构变化处在词库和句法之间的界面层次上。合成复合词的结构性质分析对于中文信息处理中讨论较多的词语切分操作提供了一种理论依据。

关键词: 合成复合词, 内部组合, 外部扩展, 词汇完整性, 词汇与句法的界面特征

The Interface Features between Lexicon and Syntax of the Synthetic Compound in Chinese

Shen Yang, Gu Yang

Peking University / Chinese University of Hong Kong
syshen@pku.edu.cn / yanggu@cuhk.edu.hk

Abstract: The paper provides evidence with the theory of Lexical Completeness to show that the various synthetic compounds are manifestations of relevant interface properties of the lexicon and syntax. The discussion captures the differentiation between word structure and sentence structure. It would also provide one of the theoretical bases for the line of demarcation of the Chinese words in the Chinese computational processing.

Key words: synthetic compound, lexical completeness, interface of lexicon and syntax

0. 合成复合词的结构性质问题

计算机中文信息处理中一个重要问题就是词语的切分, 其中对包含若干个词的较长语言片段如何切分更是令人头疼的问题。本文要讨论的“合成复合词”就是指“汽车修理工 / 工人”、“汽车修理厂 / 工具”这一类包含较多词语的复杂语言片段。这类语言片段的组成成分是词, 表面上看也可以按照句法结构形式来分析, 比如说是“主谓 / 定中”偏正词组或“定中 / 定中”偏正词组。但我们认为, 虽然合成复合词不妨可以看作是通过“S (施事或其他域外论元) - V (动词) - O (客体)”这样的论元结构变化形成的, 但这种变化其实还是一种词汇结构变化。即合成复合词不是在句法中形成的, 而只可能是句法之前的词汇层面结构变化的结果。因此这种语言片段还是应该看作一个词, 或者说按照一个词语单位来处理。支持这种分析的一个证明就是合成复合词只具有词的性质而并没有词组的性质, 或者说其结构性质只反映词库和句法之间的界面特征!

1. 合成复合词结构的传统分析

应该承认，说汉语简单的复合词是词有时已经有点麻烦；而合成复合词甚至包含若干个词，要证明它整体还是词当然就更困难些。目前定义一个较长的语言片段（除专有名词外）是不是词不外两种办法：一种办法是看语言片段中有没有不成词语素，即组成成分是否都是词。比如“服装设计师/图”中“师、图”不成词，所以整个是词；而“服装设计大师/草图”中“大师、草图”是词，而且“服装、设计”也都是词，所以整个是词组。还有一种办法是看语言片段能否扩展，即组成成分之间能不能插入结构助词“的”。比如“服装设计”与“师、图”中不能插入“的”，所以整个是词；而“服装设计”和“大师、草图”中好像可以插入“的”，所以整个就是词组。这两种办法看起来简便，其实都并不可靠。比如按照“成词说”就要把“服装设计师/图”和“服装设计大师/草图”分成两类。其实前者是词并不能证明后者不是词，因为很难说组成成分成词不成词就是词与词组的本质区别。更不用说像“服装设计人/奖”这种包含单音节成词中心语的片段就又可能无从归属。又如按照“插入说”是把不能插入“的”的片段和能插入“的”的片段看作两类。其实有“的”片段肯定是词组并不能证明没“的”片段也是词组，因为“能不能”加“的”也并不是词和词组的本质区别，而关键在于“加没加”，即能加“的”而没加还可能是词，只有加了“的”后才一定是词组。更不用说如果只考虑能不能，像刚才还肯定是词的“服装设计师/图”，因为也能在某个位置加上“的”（服装的设计师/图），其原来形式就成了词组了，这同样说不通（实际都是加“的”后才成为词组）²。可见确认合成复合词是词组还是词，仅靠这两条并不够，还应该更多注意词组和词在语法结构性质上的对立。这种结构性质可以从合成复合词的内部组合特征和外部扩展特征两方面来分析。

2. 合成复合词的内部组合特征

关于复合词的语法结构性质，朱德熙（1982）曾有一个论断，即“汉语复合词的组成成分之间的结构关系基本上和句法结构关系一致”。对这种说法要看怎么理解。事实上朱先生紧接来说的一句话就是“句法结构关系有主谓、述宾、述补、偏正、联合等，绝大部分复合词也是按照这几类结构关系组成的”。可见如果只看复合词的表面组合形式，那确实都可以分析成这样几种关系，说复合词和词组的内部结构性质一致当然就不错。但如果进一步观察成分组合的其他各种情况，那复合词和词组的结构性质就可能很不一致了。

比如“汽车修理工/工具”这样的语言片段，如果作层次分析都是“[偏正[主谓SP]NP]”，即修饰语“汽车修理”可以分析为“主谓”。笼统地说，主谓结构在词组里面当然也可以作修饰语，如“老王修理的汽车”中“老王修理”就是主谓结构。但需要注意“老王修理”这样的主谓结构和“汽车修理”这样的主谓结构有一个明显区别，即前者是“正装（S+V）”主谓，后者是述宾“倒装（O+V）”构成的主谓。事实上词组中主谓修饰语只有前一种形式，复合词中主谓修饰语却只有后一种形式。换种情况看也一样，如果说“修理汽车”是正装述宾，“汽车修理”是倒装述宾，那么作词组中修饰语只能用正装形式，作复合词中修饰语就只能用倒装形式（双音节述宾式动词除外）。这就是词组和复合词在内部结构

上的一种对立现象，或者说这一点就可以作为区别词组和合成复合词的一个重要标准。比较下面的例子：³

(1) A. 合成复合词:

- a1. 汽车修理工 / 工具 * 修理汽车工 / 工具
- a2. 房屋装修图 / 公司 * 装修房屋图 / 公司
- a3. 文物走私犯 / 手法 * 走私文物犯 / 手法

B. 偏正词组:

- b1. 修理汽车的工人 / 工具 * 汽车修理的工人 / 工具
- b2. 装修房屋的图纸 / 公司 * 房屋装修的图纸 / 公司
- b3. 走私文物的罪犯 / 手法 ? 文物走私的罪犯 / 手法

再比如词组中有定中偏正词组，合成复合词也是定中结构，二者似乎一致。但以同样表“人”的中心语为例：一般词组通常可以替换使用各种具体词语，即偏正词组在组合上只受句法规则限制，不受词汇本身的限制。而合成复合词的施事中心语却往往与修饰语中的动词有搭配选择，而且没有什么规律性，这很难说是靠句法规则控制的。就像简单复合词的组合形式也往往有特定的语素选择，很难用句法关系来说明一样，这种情况也只能看作是构词层面才存在的现象。这一点就又是合成复合词在结构形式上有别于词组的一个特点，比较：

(2) A. 合成复合词:

- a1. 汽车修理工 / *员 / *者 / *家 / *师
- a2. 汽车驾驶员 / ?者 / *工 / *家 / *师
- a3. 汽车经销商 / ?员 / *工 / *家 / *师

B. 偏正词组:

- b1. 修理（修）汽车的人 / 师傅 / 专家 / 老板 / 朋友 / 小王
- b2. 驾驶（开）汽车的人 / 师傅 / 专家 / 老板 / 朋友 / 小王
- b3. 经销（卖）汽车的人 / 师傅 / 专家 / 老板 / 朋友 / 小王

3. 合成复合词的外部扩展特征

如果说上述内部结构形式的差异特征还不足以说明合成复合词有别于词组，那么退一步说，即使假定合成复合词只是一种内部结构不同和有词语选择限制的词组，那至少也应有理由期待它的外部句法表现跟词组一致。事实上这方面合成复合词跟词组还是有明显不同。

先看跟数量词或定指词组合的情况。汉语的名词表无定指涉往往前面带数量词，表有定指涉往往前面再带定指词。需要注意的是，如果是包含多项名词成分的名词词组，那么前面带的数量词和定指词应可能分别指向词组中的不同成分，即理论上说词组中的任一名词成分都可以独立扩展成指涉性词组。因此下面（3）中第一组例子不但本来就可以分析成两种意义的结构形式（替换量词可证明）；而且还因为词组不是一个严格整体结构，甚至可以把量词直接放到词组中被修饰的成分前（不必替换量词）而突出特定的指涉意义。例如：

(3) a. 一个工人的建议

- = a1. [[一位工人]的建议] = a2. [一条[工人的建议]] / [工人的[一个建议]]
- b. 几个公司的老总
- = b1. [[几家公司]的老总] = b2. [几位[公司的老总]] / [公司的[几个老总]]

- c.那个客房的壁灯
 = c1.[那间客房的壁灯] = c2.[那盏[客房的壁灯]] / [客房的[那个壁灯]]

但合成复合词就不能这样。一方面合成复合词前面的数量词和定指词只能指向中心语名词，不能指向其他组成成分。如“一个仓库保管员”，始终只能得到“一个保管员”的意思，而得不到“一个仓库”的意思⁴；如果硬换成跟非中心语成分搭配的量词，就会造成不合格的结构。另一方面因为合成复合词是一个严格整体，所以即便数量词和定指词只修饰中心语，也不能直接插入到这个成分前，否则也会造成不合格的结构。可以跟（3）作比较：

- (4) a1.一个汽车修理工 (= 一个 /位修理工)
 a2.*一辆汽车修理工 a3.*汽车一个 /位修理工
 b1.两个眼膜捐献者 (= 两个 /位捐献者)
 b2.*两只眼膜捐献者 b3.*眼膜两个 /位捐献者
 c1.这个服装设计师 (= 这个 /位设计师)
 c2.*这套服装设计师 c3.*服装这个 /位设计师

下面例子中的形容词好像既可以修饰整个合成复合词，也可以修饰复合词中的某个成分。比较：

- (5) a.老仓库保管员
 = a1.老 /仓库保管员 (老保管员) = a2.老仓库 /保管员 (老仓库)
 b.高级汽车修理工
 = b1.高级 /汽车修理工 (高级修理工) = b2.高级汽车 /修理工 (高级汽车)

其实（5）并不构成上面说的情况的反例。前者不难理解，比如“仓库保管员”，作为一个名词当然可以受“老”修饰，就像可以受数量词修饰一样，即整个片段变成了类似“一个仓库保管员”那样的词组。后者中“老仓库”则合起来只是复合词中“保管”支配的一个域内论元，即是一个名词，这样整个片段就还是一个合成复合词，这跟“军用仓库保管员”、“重型汽车修理工”等例子类似。换句话说“老仓库保管员”既可能是词组，这时具有（3）的组合特征；也可能是合成复合词，这时就具有（4）的组合特征。分别看都不会有例外。

再看跟时态成分和状语成分组合的情况。汉语包含动词的名词词组，一般说既可能插入表时貌情态的助词，也可能带上表方式程度的副词、形容词，即词组中的动词也可以扩展成有时态特征或单独受修饰的动词词组。例如：

- (6) a.修理汽车的工人
 a1.修理着汽车的工人 a2.正修理汽车的工人 a3.紧张地修理着汽车的工人
 b.继承财产的子女
 b1.继承了财产的子女 b2.已继承财产的子女 b3.意外地继承了财产的子女
 c.走私文物的团伙
 c1.走私过文物的团伙 c2.将走私文物的团伙 c3.偷偷地走私过文物的团伙

但合成复合词就不能这样。一方面尽管合成复合词内部有动词，但却不能采用任何时态形式，可见这种动词像词中的动词性语素一样，不具有词组中的动词那种体现时间空间范畴的能力。另一方面同样因为合成复合词是一个严格的整体，如同不能插入专门修饰名词的成分一样，也不能插入专门修饰动词的成分。下面都是不合格的结构。跟（6）比较：

- (7) a. 汽车修理工
 a1. *汽车修理着工 a2. *汽车正修理工 a3. *汽车紧张地修理着工
 b. 财产继承人
 b1. *财产继承了人 b2. *财产已继承人 b3. *财产意外地继承了人
 c. 文物走私团伙
 c1. *文物走私过团伙 c2. *文物将走私团伙 c3. *文物偷偷地走私过团伙

下面例子中合成复合词前加上的词语成分，好像是修饰其中的动词的。比较：

- (8) a. 熟练的修理工（熟练修理）
 b. 合法的继承人（合法继承）
 c. 疯狂的走私团伙（疯狂走私）

其实即使不排除（8）中复合词前面的成分可以修饰动词，也可以肯定在这里并非修饰动词，至少不是专门修饰动词的。因为像“熟练、合法、疯狂”这样的词本来就既可修饰动词也可修饰名词，后者例子如“熟练工种”、“合法行为”、“疯狂少年”等；况且上例复合词的修饰语位置也绝不能出现专门修饰动词的“偷偷”、“曾经”类副词。可见这些词语是修饰整个复合词的（修饰语后面只能用“的”也是证明）。换个角度说，“熟练的汽车修理工”是“熟练”和“汽车修理工”组成的词组，“熟练”也不是进入了复合词的成分。

4. 词汇完整性理论与复合词构造的界面特征

上面所说的复合词外部组合受限制的现象在理论语言学中称作“词汇完整性”。其原则是说词语是一个完整的整体，与词汇不同层次的句法操作不得进入词语内部⁵。说得更明确一点就是：句法结构（如词组）的组成成分最小也是词组，也就是即使由单个的词构成的词组中的“词”还是具有词组特性，因为都可以独立扩展为词组；但词汇结构（如复合词）的组成成分最大也只是词，也就是即使作为合成复合词组成成分的那种“词”也不具有词组的特性，因为都不可能独立扩展为词组⁶。前面讨论的词组和合成复合词的组成成分（包括名词成分和动词成分）在扩展为独立受修饰的词组方面的区别，实际上就反映了这种情况。这也就论证了本文开头说的观点：合成复合词从内到外都不等同于词组的结构形式，因此其构造只可能是词层面的变化。同时又表明，即使承认合成复合词体现某种结构特征，充其量也只是说它比单个词语更复杂，却仍未进入句法层面，也就是说合成复合词所反映的论元结构变化只处在词库和句法之间的层面上。这也是合成复合词构造的一个基本特性。

本文虽然只讨论了定中结构形式合成复合词这样一种包含若干个词的复杂语言片段的结构性质，但实际上这一结论也完全适用于分析和确定其他复杂词语片段的性质。换句话说，这可能为计算机中文信息处理中的分词操作提供了一种理论上的依据。

附 注：

¹ Roeper和Siegel（1978）最早分析复合词的内部结构，将其定义为“词内句法”（word-internal syntax），并将其归属为论元结构的范畴。

² 我们同意把带“的”字修饰语的片断都看作名词词组的定义。详见石定栩（Shi, 1999）的讨论及所引文献。

³ 这一点参考了陆俭明先生的观点（不过他主要论证词组和句子的语序差别，相关论文尚未发表）。前述“服装设计的大师/图纸”好像可以说。不过作为词组肯定不如“设计服装的大师/图纸”好。此外如“服装的设计师”这种词组，不是主谓结构作修饰语，并不影响上述结论。

⁴ 注意这类例子不要与“一个仓库的保管员”混淆，后者可以有“一个仓库”的意思。但包含带“的”修饰语 $\frac{1}{4}$ 的实际上就是典型的词组了，此例的结构和意义相当于（3）。

⁵ 参看Chomsky（1970），Williams（1978），Lapointe（1979）。

⁶ 这一观点在Hale和Keyser（1999）的讨论中又有所强调。

参 考 文 献

- [01] 顾 阳（1994）论元结构理论介绍，《国外语言学》第1期，1-11页。
- [02] 顾 阳（1996）生成语法及词库中动词的一些特征，《国外语言学》第3期，1-16页。
- [03] 吕叔湘（1942）《中国文法要略》，上海：商务印书馆；修订本1956年，上海：商务印书馆。
- [04] 沈阳等（1995，主编）《现代汉语配价语法研究》，北京：北京大学出版社。
- [05] 沈 阳（1998）带方位处所宾语的动词及相关句式，《语言学论丛》第20辑，北京：商务印书馆。
- [06] 汤廷池（1992）《汉语词法句法三集》，台湾：学生书局。
- [07] 汤廷池、张淑敏（1996）论旨网格、原参语法与机器翻译，《中国语文》第4期，260-272页。
- [08] 徐烈炯、沈阳（1998）题元理论与汉语配价问题，《当代语言学》第3期，1-21页。
- [09] 朱德熙（1982）《语法讲义》，北京：商务印书馆。
- [10] Anderson, S. R. (1992) *A-Morphous Morphology*. Cambridge, Massachusetts: MIT Press.
- [11] Aronoff, M. (1994) *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, Massachusetts: MIT Press.
- [12] Baker, M. (1988) *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.
- [13] Bauer, L. (1983) *English Word Formation*. Cambridge, Massachusetts: MIT Press.
- [14] Botha, R. P. (1984) *Morphological Mechanisms: Lexicalist Analysis of Synthetic Compounds*. Oxford: Pergamon.
- [15] Chomsky, N. (1970) Remarks on Nominalization. In R. Jacobs & P. Rosenbaum, eds. *Readings in English Transformational Grammar*. Cambridge, Massachusetts: Ginn, Waltham.
- [16] Di Sciullo A. M., and E. Williams. (1987) *On the Definition of Words*. Cambridge, Massachusetts: MIT Press.
- [17] Hale, K. and S. J. Keyser. (1993) On Argument Structure and the Lexical Expression of Syntactic Relations. In K. Hale & S. J. Keyser, eds. *The View from Building 20*. Cambridge, Massachusetts: MIT Press.
- [18] Hale, K. and S. J. Keyser. (1999) A Response to Fodor and Lepore, "Impossible Words?" *Linguistic Inquiry* 30, 453-466.
- [19] Lapointe, S. (1979) *A Theory of Grammatical Agreement*. Doctoral dissertation. Cambridge, Massachusetts: MIT Press.
- [20] Matthews, P. H. (1991) *Morphology*. 2nd ed. Cambridge, Massachusetts: MIT Press.
- [21] Roeper, T. (1987) Implicit Arguments and the Head-complement Relation. *Linguistic Inquiry* 18, 267-310.
- [22] Roeper, T. And M. Siegel. (1978) A Lexical Transformation for Verbal Compounds. *Linguistic Inquiry* 9, 199-260.
- [23] Selkirt, E. (1982) Prosodic Domains in Phonology. In E. Selkirk ed. *The Syntax of Words*. Cambridge, Massachusetts: MIT Press.
- [24] Shi, D. X. (1999) Verbal Modifier of Chinese Nominals. Paper presented at the 8th Meeting of International Association of Chinese Linguistics. University of Melbourne, Australia.
- [25] Spencer, A. (1991) *Morphological Theory*. Blackwell.
- [26] Williams, E. (1978) Notes on Lexical Theory. Ms. Amherst: University of Massachusetts.
- [27] Williams, E. (1981a) Argument Structure and Morphology. *The Linguistic Review* 1, 81-114.
- [28] Williams, E. (1981b) On the Notions "Lexically Related" and "Head of a Word." *Linguistic Inquiry* 12, 245-274.