

“理论词”和“语素”的概念在语言学上的严重缺陷

冯志伟

我们把语言学上的词叫做“理论词”(theoretical word)，这样的理论词的概念，在语言学理论上与语素(morpheme)和词组(phrase)划水难分，存在着严重的缺陷。这样的严重缺陷，阻碍了我国语言研究的进步，也阻碍了我国自然语言处理的发展。

其实这个问题我早在20世纪80年代就认识到了，曾经在一些学术会议上提出过，可是，人微言轻，几乎无人注意到，现在，我国的现代汉语教学因为这个问题步履维艰，我国自然语言处理中单词的自动切分研究因为这个问题进展甚微。我越来越感到这个问题实在太严重了；如鲠在喉，不吐不快。因此再次讨论这个问题，希望引起学术界的关注。

在语言学中把语素分为自由语素和黏附语素两大类。“自由语素”是活动能力很强、不仅可以与其它语素组合成词、而且还可以单独成词使用的语言中的最小的造句单位。例如，“地、跑、红”等都是自由语素。黏附语素的活动能力不强，不能单独成词，它们要与其它的语素相组合而成词。如“机、劳、老、小、子、者、然”。

语言学中又把词分为单纯词与合成词两大类。单纯词是由一个语素构成的词。合成词是由两个或两个以上的语素构成的词。

由于单纯词只由一个语素构成，所以，这个构成单纯词的语素必定是自由语素。这样一来，“地、跑、红”等都是自由语素，它们同时又可以看成单纯词。从语素的角度看是自由语素，从词的角度看是单纯词。观察的角度不一样，名称不同，实质则是一样的。

在语素与词这两个集合之间，有一个交集(intersection)，这个交集就是自由语素，如果从词的角度看，它们又可以叫做单纯词。

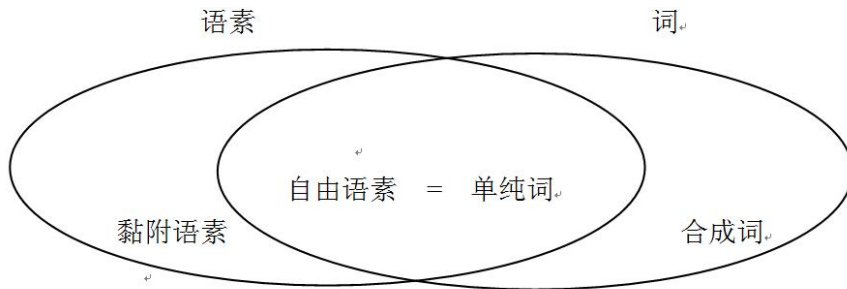


图1 语素和词之间的交集

由此可见，黏附语素和词之间的界限是可以区分清楚的，黏附语素绝不可能是词；语素和合成词之间的界限也是可以区分清楚的，合成词不可能是单个的语素。然而，在语素和词之间有一个交集，这个交集，从语素的角度看是自由语素，从词的角度看是单纯词。由于自由语素和单纯词名异而实同，导致了合成词和词组之间的界限不清。这是汉语语言学本身的内部矛盾，也是汉语语言学在理论上的一个硬伤。

下面我们进一步从结构方面说明这种界限不清的情况。

合成词的构成方式与词组的构成方式有许多一致的地方。

由语素和语素组成的合成词，构成方式主要有以下 7 种：

① 并列式：两个语素并列在一起组成合成词，形成一种并列关系。例如，

朋友，泥土，松散，东西，开关。

② 偏正式：合成词中的两个语素有主有从，后一个语素为主体，前一个语素修饰或限制后一个语素，形成一种偏正关系。例如，

火车，工业，微笑，雪白，飞快。

③ 支配式：合成词中的两个语素，前一个表示动作，后一个表示动作涉及的事物，形成一种支配和被支配的关系。例如，

领队，司机，主席，签名，悦耳。

④ 补充式：合成词中的两个语素，前一个表示动作，后一个补充说明动作的结果，形成一种补充关系。例如，

提高，说明，扩大，缩小，改善。

⑤ 陈述式：合成词中的两个语素，前一个是陈述的对象，后一个是陈述的内容，形成一种陈述和被陈述的关系。例如，

地震，日食，祖传，法定，雪崩。

⑥ 附加式：合成词中的两个语素，只有一个表示实在的意义，另一个不表示实在的意义，只是作为一个辅助成分，附加在表示实在意义的语素之前或之后，形成前缀或后缀。

老虎，第一，非法，桌子，椅子。

⑦ 重叠式：合成词是由单音节语素重叠而构成的。例如，

星星，茫茫，纷纷，巍巍，翩翩。

词组是由单词和单词组合而成的。

汉语词组的构成方式主要有以下 7 种：

① 联合结构：词组中的两个词是并列的，形成一种并列关系。例如，

风俗习惯，调查研究，勤劳勇敢，植树造林，轻松愉快。

② 偏正结构：词组中的两个词，前一个是修饰语，后一个是中心语，形成一种偏正关系。例如，

茫茫大海，劳动热情，衷心祝贺，十分灵敏，热烈欢迎。

③ 述宾结构：词组中的两个词，前一个是述语，后一个是宾语，形成一种述语对宾语的支配关系。例如，

热爱祖国，欣赏音乐，发射导弹，供养父母，行使职权。

④ 述补结构：词组中的两个词，前一个是述语，后一个是补语，形成一种补充关系。例如，

解释清楚，举起来，洗干净，讲明白，扔出去。

⑤ 主谓结构：词组中的两个词，前一个是主语，后一个是谓语，形成一种陈述关系。例如，

小孩咳嗽，姑娘唱歌，天气热，月亮圆，今天星期日。

⑥ 附加结构：“的”字结构”和“所字结构”都可以看成是附加了“的”字或“所”字的结构，形成一种附加关系。例如，

当兵的，掌柜的，当家的，所看到，所研究，所驱使。

由后缀“者”构成的一些长结构（long construction）也可以看成附加结构的词组，例如，

屡教不改者，成绩不合格者，申请移民者，患心脏病者，诺贝尔奖金获得者

⑦ 重叠结构：词组中的两个词，后一个词是前一个词的重叠，形成一种重叠关系。例如，

研究研究，练习练习，讨论讨论，总结总结，复习复习。

可以看出，汉语的合成词与词组的构成方式存在着整齐的对应，而且每种对应的结构所表示的关系是相同的。

这种对应关系可列表比较如下：

合成词	词组	表示的关系
并列式	联合结构	并列关系
偏正式	偏正结构	偏正关系
支配式	述宾结构	支配关系
补充式	述补结构	补充关系
陈述式	主谓结构	陈述关系
附加式	附加结构	附加关系
重叠式	重叠结构	重叠关系

图2 合成词与词组的对应关系比较

合成词的构成方式与词组的构成方式的这种一致性，使得汉语的语法规则易学易记，对汉语的学习是有好处的，可是，这种一致性也往往导致合成词与词组的界限不甚分明，使我们难于判断一个结构究竟是合成词还是词组。

如果一个结构由两个黏附语素构成，必定是合成词，不可能是词组。例如，“劳”是黏附语素，“损”也是黏附语素，它们结合而成的“劳损”必是合成词，不可能是词组。

如果一个结构由一个黏附语素和一个自由语素构成，必定是合成词，不可能是词组。例如，“劳”是黏附语素，“动”是自由语素，它们结合而成的“劳动”必定是合成词，不可能是词组。

含有前缀的“老师”，“老虎”等结构，也必定是合成词，不可能是词组，因为前缀是黏附语素。

含有后缀的结构“桌子”，“作者”，“忽然”，除了后缀“者”有时可以附加在多音节结构之后构成词组之外，在一般情况下，也必定是合成词，不可能是词组，因为后缀是黏附语素。

但是，如果一个结构由两个自由语素组成，问题就比较复杂。

如果组成结构的两个自由语素都是双音节语素或多音节语素，那么，它们必定是词组，不是合成词。例如，“模糊”是双音节自由语素，“逻辑”也是双音节自由语素，由它们构成的“模糊逻辑”必定是词组，不是合成词。

如果组成结构的两个自由语素，一个是双音节语素，一个是单音节语素，那么，就不容易判定这个结构是合成词还是词组。例如，“坦克”是双音节自由语素，“车”是单音节自由语素，由它们结合而成的“坦克车”，有人认为应该是合成词，因为它表示一个整体概念。但是，“开”是单音节自由语素，“坦克”是双音节自由语素，由它们构成的“开坦克”却很难认为是一个合成词，有许多人认为它是一个述宾结构的词组。

可见，当构成结构的两个自由语素中，有一个单音节语素，就可能使合成词和词组的界限变得模糊起来，难于判定。

如果构成结构的两个自由语素都是单音节语素，那么，合成词和词组的界限就更加模糊，更加难于判定。例如，当单音节自由语素“大”与另外的单音节自由语素“会，军，陆，脑，好，红”组成“大会，大军，大陆，大脑，大好，大红”时，有人会认为前后语素之间结合得很紧密，应该是

合成词。但是，当“大”与另外的单音节自由语素“鱼，河，船”组成“大鱼，大河，大船”时，可能就会有人觉得前后语素之间结合得不很紧密，它们不太像合成词，而似乎应该是词组了。

又如，表示陈述关系的结构“洗澡，鞠躬，游泳，理发”，看来似乎是合成词。可是，有时，其中的语素可以分离开来：

洗澡 -- 洗了一次澡
 鞠躬 -- 鞠了一个躬
 游泳 -- 游了一次泳
 理发 -- 理了一次发

这时，它们似乎又不像是合成词。究竟是合成词还是词组？难于判定。

我们可以把语素、词和词组的区别比较如下：

性质 结构 单元	是否有意义	是否为最小单位	能否独立运用	包含语素数	包含单词数
黏附语素	+	+	-	1	0
自由语素	+	+	+	1	1
单纯词	+	+	+	1	1
合成词	+	-	+	≥ 2	1
词组	+	-	+	≥ 2	≥ 2

图3 语素、词和词组的区别比较

从这个表中可以看出：

① 任何一个结构单元，可以根据“是否有意义”，“是否为最小单位”，“能否独立运用”，“包含语素数”，“包含单词数”等 5 个性质来鉴别。这 5 个性质之间的关系是逻辑上的合取关系(∧)，也就是说，每一个结构单元，要同时根据这 5 个性质来鉴别，如果仅仅根据其中的某一个性质或者某几个性质，是不可能鉴别清楚的。

② 自由语素与单纯词的性质完全一样，它们在实质上是一个东西。

③ 合成词与词组的前面 4 个性质都相同，只有最后一个性质（即“包含单词数”）不同，合成词只包含一个单词，而词组则包含两个或两个以上的单词；可是，由于自由语素同时又可以看成单纯词，因此，当合成词由两个自由语素组成时，也可以把它看成是由两个单纯词组成的，这样，合成词就变成词组了！

可见，在语言学的理论上，合成词与词组的分界问题并没有解决。这种理论上的缺陷，实际上也是一种理论上的硬伤，必然会在汉语书面文本自动切分的实践中，引起种种的矛盾和困难；在语言教学中，这种在逻辑上不完善的理论有可能扼杀学生的独立思考能力，影响学生按照正常的逻辑来思维，从而引起学生对于语言学理论的正确性、逻辑性和科学性的怀疑。这是我国现代汉语教学和研究应当特别注意的严重问题。特提出来，请方家指正。