

## 理论词和形式词

冯志伟

教育部语言文字应用研究所

40 多年前，在关于词儿连写的讨论中，有的学者曾经提出过形式词的概念，但是，没有引起学术界足够的重视，随着时间的流逝，人们对于形式词的概念已经淡忘了。现在，汉语拼音正词法已经公布多年，在我国的语言生活中发挥了巨大的作用，但是，还存在不少的问题。随着中文信息处理研究的发展，国家在 1992 年公布了《信息处理用现代汉语分词规范》，作为在现代汉语书面语自动切分的国家标准，在汉语书面文本的自动切分问题中，很多困难问题都与汉语拼音的正词法有密切的联系。我们发现，从形式词的角度来研究汉语拼音正词法的问题，有助于我们决定拼音词的分合，当然也就有助于汉语书面文本单词的分合，是很有好处的。今天，在纪念汉语拼音方案发布 45 周年的日子里，我们拟从汉语拼音正词法的角度，结合汉语书面语自动切分的实际，重新提出理论词和形式词的问题来进行讨论，希望能起到抛砖引玉的作用。

汉语拼音正词法的连写单位是词。但是，作为连写单位的词与语法上所说的词有联系也有区别。我们把语法上的词叫做理论词，把作为连写单位的词叫做形式词。周有光先生曾经提出，应该区分“理论词”和“连写词”，这是很有见地的。周先生所说的“连写词”，大体上就相当于我们这里的“形式词”。

汉语拼音正词法中所说的词当然是形式词，但是由于形式词与理论词有着密切的联系，我们有必要首先对传统汉语语言学中对于理论词的研究成果作一些总体性说明，以便我们进行比较。为了排版印刷的方便，我们所标的拼音一般都省去了声调，只有在涉及到声调的时候才标声调。

### 2.1 理论词

#### 2.1.1 语素

从语法理论上说，句子是由词和词组构成的，但是，词和词组并不是语言的最小单位，最小的语言单位是语素 (morpheme)。

语素是最小的语音语义结合体，是构词的材料。

汉语的语素从语音形式上看，可以分为单音节语素，双音节语素和多音节语素，其中，以单音节语素占优势。单音节语素是汉语语素的基本形式。

##### — 单音节语素

例如：天 [tian]，人 [ren]，造 [zao]，走 [zou]，鲜 [xian]，优 [you]，安 [an]，老 [lao]，子 [zi]，且 [qie]，者 [zhe]，然 [ran]。

这样的单音节语素组合起来非常方便，非常灵活，只要两个单音节语素在意义上能够搭配，组合方式符合汉语的习惯，就可以组成双音节的词。

例如，单音节语素“花” [hua] 可以同其它的单音节语素组合成许多双音节的词。

“花”在前，其它语素在后，可以组成“花瓣 [huanban]，花粉 [huafen]，花园 [huayuan]，花圈 [huaquan]，花蕊 [huarui]”等词；

“花”在后，其它语素在前，可以组成“鲜花 [xianhua]，雪花 [xuehua]，礼花 [lihua]，火花 [huohua]，荷花 [hehua]”等词。

##### — 双音节语素

汉语的双音节语素有两类：一类是从古代汉语遗留下来的“联绵词”，另一类是音译的外来语素。双音节语素中的每一个单独的音节不能表达独立的意思，只有两个音节结合起来才能作为一个最小的语音语义结合体。表达独立的意思。

所谓“联绵词”是指声母相同或者韵母相同的双音节语素。声母相同的双音节语素叫做“双声联绵字”，韵母相同的双音节语素叫做“叠韵联绵字”。

双声联绵字：

例如：蜘蛛 [zhizhu]，枇杷 [pipa]，琉璃 [liuli]，伶俐 [lingli]，仓促 [cangcu]，慷慨 [kangkai]，吩咐 [fenfu]，参差 [cenci]，拮据 [jieju]，坎坷 [kanke]，崎岖 [qiqu]，惆怅 [chouchang]。

叠韵联绵字：

例如，葫芦 [hulu]，蜻蜓 [qingting]，玫瑰 [meigui]，喇叭 ([laba]，霹雳 [pili]，苗条 [miaotiao]，逍遥 [xiaoyao]，汹涌 [xiongyong]，怱怱 [songyong]，伶俐 [lingding]，朦胧 [menglong]，荡漾 ([dangyang]。

在汉民族与其它民族的交往过程中，从其它民族的语言中吸收了一些词语来作汉语的语素，形成了音译外来语素。

双音节的音译外来语素的例子：

马达 ([mada]，motor)，逻辑 ([luoji]，Logic)，加仑 ([jialun]，gallon)，，摩登 ([modeng]，modern)，坦克 ([tanke]，tank)，幽默 ([youmo]，humor)，扑克 ([puke]，playing card)，瓦特 ([wate]，Watt)，欧姆 ([oumu]，Ohm)，柠檬 ([ningmeng]，lemon)。

—多音节语素

普通话中的多音节语素全部都是音译外来语素。

例如，白兰地 ([bailandi]，Brandy)，巧克力 ([qiaokeli]，Chocolate)，凡士林 ([fanshilin]，vaseline)，托拉斯 ([tuolasi]，Trust)，奥林匹克 ([aolinpike]，Olympic)，歇斯底里 ([xiesidili]，hysteria)，阿司匹林 ([asipilin]，aspirin)。

现代汉语的语素发展呈现出一种单音节化的趋势，来源于古汉语联绵字和来源于其它语言的音译外来语素日趋减少。这是现代汉语语素发展的特点。

汉语的语素从功能上看，又可以分为自由语素和黏附语素两类。

—自由语素：这一类语素活动能力很强，不仅可以与其它语素组合成词，而且，还可以单独成词使用，成为语言中最小的造句单位。

例如，“地 [di]，跑 [pao]，红 [hong] 等单音节语素都是自由语素。

这些自由语素可以单独使用。

例如：

天冷了，地冻了



他跑得快



脸上白一阵，红一阵



自由语素也可以同其它的语素组合成词。

例如，

地：“地”在前，其它语素在后。

地板 [diban]，地层 [diceng]，地带 [didai]，地基 [diji]，地震 [dizhen]。

“地”在后，其它语素在前。

高地 [gaodi]，陆地 [ludi]，山地 [shandi]，田地 [tiandi]，荒地 [huangdi]。

跑：“跑”在前，其它语素在后。

跑步 [paobu]，跑表 [paobiao]，跑道 [paodao]，跑电 [paodian]，跑鞋 [paoxie]。

“跑”在后，其它语素在前。

逃跑 [taopao]，赛跑 [saipao]，飞跑 [feipao]，奔跑 [benpao]，小跑 [xiaopao]。

红：“红”在前，其它语素在后。

红榜 [hongbang]，红茶 [hongcha]，红尘 [hongchen]，红人 [hongren]，红晕 [hongyun]。

“红”在后，其它语素在前。

鲜红 [xianhong]，粉红 [fenhong]，口红 [kouhong]，花红 [huahong]，橘红 [juhong]。

所有的双音节语素和多音节语素都可以单独成词，它们都是自由语素。

一黏附语素：这一类语素都是单音节语素，它们的活动能力不强，不能单独成词，它们要与其它的语素相组合而成词。如“机 [ji]，劳 [lao]，老 [lao]，小 [xiao]，子 [zi]，者 [zhe]，然 [ran]”。

有的黏附语素可以黏附在其它语素之前，也可以黏附在其它语素之后。如“机，劳”等。

机：“机”在前，其它语素在后。

机场 [jichang]，机车 [jiche]，机床 [jichuang]，机会 [jihui]，机密 [jimi]。

“机”在后，其它语素在前。

时机 [shiji]，飞机 [feiji]，动机 [dongji]，心机 [xinji]，电机 [dianji]。

劳：“劳”在前，其它语素在后。

劳动 [laodong]，劳力 [laoli]，劳神 [laoshen]，劳累 [laolei]，劳损 [laosun]。

“劳”在后，其它语素在前。

功劳 [gonglao]，疲劳 [pilao]，操劳 [caolao]，勤劳 [qinlao]，烦劳 [fanlao]。

有的黏附语素只能黏附在其它语素之前，叫做前缀。例如：

老：老板 [laoban]，老师 [laoshi]，老鼠 [laoshu]，老虎 [laohu]，老鹰 [laoying]。

小：小吃 [xiaochi]，小丑 [xiaochou]，小费 [xiaofei]，小工 [xiaogong]，小结 [xiaojie]。

有的黏附语素只能黏附在其它语素之后，叫做后缀。例如：

子：儿子 [erzi]，桌子 [zhuozi]，车子 [chezi]，棍子 [gunzi]，法子 [fazi]。

者：作者 [zuozhe]，读者 [duzhe]，强者 [qiangzhe]，前者 [qianzhe]，后者 [houzhe]。

然：忽然 [huran]，显然 [xianran]，欣然 [xinran]，悄然 [qiaoran]，岸然 [anran]。

汉语的语素从意义的虚实来看，可分为实语素和虚语素。此外还有少数语素是处于实语素与虚语素之间的，可以称为半实素。

#### 一 实语素

实语素包括名素、动素、形素等。

名素表示实体事物、数量、时间、方位等。例如：

事物名素：天 [tian]，地 [di]，石 [shi]，木 [mu]。

数量名素：一 [yi]，十 [shi]，尺 [chi]，斤 [jin]。

时间名素：年 [nian]，月 [yue]，日 [ri]，晨 [chen]。

方位名素：上 [shang]，下 [xia]，左 [zuo]，右 [you]。

动素表示动作行为、心理活动、发展变化等。例如，

行为动素：走 [zou]，飞 [fei]，见 [jian]，闻 [wen]。

心理活动：想 [xiang]，忘 [wang]，盼 [pan]，思 [si]，

变化动素：增 [zeng]，减 [jian]，扩 [kuo]，缩 [suo]。

形素表示事物的性质、状态等。例如，

性质形素：高 [gao]，大 [da]，美 [mei]，红 [hong]。

状态形素：谦 [qian]，恭 [gong]，愉 [yu]，速 [su]。

实素中的自由语素可以单独成词。例如，“天，地，年，月，上，下，左，右”等可以单独成为名词，“一，十”等可以单独成为数词，“尺，斤”等可以单独成为量词，“走，飞，见，想”等可以单独成为动词，“高，大，美，红”等可以单独成为形容词。但是，实素中的黏附语素要

与其它语素结合才能组成一个词。例如，名素“晨”要与语素“早”结合才能组成名词“早晨” [zaochen]，动素“思”要与语素“念”结合才能组成动词“思念” [sinian]，形素“谦”要与语素“虚”结合才能组成形容词“谦虚” [qianxu]。

— 虚语素：

虚语素不表示实在的意义，只是表示某种语法关系。例如，“和 [he]，的 [de]，着 [zhe]，了 [le]”。

— 半实素：

半实素有一定的意义，但不像实素那样实在。例如，“第 [di]，把 [ba]，员 [yuan]，被 [bei]”。

### 2.1.2 词

词是由语素构成的，可以单独使用。

词可以由一个语素构成，也可以由多个语素构成。根据词的构成情况，可以分为单纯词和合成词两类。

— 单纯词：由一个语素构成的词，叫做单纯词。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。

绝大部分单纯词是由单音节语素构成的。如 2.1.1 中所述的单音节自由语素“地，下，跑，红”等同时也是单纯词。

2.1.1 中所述的双音节自由语素“蜘蛛，枇杷，马达，逻辑”等以及多音节自由语素“白兰地，巧克力”等，也是单纯词，不过，这样的词为数不多。

— 合成词：

由两个或两个以上的语素构成的词，叫做合成词。这种词绝大部分是双音节的。例如，

人 + 民 → 人民 [renmin]  
学 + 生 → 学生 [xuesheng]  
劳 + 动 → 劳动 [laodong]  
热 + 爱 → 热爱 [re'ai]  
美 + 丽 → 美丽 [meili]

此外，还有一部分由两个以上的语素构成的多音节合成词。例如，

运 + 动 + 员 → 运动员 [yundongyuan]  
录 + 音 + 机 → 录音机 [luyinji]  
火 + 车 + 站 → 火车站 [huochezhan]  
现 + 代 + 化 → 现代化 [xiandaihua]  
非 + 金 + 属 → 非金属 [feijinshu]

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。

有时，两个并列的语素的意义是相同的或相近的。例如，

朋友 [pengyou]，泥土 [nitu]，帮助 [bangzhu]，松散 [songsa]，舒畅 [shuchang]。

有时，并列的两个语素的意义是相对的或相反的。例如，

东西 [dongxi]，开关 [kaiguan]，矛盾 [maodun]，买卖 [maimai]，来往 [laiwang]。

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 [huoche]，工业 [gongye]，微笑 [weixiao]，雪白 [xuebai]，飞快 [feikuai]。

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 [lingdui]，司机 [siji]，主席 [zhuxi]，签名 [qianming]，悦耳 [yue'er]。

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 [tigao]，说明 [shuoming]，扩大 [kuoda]，缩小 [suoxiao]，改善 [gaishan]。

⑤ 陈述式：合成词中的两个词，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 [dizhen]，日食 [rishi]，祖传 [zuchuan]，法定 [fading]，雪崩 [xuebeng]。

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。可参看 2.1.1 中说明“老，小，子，者，然”等黏附语素的例子。

⑦ 重叠式：合成词是由单音节语素重叠而构成的。例如，

星星 [xingxing]，茫茫 [mangmang]，纷纷 [fenfen]，巍巍 [weiwei]，翩翩 [pianpian]。

2.1.3 词组：词组是由词和词组合而成的。

汉语词组的构成方式主要有以下 6 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 [fengsu xiguan]，调查研究 [tiaocha yanjiu]，勤劳勇敢 [qinlao yonggan]，植树造林 [zhishu zaolin]，轻松愉快 [qingsong yukuai]。

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 [mangmang dahai]，劳动热情 [laodong reqing]，衷心祝贺 [zhongxin zhuhe]，十分灵敏 [shifen lingmin]，热烈欢迎 [reliehuanying]。

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 [re'ai zuguo]，欣赏音乐 [xinshang yinyue]，发射导弹 [fashe daodan]，供养父母 [gongyang fumu]，行使职权 [xingshi zhiquan]。

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 [jieshi qingchu]，举起来 [ju qilai]，洗干净 [xi ganjing]，讲明白 [jiang mingbai]，扔出去 [reng chuqu]。

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 [xiaohai kesou]，姑娘唱歌 [guniang changge]，天气热 [tianqi re]，月亮圆 [yueliang yuan]，今天星期日 [jintian xingqiri]。

⑥ 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 [yanjiu yanjiu]，练习练习 [lianxi lianxi]，讨论讨论 [taolun taolun]，总结总结 [zongjie zongjie]，复习复习 [fuxi fuxi]。

2.1.4 语素、词和词组的界线

黏附语素不能独立使用，不是词，自由语素可以独立使用，所以，自由语素同时又是单纯词。从语素的角度看，它是自由语素，从词的角度看，它是单纯词。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。

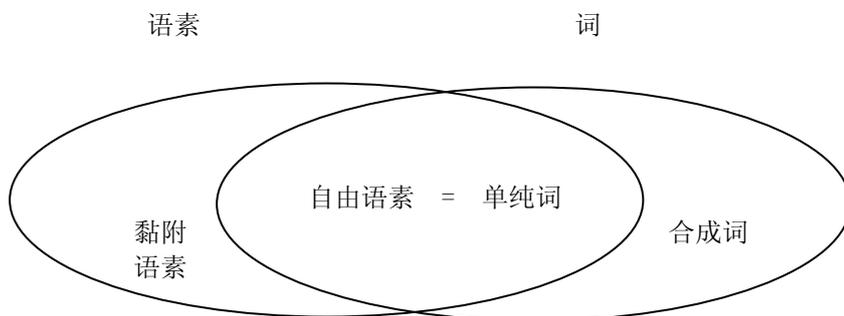


图 2.1 语素和词之间的交集

由此可见，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，合成词也不可能是单个的语素，语素和词之间的交集，从语素的角度看是自由语素，从词的角度看是单纯词。

然而，合成词和词组之间的界限就不是那么容易区分清楚了。

从前面所述可以看出，合成词的构成方式与词组的构成方式有许多一致的地方。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
重叠式	重叠结构	重叠关系

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” [lao] 是黏附语素，“损” [sun] 也是黏附语素，它们结合而成的“劳损” [laosun] 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” [lao] 是黏附语素，“动” [dong] 是自由语素，他们结合而成的“劳动” [laodong] 必定是合成词，不可能是词组。

含有前缀的“老师” [laoshi]，“老虎” [laohu] 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” [zhuozi]，“作者” [zuozhe]，“忽然” [huran] 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” ([mohu], unklar) 是双音节自由语素，“逻辑” ([luoji], Logik) 也是双音节自由语素，由它们构成的“模糊逻辑” ([mohu luoji], Fuzzy Logik) 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” [tanke] 是双音节自由语素，“车” [che] 是单音节自由语素，由它们结合而成的“坦克车” [tankeche]，有人认为应该是合成词，因为它表示一个整体概念。但是，“开” [kai] 是单音节自由语素，“坦克” [tanke] 是双音节自由语素，由它们构成的“开坦克” [kai tanke]，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” [da] 与另外的单音节自由语素“会 [hui]，军 [jun]，陆 [lu]，脑 [nao]，好 [hao]，红 [hong]”组成“大会 [dahui]，大军 [dajun]，大陆 [dalu]，大脑 [danao]，大好 [dahao]，大红 [dahong]”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 [yu]，河 [he]，船 [chuan]”组成“大鱼 [da yu]，大河 [da he]，大船 [da chuan]”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 [xizao]，鞠躬 [jugong]，游泳 [youyong]，理发 [lifa]，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

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，从理论上说，合成词与词组的分界问题并没有解决。这种理论上的缺陷，必然会在汉语拼音正词法的分词连写实践中，引起种种的矛盾和困难。

## 2.2 形式词

### 2.2.1 形式词的定义

汉语拼音正词法要求“分词连写”，所谓“分词连写”，就是在拼音连写之前，要考虑如何分词，把词分出来之后，一个词中的各个字母要连写。

从分词连写的角度，我们可以把词定义为“汉语拼音中经常连写在一起的一组字母”。这样定义的词，叫做形式词。

其实，即使在用汉字表示的词中，人们也常常给词赋予某种形式，使之更加鲜明醒目。

例如，《说文解字》中说，“蝼，蝼蛄也”。蝼蛄 [lougu] 中的“蝼”字和“蛄”字都是“虫”字旁，以表示它们是一个词。

又如，“伙伴” [huoban] 原来写为“火伴”，后来在“火”字上仿照“伴”字加了“亻”字旁，以表示它们是一个词。

“婚姻” [hunyin] 原来写为“昏姻”，后来在“昏”字上仿照“姻”字加了“女”字旁，以表示它们是一个词。

“凤凰” [fenghuang] 原来写为“凤皇”，后来在“皇”字上仿照“凤”字加了个帽子，写成“凰”字，以表示它们是一个词。

在汉语拼音正词法中，我们给词赋以特定形式的方法，就是分词连写，也就是把形式词与其前后的其它形式词分开，而在形式词内部的各个字母之间或各个音节之间连写。“分词连写”是形式词的形式标准，是把形式词从形式上表现出来的形式手段。

### 2.2.2 分词连写的学术依据

由于在理论上合成词与词组的界限问题没有彻底解决，我们在讨论形式词的分词连写问题时，只有从实践中逐步摸索它们的学术依据。

学者们曾经尝试从三个方面来寻找分词连写的学术依据，提出了各种不同的测定合成词与词组的界限的方法：第一是语法依据，第二是语义依据，第三是语音依据。

在语法方面，提出了如下的测定方法：

#### ① 替代测定法

用性质相近的别的自由语素来替代待测结构中的自由语素，如果能够替代，就可以判定为词组，而不是合成词。

例如，要测定“吃饭” [chi fan] 是合成词还是词组，先用与“吃” [chi] 性质相近的自由语素“盛” [cheng]，煮 [zhu]（它们都表示动作）替代“吃”，说成“盛饭” [cheng fan]，“煮饭” [zhu fan]；再用与“饭” [fan] 性质相近的自由语素“面” [mian]，“粥” [zhou]（它们都是食品）说成“吃面” [chi mian]，“吃粥” [chi zhou]，由于前后两个自由语素都能被替代，就可以判定“吃饭”是词组，而不是合成词。

替代测定法是不可靠的，这种方法容易引出不合常识的错误结论。

例如，“驼绒” [tuorong]，可以被替代之后成为“驼毛” [tuomao]，“驼肉” [tuorou]，“鸭绒” [yarong]，“鹅绒” [erong]，但是，从语感上显然“驼绒”不可能是一个词组，而是一个合成词。替代测定法得出的结论，与人们的语感差别太大。所以，替代测定法只能作为分词连写的一种参考，不能作为可靠的依据。

#### ② 插入测定法

用特定的自由语素（如“的”）插入待测定的结构中，如果能插入而不改变该结构的意义，就判定为词组，而不是合成词。

例如，“新鞋” [xin xie] 中插入特定的自由语素“的” [de]，形成“新的鞋” [xinde xie]，意义没改变，可判定“新鞋”为词组，不是合成词。

“白菜” [baicai] 中插入特定的自由语素“的”，形成的“白的菜” [baide cai]，其意义与“白菜”不同，可判定“白菜”不是词组，而是合成词。

插入测定法比较客观，适用范围比较广，但是，有时也会得出一些不合常识的结论。

例如，北京话中可以说“鸡” [ji]，不可说“鸭” [ya]，而要说成“鸭子” [yazi]。如果我们用插入自由语素“的”的方法来测定“鸡蛋” [jidan] 和“鸭蛋” [yadan]，“鸡蛋”可以改说成“鸡的蛋”，“鸭蛋”不可以改说成“鸭的蛋”，于是得出结论：“鸡蛋”是词组，“鸭蛋”是合成词，这种结论与人们的语感相差太大。事实上，人们普遍认为“鸡蛋”和“鸭蛋”都不是词组，而是合成词。

可见，插入测定法并不是万能的，使用时要考虑到各种复杂情况。

在语义方面，提出了如下的方法：

### ① 意义单纯性判定法

根据待测结构中两个语素意义结合而成的总体意义的单纯性来判定。总体意义单纯的判定为合成词，总体意义不单纯的判定为词组。

例如，“城市” [chengshi] 的总体意义单纯，是合成词；“夫妻” [fu-qu] 的总体意义不单纯，它的意义等于“夫” [fu] 与“妻” [qi] 的意义的总和，是词组。

“东西”这个结构有歧义。当它的意义表示事物时，意义单纯，是合成词，当它的意义表示“东边和西边”时，这个意义等于“东” [ost] 和“西” [xi] 的总和，意义不单纯，是词组。

### ② 意义紧密性判定法

根据待测结构中两个语素意义结合的紧密性来判定，意义紧密的判定为合成词，意义松懈的判定为词组。

例如，“爱国” [aiguo] 中的两个自由语素“爱”与“国”中间不能插入别的成分，意义结合得很紧密，判定为合成词。“读书” [du shu] 中的两个自由语素“读”和“书”之间可以插入别的成分：“读了一本书”，意义联系松懈，判定为词组。

### ③ 引申意义判定法

根据待测结构的意义是否为引申意义来判定，是引申意义的判定为合成词，而保持本义的就可判定为词组。

例如，“吃饭”的本意是进餐，判定为词组，但是，在句子“靠自己的劳动吃饭”中，“吃饭”的意义引申为“生存”，就判定为合成词。

### ④ 常用性判定法

根据待测结构的常用性来判定，常用的判定为合成词，不常用的判定为词组。

例如，“国外” [guowai] 常用，判定为合成词，“省外” [sheng wai] 不常用，判定为词组。

在语音方面，提出了如下的方法：

#### ① 停顿判定法

构成词组的自由语素之间常有停顿，可以作为分词连写的参考。

例如，“中华人民共和国”这个结构中的停顿情况是：“中华 · 人民 · 共和国”，语素之间有停顿，判定为词组，分写为 [Zhonghua Renmin Gongheguo]。

#### ② 双音节化判定法

汉语的单词有双音节化 (disyllabism) 的倾向。双音节化导致音节之间出现两种相反的现象：一种是相吸，另一种是相拒，汉语双音节化的基本规律是：“单单相吸”，“双双相拒”，“吸单拒双”。

所谓“单单相吸”，是指两个单音节的语素相吸而连结成一个合成词。例如，“人”和“民”相吸而连结成合成词“人民” [renmin]，“香”和“烟”相吸而连结成合成词“香烟” [xiangyan]。

所谓“双双相拒”，是指两组双音节结构往往有相拒的倾向而分写为词组。

例如，“讨论” [taolun] 是一个双音节结构的合成词，它的重叠形式是“讨论讨论”由两个双音节结构组成，这两个双音节结构彼此相拒，应分写为 [taolun taolun]。

所谓“吸单拒双”，是指当双音节结构与单音节结构相遇时，这个双音节结构能够把单音节结构吸引过来而形成合成词，而当双音节结构与另一个双音节结构相遇时，这个双音节结构往往会排斥另一个双音节结构而新形成词组。例如，“图书” [tushu] 是个双音节结构的合成词，当它与单音节语素“馆” [guan] 相遇时，能够把这个单音节语素“馆”吸引过来，形成“图书馆”这个合成词，应连写为 [tushuguan]；但是，当它与双音节结构“目录” [mulu] 相遇时，却排斥这个双音节结构，而形成一个词组“图书目录”，应分写为 [tushu mulu]。有时，三音节结构也会把它后面的单音节语素吸引过来而形成合成词，也具有“吸单拒双”的规律。例如，“天文学” [tianwenxue] 这个三音节结构，与单音节语素“书” [shu] 相遇时，会把这个单音节语素吸引过来而形成合成词“天文学书”，应连写为 [tianwenxueshu]。

在汉语拼音正词法的研究中，我们不能只采用一种方法来解决分词连写问题。比较切合实际的办法是综合运用上述各种方法来进行判断，各种方法之间应该相互补充，相互校正。

### 2.2.3 形式词的其它形式因素

形式词是理论词的汉语拼音书写形式，除了 2.2.2 中所述的学术依据之外，还应该考虑以下的形式因素。

#### ① 视读原则

汉语拼音是一种视读实体，它必须满足视觉形象方面的要求。

但是，词的视觉形象和听觉形象往往不一致。“鹅”（[e]，Gans）和“床”（[chuang]，Bett）的听觉形象都是一个音节，但在拼音词的视觉形象上差别很大，“鹅”只含有一个字母 [e]，而“床”却含有六个字母 [chuang]。“庄重”（[zhuangzhong]，ernst）和“积极”（[jiji]，aktiv, positive）的听觉形象都是两个音节，但拼音词的视觉形象差别却很大，“庄重”含有 11 个字母 [zhuangzhong]，而“积极”只含有 4 个字母（[jiji]）。视觉会本能地接受拼音形式较短的词，而觉得拼音形式太长的就不像词。

根据认知心理学的研究，人对信息的感知广度以 7 左右为限。我们数苹果，五个五个地数比较容易，十个十个地数就很难。据说象棋大师对于不成布局的、阵势较乱的棋盘，粗看一下之后，至多也只能记住 7 个棋子的位置。根据这样的原理，汉语拼音词的字母数以不多于 7 个为佳，在汉语拼音正词法中，要尽量使字母数目超过 7 个的词不要太多。例如，单音节动词重叠式连写，双音节动词重叠式分写，就是因为双音节动词重叠之后，其字母数目往往多于 7 个，超出了人们的感知广度。又如，“同步回旋加速器”如果连写为一串长龙 [tongbuhuixuanjiasuqi] 不便阅读，根据视读原则，可分写为“同步”、“回旋”、“加速器”3 个部分，拼写为 [tongbu huixuan jiansuqi]，这 3 个部分，称之为语节（segment）。

认知心理学还证明了，字母序列中首尾两头的字母比较容易辨认。个别的一些长词，如果我们看一看它们的两头，再加上前后文的提示，则中间的字母不必细看也可以辨别出这个词来。根据这样的原理，汉语拼音正词法中，可以把多音节后缀“-主义”、“-主义者”同前面的词根连写，反而比分写容易辨认。当然，这样的长词不宜过多，长词的数目要加以严格的控制。如果长词数目太多，其可辨识性就会随长词数目的增加而降低。

在确定形式词的时候，我们应该考虑到这些视读方面的原则。

#### ② 分写、连写和半连写

在汉语拼音正词法中，词与词之间的间隔用空格表示，这是“分写”，词内各字母之间不留间隔，这是“连写”，除了“分写”和“连写”两种手段之外，还有第三种手段——“半连写”，也就是用短横“-”相连接。因为语词的结构是复杂的，在许多场合，“分写”不妥当，“连写”也不妥当，就需要采取中性的妥协手段来处理，使用短横“-”来半连写，以便阅读和理解。

有些新产生的缩写词，一时不易理解，可用半连写。例如，“环境保护” [huanjing baohu] 的缩写词“环保”，可写为 [huan-bao]，“公共关系” [gonggong guanxi] 的缩写词“公关”，可写为 [gong-guan]。

用两个连续数字相连表示约数的数词，可用半连写。例如，“八九天”写为 [ba-jiu tian]，“十七八岁”写为 [shiqi-ba sui]。

有些并列结构的词，并列部分可用半连写。例如，“夫妻”可写为 [fu-qi]，“人机对话”可写为 [ren-ji duihua]，“中小学”可写为 [zhong-xiao xue]，“陆海空军”可写为 [lu-hai-kong jun]，“辩证唯物主义”可写为 [bianzheng-weiwu zhuyi]。

#### ③ 只标原调，不标变调

汉语的声调有原调和变调之分。例如，“一”这个语素单用或处于一词一句末尾时念阴平，如“十一”念 [shiyī]，“二十一”念 [ershīyī]，在去声字前念阳平，如“一半”念 [yíban]，“一共”念 [yígong]，在阴平、阳平、上声字前念去声，如“一天”，“一年”，“一点儿”中的“一”念为 [yì]，如果按照变调拼写，“一”会有多种拼写形式，不便阅读。汉语拼音正词法规定，“一”一律标阴平调，凡有变调的字，一律只标原调，不标变调。这样，上述词应该拼写为：

十一	[shiyī]	二十一	[ershīyī]
一半	[yíban]	一共	[yígong]
一天	[yìtian]	一年	[yìnian]
一点儿	[yìdiǎnr]		

根据同一原理，“七、八”也有变调，但都标为阴平。例如，

“七万”标为 [qī wan]，“七本书”标为 [qī ben shu]，“八个学生”标为 [bā ge xuesheng]，“七上八下”标为 [qī shang bā xia]。

“不”也有变调，但都标为去声。例如，

“不去”标为 [bù qu]，“不对”标为 [bù duì]，“不至于”标为 [bù zhīyú]。

从理论的角度说，词当然是有变调的，但从形式词的角度说，为了便于阅读和理解，只标原调，不标变调，更有好处。只是在语音教学时，可以根据需要按变调拼写。

#### ④ 音译转写，名从主人

非汉语的人名、地名译为汉语时，如用汉字注音，由于语音体系的不同，与原来的读音会有许多差异。例如，著名物理学家 Einstein 用汉字译为爱因斯坦，按汉字转写成拼音应该为 [Aiyinsitan] 用这样的汉字注音方法拼写成的外国人名，外国人看了不知道是谁，中国人看了也不知道真姓名是什么。汉语拼音正词法规定，非汉语人名、地名转写为拼音时，应本着“名从主人”的原则，按照拉丁字母原文书写，非拉丁字母文字的人名、地名，按照该文字的拉丁字母转写法拼写。

例如，非汉语人名：乌兰夫 [Ulanhu]，赛福鼎 [Seypidin]，阿沛·阿旺晋美 [Ngapoi Ngawang Jigme]，马克思 [Marx]，达尔文 [Darwin]，牛顿 [Newton]，爱因斯坦 [Einstein]，芥川龙之介 [Akutagawa Ryunosuke]。

非汉语地名：乌鲁木齐 [Urumqi]，呼和浩特 [Hohhot]，拉萨 [Lhasa]，伦敦 [London]，巴黎 [Paris]，华盛顿 [Washington]，东京 [Tokyo]。

对于不懂外文的中国读者，这样的拼写形式念起来可能不方便。因此，汉语拼音正词法还作了变通的规定：为了便于阅读，可以在原文后注上汉字或汉字的拼音，在一定的场合，也可以先用或仅用汉字的拼音。

但是，对于已经汉语化的音译人名、地名，可按汉字译音拼写。例如，

莎翁 (Shakespeare) 可拼写为 [Shaweng]，南美 (South America) 可拼写为 [Nanmei]。

这些从形式词的角度做出的规定，可使汉语拼音正词法更好地满足阅读的需要。

拼音的习惯是要在许多年的时间内才可以养成的，而这些习惯，只有到了印刷事业发达之后才有可能稳定下来，正词法的形成是一个很长的历史过程。国家教育委员会和国家语言文字工作委员会于 1988 年 7 月公布了《汉语拼音正词法基本规则》，这个正词法规则不可能是一劳永逸的。这样的正词法规则只有示范作用，没有强制作用。正词法必须在群众的拼写实践中不断地加以修正，才有可能逐渐完善起来，最后达到相对的稳定状态。在拼音正词法不断完善的过程中，如果我们注意理论词和形式词的区别，充分考虑确定切词单位的各种语法因素和非语法因素，将会对我们的研究工作有所裨益。

### 参考文献

1. 冯志伟，确定切词单位的某些非语法因素，《中文信息学报》，2000 年
2. Feng Zhiwei, Some grammatical factors to determine the segmentation element 《Journal of Chinese Language and Computing》, An International Journal of The Chinese & Oriental Language Information Processing Society, Dec 2001, Volume 11, Number 2, Singapore.
3. 周有光语文论集，第二卷，上海文化出版社，2001 年。
4. 国家标准 GB13715, 信息处理用现代汉语分词规范，北京，中国标准出版社，1992 年。