约定用法和“词”的定义

陈嘉映

Word and its Counterparts in Chinese

关键词 字　词　短语　　自由词组　约定　类推

提要：‘词’的定义一直是个难题，在汉语中尤甚。本文分析了‘最小意义单位’、‘基本造句单位’等传统定义并指出这些定义的不足。进而讨论了约定用法和自由表达这组基本概念，指出‘约定’这一概念是探讨‘词’的定义的主要线索，并讨论了基础约定和次级约定。检验是不是约定的形式标志是类推，不能充分类推的是约定用法，能够充分类推的是自由词组或自由表达。本文最后指出，日常汉语里所说的“词”不仅要从约定来判断，而且和音节数有关。在讨论的展开过程中，本文也考察了‘定义’、‘概念’、‘说话人与听话人’等概念。

作者 陈嘉映，1964—1968 年上中学，1968—1976 年插队，1978—1981 年在北京大学西方语言文学系和北京大学外国哲学研究所读书，毕业后留校，1984—1991 年在美国宾夕法尼亚州立大学哲学系读博士学位，1993 年回国，现在是北京大学哲学系的副教授。主要译著是海德格尔的《存在与时间》，主要著作是《海德格尔哲学概论》。

> 词是‘语言中的最小意义单位’。（王力，一，16 页）
> “一个词，就是一个简单的意义单位。”（王力，二，171 页）
> 词是‘能够单独成句的单位’。（Bloomfield，178 页）
> “词是概念的个别形象……”（洪堡特，116 页）

“词”的定义一直是个难题。“我们怎样知道哪些是词？语言学家进行了长期探索，试图设计出可行的标准，但还没有一个完全成功的”（克里斯特尔，145 页）。

但这个难题不是专属语言学的，它也一直为哲学所关心。“词”的定义涉及“意

需要全本请访问 www.ertongbook.com

义”问题，而意义理论通常认作是语言哲学的核心，仅此一点就可说明为什么“词”的定义问题是语言学和哲学的交汇点。

最小的意义单位

“词”的最常见的定义之一是“最小意义单位”。有人反对用意义、观念来定义“词”，因为这些似乎是内在的东西，没有提供形式特征，不宜用来下定义①。一个明显的问题是，意义怎样分解为单位呢？固然，意义是可以分析的，例如我们常说“分析一篇文章的意义”，然而，分析不见得是把一样东西分解成较小的部分、分解为要素(永井成男,序论,一;“什么是分析”),例如把‘三角形’分析为‘由三条线段组成的封闭图形”,其中‘三’‘封闭’‘图形’等显然不是比‘三角形’更小的‘意义单位”。奥斯丁曾指出“意义的一部分”是个没着没落的用语（Austin,31 页),那么自然也就谈不上“最小的意义单位”了。

简单的意义单位

“最小的意义单位”这个表达式含义不明，于是有人把它进一步理解为“简单的意义单位”(王力,二,171 页)。刘叔新把词定义为‘意义单纯的语言单位”,并举例说‘知识’表示一个单纯的概念”,与此相对;“知识分子’是个词组,其意义是‘复杂’的(邢公畹,118 页)。这是个更糟的定义;“知识”“法”“衍射”,这些都是单词,但何尝意义简单?在我看;“知识’是个很复杂的概念,至少和‘知识分子’同样复杂。况且，决定一个概念是否复杂，不是件容易的事，并不比确定什么是意义的单位容易。

最小音义结合体：语素

其实，采用这一定义的论者多半指的是“语言中独立地具有意义的最小声音单位”、“最小的音义结合体”或诸如此类。这个定义不问词义是否还能分解为更小的单位，问的是一个语音是否还能分解为具有含义的语音单位。Beacon 固然可以分解为 bea 和 con 两个音节，但这两个音节各自或者没有意义，或者其意义不能说明 beacon 的意义。“醉’是‘承载意义的最小单位”,一方面,若把它分解为更小的单

① 王力即采用这一定义（王力，一，16 页），但他自己也承认“这不是完善的定义……怎么才算‘最小’什么是‘意义单位’?”(王力,三,238～239 页）

位，z 和 ui那是些没有意义的声音① 另一方面，“他喝醉了”固然承载意义，但它不是承载意义的最小单位，因为它是由“他”、“醉”等更小的音义结合体组成的。用同样的方法来分析“囫囵”，可知“囫”、“囵”都不是词，“囫囵”才是单词。

“最小音义结合体”是一条值得追循的线索，但拿它来作“词”的定义则有疑问。次要的一点是它仍然和意义相连，缺少外在判别式，例如作为词缀的“儿”和“子”，怎样判断它们独立具有意义，抑或只是附加的符号？② 但这个定义的主要困难在于：语言学里普遍把承载意义的最小单位称为“语素”（morpheme）。语素是比单词低一级的音义结合体，固然，多数语素本身就是单词，如 he，top，但也有些语素不成其为单词，如 co-，pro-，trans-、-ness、-ism，它们符合“音义结合体”这个定义，但不是词。语素概念也早被引入来分析现代汉语，按照这种分析，虽然汉语里很多语素同时就是词，但也有大量的字，如“英”、“始”、“鞠”，有音有义，是语素，但不是词。

基本造句单位：独立成句

用“最小意义单位”来定义词看来行不通。另一种通行的定义是“最小的或基本的造句单位”（高名凯、石安石，105 页；张志公，126 页）。“基本造句单位”又有两种解释，一种是理解为“能够单独成句的单位”（Bloomfield，178 页），一种是“自由形式”。

赵元任基本上同意“单独成句”这个定义，但他提出了不少附加标准（赵元任，二，79 页、86 页、101 页）。王力则反对这个定义，因为这样一来至少得把“吗”“之”这些单词都排除在单词之外（王力，一，17 页）。对此，吕叔湘提出一个补充方案：把一个句子里能单独成句的成分拣走，剩下的成分，若不是一个词的一部分，就是词，例如“我下午再来”，把“我”、“下午”、“来”拣走后，剩下的“再”字也是词。但“比赛现在开始”里的“始”字不属此列，因为“始”是“开始”这个词的一部分（吕叔湘，17 页）。

这个补充方案虽解决了部分困难，却不能解决根本问题，那就是很多词不能单独成句，不仅很多虚词不能单独成句，多数副词都成问题，其它的词类中也有不少例子：“都”、“已经”、“满地”、“得到”、“度数”，按照我们的实感应是单词，却从不独立成句。吕叔湘自己还列举了一些更复杂的情况：1. 有时不能单独说，有时却能

① 这里不考虑极端的声训家的主张。的确，有些声母和韵母似乎本身就有特定的含义（参阅齐冲天，207～208 页，227～228 页），或更确切说，声母和韵尾发音部位相同而意义又相近。赵元任把这些称作“准语素”（赵元任，二，95～96 页）。但这不是普遍的情况，通常研究现代汉语的都不考虑这一层音义结合体。【除了极少数的例外如感叹词，爸爸妈妈这些词以外，他们的主张没什么根据。】

② 此外还有“专名问题”：“关羽”说不上有什么意义，但它显然是个词。不过专名问题是另一层次上的问题。

单独说，例如三号楼、院领导；2. 科技文献中可以单独说，例如氧、叶；3. 熟语里可单独说，例如前怕狼后怕虎、你一言我一语；4. 口语中不单独说，文章中却可以单独用，例如云、时（吕叔湘，18 页）。英语里也有 the 之类的难题。不仅有这些实例为证，而且，从道理上说，也看不出为什么把“是一个词”和“能够单独成句”联在一块儿。从字面上说，“造句”和“单独成句”显然是两码事，让学生用“满地”造句，不是让他把“满地”当作一个句子。单词独立成句本身是个值得讨论的现象，但它与词的定义没有直接关系。

基本造句单位：自由形式

另一种理解不把“基本造句单位”看作“能单独成句”，而是理解为“自由成分”或“自由形式”（free form）。自由形式是不粘附于其它语言成分的单位，它不必和某个其它成分连用而能和别的单位自由组合。“我”“是”“看见”可以出现在无数种前后文里，在合乎句法的条件下可以和任何语词搭配，是典型的自由形式。相反，“英”这个字只出现在“英雄”“英俊”“群英会”等特定的搭配之中，所以它不是自由形式。至于“龌龊”，“龌”只带着“龊”，“龊”永远跟着“龌”，两个字分开来都没什么含义，自然更不是自由形式了。另有一些字，如“取”和“笑”，在“取笑”这话里粘连在一起，不能自由移动，在“取订单”“笑他弟弟”这些话里则是自由形式。

“自由形式”和“单独成句”是对“基本造句单位”的不同理解。“吗”符合“自由形式”这个定义，但不符合“独立成句”这个定义。然而有论者不加辨别，已经把“基本造句单位”明确界定为“在语句中能独立活动的基本语法单位”，接下来立刻又把这话解释成“能够单独成句”（陈光磊，9 页），或干脆把两者混在同一个定义里，说词是“能够单独说的、自由运用的造句单位”（李荣，14 页）。我认为，“最小造句单位”只能由“自由形式”来定义，而不能由“独立成句”来定义。

“自由形式”和“最小音义结合体”显然是不同的定义，“英”是音义结合体，但“英”在现代汉语里却不是自由形式。然而有些论者同时采用这两种定义：“词是由句子中分出来的最小意义单位，这就是说，词是句子所组成的各个可以独立运用的部分”（王力，三，238 页）。这种“兼收并蓄”不能解决而最多是掩盖了我们的困难。我们希望找到把不同线索结合起来的结构，而不是在两条线索之间摇摆的折衷主义。

“自由形式”这个定义有自己的困难。驼和鸭都不是自由形式，羊和鸡都是，推论下来，似乎“驼毛”和“鸭蛋”都是单词，而“羊毛”和“鸡蛋”都是自由词组（吕叔湘，20～21 页），这未免有点奇怪。不过我们先放过这层困难，来看一看怎样确定自由形式。

扩展法

比起从意义方面来定义的“音义结合体”，“自由形式”这个定义看来有一个优点，那就是可以由外在的标准来判别。最通用的判别方法是‘扩展法’或‘插分法’：我们想知道 hit the man 是一个单词还是三个单词，我们可以试着在 hit 和 the 之间以及 the 和 man 之间插入另一些语言成分，看看成不成话，例如把它扩展成为 hit and killed the young man，现在我们看到这仍然成话，于是就确定 hit the man 是三个单音节的单词而不是一个三音节的单词。相反，在 youngest 中间，在 young 和 -est 之间，无法插入任何别的成分，由此可知 youngest 是个单独的词。同理，some thing 是词组而 something 是单词，因为我们说 some good thing，但只能说 something good。由此也可知，虽然 un-、-est、-ness 等等可以加在很多词上产生新词，但它们本身只是词缀，不是单词。

很多研究汉语的学者也采纳扩展法来区别单词和词组（王力，一，46 页；高名凯、石安石，106 页），有些在运用时有所变通（参考 陈光磊，13～14 页）。根据扩展法，可以判定‘老婆’是单词而‘老人’是词组，在‘老人’之间可以插入别的字，扩展为‘又老又丑的人’，但‘老婆’中间就插不进任何别的字了；‘马车’可以扩展成‘马拉的车’，所以是词组，不是单词；‘说话’和‘取笑’都是动宾结构，但一个是词组，一个是单词；‘说话’可以扩展为‘说大话’‘说废话’；‘取笑’中间则不能再插入任何语词（王力，一，46 页）。

我们也可以用扩展法来检验包括‘英’字的所有语词，如‘英雄’‘英俊’‘群英会’等等，结果发现“英”与另一个语素之间都插不进别的成分，由此可知在现代汉语里“英”是语素但不是单词。

扩展法的缺陷

扩展法的一个优点，是它能提供外部标准。但这个优点其实十分有限，我们说允许在几个语言成分之间插入其它成分，当然是说插入了其它成分而意思没有大变。我们凭什么说‘老婆’不能扩展成‘老年的婆婆’？因为‘老婆’和‘老年的婆婆’意思不一样。然而，把‘马车’换成‘马拉的车’，意思就当真没变吗？一辆车现在由骡子拉着，我们仍然叫它“马车”，用马来拉一辆小汽车，这辆小汽车并不因此叫作“马车”。这个反诘虽然有点刁钻，但的确需要回答。总的说来，‘意思不变’‘意思一样’是极难澄清的概念。但无论意思变了还是没变，总要对意思有个了解，而这种了解恰恰又回到了“内在”方面去了。

而且，我们根据什么规则允许把‘马车’换成‘马拉的车’，而不是仅仅换成‘马的车’？‘马车’可以扩展为‘马拉的车’，那么‘火车’‘卧车’‘机车’‘大车’‘铁路’呢？我们能否说火车是火力推动的车，铁路是铁轨铺成的路？如果提供不出一

需要全本请访问 www.ertongbook.com

定的规则,说不定所有的复音词都能扩展,即如‘取笑’,亦可扩展为‘取其短处加以嘲笑’。

而且，在西语分析中采用扩展法就会碰到不少模棱两可的案例，应用于汉语分析，疑难的例子更多。王力承认有时会碰上分不清的例子，如“干儿子”、“帝国主义”等(王力,一,14 页)。进一步的验试发现疑难实例远比“有时”更经常。“物理”(物之理)、“宇宙”(上下四方为宇,古往今来为宙)、“人心”、“六边形”、“法律学”,算不算合乎扩展法?打倒’等动补式组合通通可以插入‘得’字‘不’字,构成了另一大类难题。再如‘鞠躬’,人人都会认之为单词,却可以扩展为‘鞠了个躬’。另一方面,‘含沙射影’、‘天将降大任于斯人也’等成语、熟语在句子里总是连在一起用的,不能在中间插入别的成分，但因此可以说它们是单词吗？

最后,采用扩展法得出的结果常和我们对语词的实感相去甚远,例如‘英~烈’被认作词素①,‘稳~定’则本身是词,‘英烈’和‘稳定’的构词就实感而言有多大差别?再则,拿汉语里的‘英’、‘扩’、‘始’来和英语里的 trans- un- -ness 一类比照，不同之处大于相同之处。1. 英语里的这些亚单词语素,主要是前缀、后缀,汉语里是否存在词缀，则有争议，即使存在，也很不发达。2. 作为词缀的语素，含义极抽象,而‘英’、‘扩’、‘始’等汉字,含义像其它单词一样具体。3. 现在所谓语素，在古汉语里都是单词，而古汉语和现代汉语可以交叉使用，“始于”、“自某月某日始”这类表述现在仍很常见。最近北约开始考虑东扩，这里的“扩”显然是个单词。

扩展法看来有很多缺陷,所以王力说它是‘在没办法之中想办法’(王力,一,46 页)。

仂语

即使‘马车’、‘说话’等等不是单词,但它们总比一般自由词组结合得紧密,我们的实感总觉得它们和‘大车’、‘取笑’更近些,和‘一匹白马拉的破车’或‘说了一上午的话’远些。王力采用“仂语”这个概念，就是为了解决这个问题。貌似单词而经不起扩展法检验的复音语词，都是仂语。仂语是汉语里一个庞大的中间体，一边与复音词邻接，另一边和自由词组邻接。

然而，我们应当怎样定义“仂语”呢？我们看来只是把给词下定义的困难转移

① 陈光磊,12 页。“英”虽然不是自由形式，但在构词方面今天仍有很强的孳生能力。这种不能直接用来造句而只能用来构词的语素，作为某些单词的一个成分，有时称作“词素”。另有一些语素可能有时作为单词出现,有时作为词素,如‘取’字在‘取药方’和‘取笑’这两种情况中、‘龌’、‘龊’根本不是语素,所以既不是单词也不是词素、“词素”是 morpheme 的另一个译名。morpheme 在西语语言学里有时也意味‘词素’,不过现在多半都在‘语素’的意义上使用。关于‘语素’和‘词素’的讨论,可参考吕叔湘 93。此外,陈光磊把‘烈’视作非自由形式是错误的,‘烈’是可以单独使用的,例如‘这马的性子真烈’。

过来而已。“凡词群没有句子的作用者,都是仂语”(王力,一,42 页)。这个定义实在糟糕,按照这个定义,“马车”、“马拉的车”、“这辆从来没有任何人动过的车”,都该归在仂语里。但显然,“马车”的结构、功能和具有意义的方式都和“挡车”、“大车”这些标准的单词更为接近,和“这辆从来没有任何人动过的车”则相去甚远,故有论者说:“我们把它(‘马车’)当作一个单词看待,也不能算错,而且在习惯上是很自然的”(文炼,226 页)。可以肯定,王力心目中的仂语,只有“马车”、“说话”一类,而不包括“这辆没人动过的车”一类,所以他会说“仂语在原则上等于一个单词的用途”(王力,一,45 页)。可按照他对“仂语”的定义,偏偏要把“马车”和“这辆没人动过的车”划在类。看来,仂语概念无助于明确词的定义。①

外部标志:拼写与语音

科学偏爱具有形式标志的定义,在这一点上,语言学殊非例外。所以很多语言学家反对从意义方面来下定义。“造句单位”可由扩展法判别,但不仅例外太多,而且再追一步,扩展法仍然依赖我们对意义的了解。在哪里能找到较为确实的形式标志呢?

在拼音文字里,似乎有个简捷的办法:看一看语词是怎样书写的。然而,谁能单凭这一点来确定 roadblock 是一个词而 road machinery 是两个词呢?“海滩”在英文里有时写成 sea shore 有时写成 seashore,那么它是一个词还是两个词?至于中文这样的非拼音文字,书写特征就更不足为凭了。

其实,“语言学所谓形式就是音,并不是文字学所讲的‘形’”(赵元任,三,64~65 页)。那么,我们也许会期待语言学专家通过实测语音之间的间隔来确定词这个单位。可是这项工作得不出什么决定性的结论,因为“在实际说话时每个词之间不会有停顿”(克里斯特尔,145 页)。虽然我们可以找到词这个单位的一定语音特征,但无法单凭语音特征来确定单词之间的界限,实际上,“语义学乃是音位分析的必要条件”(乌尔曼,195 页)。有教科书称“词的语音形式独立、完整而固定”(邢公畹,119 页),然而却不曾提供任何证据资料,所以不足为凭。其实,单靠语音无法界定单词,这一点早有定论:“词是语言单位,而不是语音单位,仅仅对连续的话语进行语音分析不能告诉我们这段话是由多少个词组成的,也不能告诉我们词与词之间的界线。这是语言学家早就承认的、无可争议的事实”(叶斯柏森,108 页)。近几十年的继续测试并未改变这一结论:“如果全然不知道词的意义,简直不可能确定一个词在哪里停止而下一个词在哪里开始”(Bolinger and Sears,42 页)。

陆志韦很早就指出单根据声音来给词下定义是不可能的,而且这个办法在汉

(1) 王力曾表示:“仂语和词的界限虽然有时候分不清,在语法是不关重要的”(王力,一,46~47 页),而且他也考虑过取消“仂语”这一提法(王力,三,241 页)。

语尤其行不通,他举的一个例子是‘代表’和‘戴表’(陆志韦,14 页)。王力的看法也是这样(王力,三,240 页)。陆志伟自己采用的方法叫作同形替代法(陆志韦,15~26 页),这个方法没有被后人普遍认同,他的工作可说是后继无人。赵元任曾对这个方法的主要缺点做了说明(赵元任,三,236 页),不过他认为把某些语音特点和同形替代法综合起来考虑,“也许会产生出一个跟其它语言的 word 很相似的概念”(赵元任,三,239 页),“这些因素(重音声调等)看来至少能给出部分答案”(赵元任,三,237 页),但他同时也指出,汉语复合词固然有轻重音型这一特定的音型,但‘在这一方面,词组跟复合词没有区别’(赵元任,二,79~80 页、190 页);‘袖手’是词而‘就走’不是;‘图书馆’是词而‘他不管’不是。

词和字

为“词”下定义的工作陷入困境,根本的原因也许在于‘word 这一级单位……在汉语里没有确切的对应物’(赵元任,三,233 页)。在西语里,语言的基本的单位是词,在汉语里,是字:‘字’这个概念的外延很清楚,字读出来是一个音节,写出来是一个方块。小学校里最初一个一个教的,任何语文工具书必须一个一个解释的,就是‘字’。在中国人的观念中,‘字’是中心主题,‘词’则在许多不同的意义上都是辅助性的副题(赵元任,三,248 页)。

在一个基本的意义上,汉语的字和西语的 word 大致相当。在西语里,绝大多数的语素本身就是单词(元素词),而在汉语里,绝大多数的语素都是字。在西语里,复合词是由元素词构成的,汉语里的复合词则是由字组成的。语言学的中心任务是研究语音和语义的联系,因此,确定最小音义结合体始终是语言学的一项中心任务,而在汉语里,音义结合体就摆在字这个自然层面上,那我们似乎是该把字当作“中心主题”。

但多数汉语研究者的看法不是这样。吕叔湘说:‘用传统的‘字’来分析古汉语的语法结构,也许还可以试试,用来分析现代汉语,显然行不通了’(吕叔湘,14 页)。王力明确反对以字为基本单位,主要根据是“字在语法里是没有地位的”(王力,三,237 页);‘葡萄’的‘葡’;‘刀子’的‘子’;‘铁路’的‘铁’,都是一个字,但这里的‘铁’是一个词;‘子’是一个词缀;‘葡’什么都不是,语法地位完全不同。王力坚持以词为基本单位,是从语法的普遍性着眼的,甚至包括对中文拼音化的预期。同样,陆志韦对“词”这个概念的研究,可以说根本就是由推崇白话和拼音化促动的(陆志韦,3 页,12 页)。的确,语言学作为一门科学似乎总应力求获取普适性。字是单属于汉语的东西,“词”这个单位却是各种语言里都有的,以词为基础,似乎有利于把汉语研究和一般语言学研究统一起来。何况,现代汉语正在向复音的语言单位转变,无论在理论上还是在教学实践中,即使第一步确定了以字为基本单位,下一步仍要面临区分复音词和自由词组的问题。所以,赵元任虽然断定“在中文

里，跟英文的 word 相当的社会学上的单位就是字”（赵元任，二，72 页），但他还是下了好大的功夫来界说‘词’这个‘辅助性的副题’（赵元任，二，第三章）。再说，倘若汉语里比较接近 word 这一级的单位是字，那么，我们为什么不径用‘字’来翻译 word 呢？[①]

综观前面对“词”这个概念所作的分析，主要得出“语素（最小音义结合体）”和“自由形式（最小造句单位）”这两条线索。在西语 word 这个层次上，这两条线索大致重合：绝大多数语素同时也是自由形式，只有 trans－、－ism 等少数例外。所以，西语研究把 word 作为语言的基础层次，颇为自然。在古汉语里，在“字”这个层次上两条线索也基本重合，只有“囫囵”这样少量的连绵词是例外。如果讨论的是古汉语，我们满可以用‘字’来翻译 word。但上古以后，通过意义兼并、近义词结合、习惯用法等等，多音语词渐渐多了起来；‘语素’和‘自由形式’不再相合：字大都是语素，但其中一半是自由形式，一半不是；词汇多半是复音词，是自由形式，但不是“最小音义结合体”。所以；‘字’或‘词’都不能和 word 对上。‘语素’这条线索大致落在‘字’上；‘自由形式’这条线索大致落在‘词’上。这样我们也就清楚了：以字还是以词为基础单位，实际上争点在于侧重于音义结合体还是侧重于造句单位。像赵元任那样建议以字为基础单位，等于是说音义结合体是主导线索，像王力那样坚持以词为基础单位，等于是说造句单位是主导线索[②]。人们更经常把 word 译作“词”，这似乎表明人们更偏重于从‘造句单位’来理解 word。

依我的看法，这两条线索的重要性差不多。后面我将表明这一看法的依据。其实，虽然汉语研究者有的赞成以字为基础单位有的赞成以词为基础单位，但在具体的研究工作中差不多都是两者并重的。

① 赵元任认为汉语里没有哪个单位正好和 word 相当，但是有‘许多种形式类似 word 的单位’（赵元任，二，73 页），有时他把这些形式统称为‘word－like units’。这给翻译设了个小陷阱，word－like units 被译作‘似词单位’（赵元任，二，104 页），转了一圈，就像说‘汉语里的词（等等）只是和词相似’。这话当然不通，因为词就是词而不是类似于词。这里要绕的弯子何止于“词”？研究政治学的人岂不要说，中国社会里并没有什么东西正好和 law 相当，但有不少 law－like 的规章和建制。这个小陷阱里还藏着汉语学术的一种尴尬，即我们在学术讨论中一忽儿从汉语日常语境加以理解某个概念，一忽儿又把它当作某个外语词的译名。人为禁止这样作于事无补，不过须得意识到这一现象。

② 侧重这一条线索还是那一条线索，也体现在对其它相关问题的分歧上。当不好认定某一复音语词是单词还是“仂语”的时候，王力建议宁可把它认作仂语而不认作单词，因为汉语毕竟是以单音词为主的（王力，一，47 页）。相反，陈光磊则建议把这样的语词认作单词，这样会对句法分析更方便些（陈光磊，13 页）。

定义和概念分析

我们行到中途，有必要停一下，看看自己走到哪里了，看看我们是否已经接近了‘词’的定义。

对定义工作有一种极常见也极幼稚的误解，以为给“词”下定义就是要找出一个或一些特征，为一切词所共有，为一切非词所不具有，似乎这样的特征就构成了词的本质。。“词”这个概念从来不是从所有词的共同点归纳出来的共相，我们原不可能把天下的单词或至少一种语言里的单词都摆到眼前，以便抽绎出它们的共同之处。所谓经验主义的归纳法实是完全无视经验的主张。“词”是日常用语，我们平时用“词”所称的，原不一定具有某一或某些共同点，这一点已由维特根斯坦的“家族相似”概念讲得很清楚了。退一步讲，如果是由于所有的词都有一个或一些共同点所以都称作‘词’,像所有的兔子都具有一些共同点那样,那只说明‘词’这个词没什么概念内涵，不过是个名称而已。而本质却是属于概念的，我们不谈论“兔子的本质”。

要理解词的本质，首先需要的不是定义，而是概念分析。对“词”作所谓“概念分析”，说得最简单，就是弄清我们把这些语言单位叫作“词”而不把另外一些语言单位叫作‘词’道理何在。‘词’的概念分析回答‘词’这个概念是根据什么道理形成的,而不是回答‘所有被称作‘词’的东西的共同特征是什么’。‘音义结合’和‘造句单位”就是这样的道理。

但其中每条道理，乃至各条道理的联合，通常都不能充分界定“词”的实际用法。首先，我们不能保证我们已经穷尽了实际用法所基的全部道理；其次，一些新道理可能在不断改变实际用法；最后也是最重要的，我们不能保证一切实际用法都必定符合某种道理。

那么，“词”岂非不可能得到充分的定义了？自然概念本来就不可能得到充分定义。像‘音元’、‘语素’这样的理论概念是可以充分定义的,所有语素都是‘最小音义结合体”,没有例外。这并不奇怪,因为‘语素’不是一个自然概念,它本来就是其定义的缩写。与逻辑主义者的论断相反，概念的可充分定义性并非泛泛的优点，而只是建构精密理论的一条必经之路。

如果把‘音义结合’、‘造句单位’这样的分析结果称作‘定义”,那它就是揭示性的或概念性的定义，与理论概念的定义不是一回事。概念性的定义无需提供形式标志。本文第一节提到,从意义、观念、概念方面来定义词,人们疑虑重重,因为‘意义”等概念缺少形式标志。在概念分析的范围之内，这种疑虑是多余的。

“词”、‘句’、‘意义’等,一方面来自自然语言,同时又是语言学理论中的概念；它们是语言学的基本概念，也可说是桥梁概念，是沟通自然理解和专业理解的桥梁。。我们在为理论目的定义这些语词的时候，既要梳理其原始概念结构，又要考虑

其构建理论的功能。在这些基本概念的基础上,进一步构建起“语素”“音元”等纯粹理论性的术语。这些术语的整体构成了一门科学的语言,即建立在自然理解基础上的一种亚语言。

从建构理论的目标来看,可以把概念分析看作定义的准备工作。但“严格”的定义工作不是用来界定“词”的实际用法的,而是用来构建语言理论的。我们会纳闷,即使能够通过重音声调等等确切地界定什么是词,对平常人怎样确定什么是一个词也并无帮助。这些技术性的方法,十分繁难,我们未经训练,没有设备,所以无法应用,其结果也只是一些技术术语,完全不能增进我们对“词”这个概念的理解。然而,这些定义本来不是为了直接增进我们的理解,而是为了建构科学理论的。

概念分析是在自然理解的层面上进行的,始终依托于感觉。在这个园地里,我们经常听到研究者互相询问:这种提法是否符合实感。然而,概念的技术性越强,我们就越不可能诉诸实感来检验它。我们平常看作静止的,牛顿说“(匀速直线)运动”,从前,光和明晰可见连在一起,伦琴以来却有了“可见光”和“不可见光”的区分。技术性概念的功能是建构理论,而不是为“模糊的自然概念”提供精确的界说。我们也许会抗议说,如果科学对“运动”“光”等等的定义和我们对这些词的日常理解相悖,科学干嘛还要说它讲的是“运动”和“光”呢?它满可以为它所界定者另起一个名字,就像“语素”“夸克”这类与日常语言无关的语词。然而,科学所理解的“光”“运动”“词”,并不是与日常理解的“光”“运动”“词”完全无关的全新概念,它们是日常概念的某种变形、伸张、深入,虽然在不断的理论构建过程中最后变得和日常概念完全无关了。

有些论者完全不了解理论概念的功能,一心希望做得面面俱到,似乎面面俱到保障了定义的“科学性”。于是我们读到:“词,是意义单纯,语音形式独立、完整、固定,而且其中没有停顿的语言建筑材料单位”(邢公畹,119 页),又如:“词是代表一定的意义、具有固定的语音形式、可以独立运用的最小结构单位”(胡裕树,232 页)。这样的定义把意义的、语音的、造句的三个平面硬拼到一起,既缺乏概念分析的思想性,不能加深我们对“词”的意义的理解,也无助于构建任何理论。完备不是科学的鹄的,语言理论给“词”下定义,其目的从不是找到一条界限不多不少把日常称作“词”的东西圈在一处,不是要在各种各样的词里找出共同点来,而是要“营建”一个关于词的科学概念,可借以有效地建构某种语言理论。

语言的基本单位

回顾历来对“词”的各种各样的定义,音义结合体也好,造句单位也好,以“字”代“词”也好,不难看到,都是要确定某种“基本单位”。那么,我们径直把词定义为“语言的基本单位”行不行?这个笼统的提法也许不错,但它实在太笼统了,行得通行不通仍然要看下面怎样解释这一定义。在采取这一定义的论者那里,所谓“基本

需要全本请访问 www.ertongbook.com

单位”仍被理解为“最小音义结合体”或“最小造句单位”,或无原则地在两者之间跳跃。

我们可以换一个角度来思考这个问题。若说词是基本单位,那么,什么是不基本的单位呢?自由词组和句子也是语言里的单位,但它们既然由单词这个更小的单位组成,所以它们不是“基本的”单位。然而,句子和单词却不是同一个意义上的单位,正是由于认不清这一点,引起了经久不息的争论和广泛的混乱。从一个角度着眼,人们主张词是基本单位,从另一个角度着眼,人们又会主张句子是基本的单位。我颇喜读的一位论者,竟在同一本书里一时明称“词是语言的基本单位”,一时又称“(语言的)基本单位是句”(朱星,5 页、62 页)。

词是构成句子的单位,句子又是什么的单位?是一篇讲话的单位?一篇文章的单位?这样的思路错失了关键之点:词的用法是约定的,而句子、段落、文章等等都不是约定的。人们常说语言是交流、交通。那么,词就是我们进行交通的设施,而句子则是交通本身。交通设施是建好的,交通则依各人每天的需要进行。单词是约好的表达手段,句子则是自由表达。

约定用法与自由表达

约定用法与自由表达,即叶斯柏森所称的 formulas and free expressions,这一组范畴提示出了语言的本质,恰如叶斯柏森所说的那样,“约定用法和自由表达①的差异渗透到了语法的各个部分”(叶斯柏森,7 页);“约定用法可以是一个完整的句子,也可以是一个词组,或是一个词,或一个词的一部分”(叶斯柏森,13 页)。所以,他在《语法哲学》里开篇就讲解这对范畴。最重要的对照不是语素和单词的对照,而是约定表达和自由表达的对照:“喷饭”不能说成“喷哺”;“吐脯”不能说成“吐饭”,这是典型的约定;“我明天来”是典型的自由表达。有了“心细∞心碎”的成说,心思细碎就只好说心思细碎或心思琐碎,不能再说“心细∞心碎”;就连“吃了吗?”也应该视作约定用法。叶斯柏森就提供出很有力的理由,把 How do you do 和 Beg your pardon 都叫作约定用法(叶斯柏森,5 页)。

音义结合体、单词、以及短语等等,它们都是约定的用法,它们之间的区别在于约定/自由的程度各个不同。语言在各个层次上都有着程度不同的约定,整个语言可看作是使用完全约定的符号编制约定性质较弱而自由程度较大的符号,直至于自由的表达。

我们可以把约定大致分为两个等极。第一级的约定是语素,大致就是字,如“始”、“走”、“大”,它们是没有什么道理可讲的纯粹约定。第二级的约定依据一定

① 《语法哲学》的译者把 free expression 译作“自由用语”,这个译名不太好,因为“用语”通常指不自由的表达,近于“约定用法”。

的道理乃至某些规则，但这些道理和规则是不充分不完整的。“地震”是约定的说法，不能说成‘地动’，但‘地震’这个约定是以‘地’‘震’这两个字原有的含义为基础的，所以不是纯粹的约定，我称之为次级约定。次级约定里有一种特别的约定，如“吐脯”、“瓜田李下”、“爱你没商量”等等，它们的意思不仅有赖于其组成成分的语义，而且也有赖于某种特殊的文化、历史情况。在约定用法之外是不经约定的自由表达或自由词组。“飞机”是个约定用法；“飞来”和“大飞机”是自由表达。

基础约定，语素、字、连绵词、混一词

最基本的约定是最小音义结合体，或语素。“大”就是‘大’的意思，把‘大’分解成 d 和 a，这两个音要就没意义，要么其意义和‘大’没有关系。“雨”不指雪；“跑”不指跳，这些完全是约定的，没有道理可讲，只有单独学过才懂得其意义。

汉语里的语素大致就是字。有些字是自由形式，如‘大’和‘我’；有些字不是自由形式，如‘英’和‘始’。英语里的 trans- 不是自由形式，但其中文对应物‘跨’则应视作自由形式。

无论是不是自由形式，语素都是造词的元素；“始终”这个词是由‘始’和‘终’这两个非自由形式的语素构成的；“大话”这个词是由‘大’和‘话’这两个自由形式的语素构成的。一个语素尽管可能不是自由形式，不能自由地用来造句，但它造新词的能力体现了一定的自由度（Hockett，307～308 页）。“迢”和“遥”都不是自由形式，但‘遥’的自由程度较高，因为它具有更强的造词能力，如‘遥望’、‘遥盼’、‘遥祝’，而‘迢’已大致丧失造词能力。“英”、“始”这样的字自由程度更高，它们虽不是完全的自由形式，但在多种语言游戏里可以单独使用，包括玩笑、对联。

基础约定还包括‘鸽鹇’、“囫囵”、‘蝴蝶’这样的连绵词。连绵词是双音的，就是说，是两个字①，但也是最小的音义结合体，因为它与复合词不同，其成分没有独立的意义。

各种连绵词的约定程度稍有不同。“囫”和‘囵’都不能单独使用；‘蝴蝶’中的“蝶”却是可以单独使用的，如‘庄生梦蝶’。在现代汉语里，‘蝶’是个独立成分，构成‘粉蝶’、‘蛱蝶’等词，颇类于‘鲸鱼’、‘鲫鱼’里的‘鱼’字。不过，‘鲸鱼’、‘鲫鱼’里的‘鲸’和‘鲫’本身是语素，而‘蝴蝶’里的‘蝴’不能拆出来使用。连绵词里至少有一个字不是语素。概括说，复音词里只要有一个字不是语素，我就把它归入一级约定。依此，‘蝴蝶’是一级约定而‘鲸鱼’是次级约定。

另有一些复音语词，如打尖、扒灰、叨唠、马虎、知道、捭划、马大哈，其含义与其

① “字”是否只表示一个音节而不表示两个，这要看大多数说汉语的人是否接受“‘囫囵’这个字”这种说法。也许人们通常从单音节来理解“字”，但多少也从“最小音义结合体”来理解“字”。

组成部分没有联系[①]，或其联系已不可考，可说间于纯粹约定和二级约定之间。[②]正因为如此，所以连给它们起个名称也不容易。赵元任有时称作“不能分析的字组”或“不能分析的复合词”（赵元任，二，106），但若不能分析，怎么说是复合呢，那就干脆叫作“纯多音词”、“真正多音节词”（赵元任，三，242）。正因为它们实在“很像多音节语素”，我倾向于把它们视作一级约定，似乎可以称作“混一词”。当然，若通过研究澄清了其构成成分各自的来源，那我们就得承认它是个复合词而不是个混一词。

外语词的音译通常也是第一级的约定，如“坦克”、“逻辑”、“托拉斯”。有些译名混合了音译和意译，如“拖拉机”、“来福枪”、“乌托邦”，按照上述原则，这些仍该算作一级约定。

次级约定、派生词、复合词

“牙”和“齿”意思相同，但“齿冷”不能说成“牙冷”。同理，“人造”，不说成“人制”。“地震山摇”和“地动山摇”都说得，“七级地震”却不可以说“七级地动”。“人造”、“地震”、“齿冷”，这些都是约定的说法。不过，它们是次级的约定，因为它们以“人”、“造”这些一级约定为基础，是由其成分的意义促动的约定。

汉语里最基本的约定至少已经是一个字，那么次级约定当然两个字或更多，所以次级约定都是复音的。[③]

次级约定的约定程度有深有浅，相差甚多。“打听”的约定程度明显高于“听说”，“口吻”的约定程度明显高于“口译”。“发轫”的约定程度甚深，近乎混一词，我们很可能只知道它的整体意思而不清楚其词素的意思，或不尽能看清“发”和“轫”这两个字怎样合成了“发轫”一词的意思。另一端，“闭上”的约定程度很浅，但仍应看作是约定，我们不说“闭下眼睛”，虽然实际上我们很难说眼睛是闭上的而不是闭下的。

次级约定主要包括通常所说的派生词和复合词。

汉语的派生词结构比较简单，由一个词根和一个词缀合成，例如“桌子”、“肘

① 连绵词有声母韵母的联系，和“不可分析的复合词”不同。

② 这样的词，常常是文盲认它们为元素词，读书人认它们为复合词，所以在汉语里，“不可分析的多音节词跟可分析（或已分析）的复合词之间的界线就比其它语言都更加飘忽不定”（赵元任，三，243）。

③ “别”这个字，来自“不”和“要”这两个语素，但已经过了变形，汉语里的 b 和 ie，并不单独表示“不”和“要”的意思，所以，我们把“别”本身看作一个语素，不算二级的约定。“甭”（大概来自“不”和“用”，但 b 和 eng 单独不表示“不”和“用”的意思）、“有”等也属于这一类，但总的说来这样的情况在汉语里是很稀少的（赵元任一，142）。

子”、“老师”、“老虎”、“大李老李和小李”。词缀本身不是一个语素,但具有构词功能[①]。

借词缀构成的单词不多，词缀的数目也不大。有一些字词，该不该算词缀，尚有争论,如作者、智者,药性、革命性,氧化、现代化,激进主义、自由主义,我们、人们。我认为只有那些大致不具有意义的构词成分才应被当作词缀，所以，上面列举的这几种成分依我看都不是词缀。

组成复合词的成分，本身各是语素，复合词的含义建立在这些语素的含义之上，但这几个语素如此结合，却有约定在内，不能完全推论出来。我们后面将详细讨论这一点。

我们所说的复合词,不仅包括王力所说的复合词(“复音词”),而且也包括他划作仂语的“说话”、“扩大”等，这些语词也有约定的成分，只不过约定程度较低而已[②]。在我看，既然仂语是与自由词组对照提出的，那么无论就用途而言还是就语法而言，它们本来就该归属于复合词的范畴。

三级约定，短语

次级约定里有一些词语，它们之所以成为约定用法，不仅基于语义，而且含有特殊的历史文化因素。“吐脯”、“喷饭”,从结构上看就像“喷水”那样是普通的动名结构，两者字面上的意思也差不多，但它们的实际含义来自特定的典故，结果意思相去甚远。“玩的就是心跳”结构上是个普通句子，一旦被视作某种时代心情的标题，就成了约定用法。若加细分，可以把这类语词称作三级约定。

英语表示速度之快的词如 quick, fast 等,都和心情之欢畅没有什么联系,但学习汉语的英国人不必了解中国哪一个特别的故事，就能够理解“快”这个字同时含有速度之快和心情之畅快。然而，他不读古典或注解就不可能知道为什么“吐脯”意指求贤若渴，“高山流水”意味着知音难觅，我们也无法从任何或全或偏或显或隐的道理向他解释“断袖”的同性恋倾向。从瓜田李下和天上地下的相同结构我们也

① 词缀多半是从语素转来的,就是说,它从前是有意义的。《窦娥冤》里有句话说“不想撞见张驴儿父子两个，……那老张问道……有张驴儿数次调戏你女孩儿”，何乐士评论说，由此可见“‘老’还有些形容词的作用”(何乐士,54 页)。直到今天,“老”也没有完全失去意义,“老张”也许不太老,但也不会很年轻。在“老师”、“老兄”、“老弟”、“老茧”这些词里多多少少能体味出某种“老”的含义。不过,词缀原有的含义,即使仍有所保留,通常也比较浅淡,有时则已完全消失。“老弟”之“老”,和“老”的本义已相去甚远 要在“老师”、“老虎”、“老鼠”这些词里去挖掘共通的意义，近乎徒劳。所以我们在这里还是应当把“老”看作词缀而不看作一个独立具有意义的单位。

② 王力深受叶斯柏森的影响，奇怪的是他不曾注意到，只有从是不是约定用法着眼，才能把“用如单词的仂语”和自由词组区别开来。

推导不出同类的含义结构。

上节说到语词的约定程度有深有浅。三级约定既然在语义之外还含有文化积淀，像“蛊惑”“杀青”，约定程度通常都更深，人们常只明白整个这个词的意思，却不清楚各个词素的意思，以及这些词素如此组合怎么一来就有如此含义。

三级约定中有些是普通的复合词，如“吐脯”。三级约定中最大量的是成语、谚语、熟语、流行语、歇后语等，我总称为“短语”。成语一般指书面上的短语，谚语则指民间的短语，日常生活中新近涌现的短语，七八年搞一次、你别无选择、过把瘾就死，我称之为“流行语”。我称之为“短语”的，别的书里常称作“固定语”“固定词组”“现成说法”“习语”“习用语”等。

教科书对短语主要有两点界说，一是结构的完整性，二是意义的完整性（高名凯、石安石，107、108 页）。从结构上看，多数短语并没有什么特点，“玩的就是心跳”，原是再寻常不过的话。如果“结构的完整性”是说“瓜田李下”不能说成“瓜田桃下”；“爱你没商量”不能说成“爱你不磋商”[①]，那么这显然不是短语的特点，而是所有约定说法的共性：“口译”不能说成“嘴译”；“闭上”不能说成“闭下”。至于短语意义的完整性，仍定义为“不是它的各个结构成分的意义的简单总和”（高名凯、石安石，108 页），这也一样不成立，因为这说法也可套在复合词上。另有教科书说“组成短语的词的意义明显，……由词所组成的短语的意义则显得复杂”（邢公畹，118－119 页）[②]。我们在“简单的意义单位”一节已经批评过这种说法，“玩的就是心跳”，意义有什么不明显的？再说，“显得复杂”这种话也实在不适于用来对概念进行界定。至于说短语是“极常用的”词组，那就更不着边际了。短语不是由频率决定的；“今天下午”“看电视”，这些话出现的频率极高，却不是短语；“藕断丝连”出现频率要低得多，却是短语。其实，与复合词一样，短语之有别于自由词组，只因为它是约定的说法，只不过短语的约定牵涉到特定的历史文化来源。

约定与规则，类推

我们根据什么说“理想”是个词；“万一”是词；“你讲”和“两万”则不是单词；“一唱一和”是成语而“一大一小”不是？概括说来，我们根据什么区分约定用法和自由词组？

在谈到约定用法与自由词组的区别之时，最常见的说法是，复合词等等约定用

① 为方便、玩笑等等改动短语中的一两个字，如“拔苗助长”“有的放矢”，这不反证短语的约定性，反倒提供了额外的证明。改动过的用法不是与原约定并列的、而是以原约定为前提的语言游戏。

② 更有论者笼统地称词组（包括自由词组）表示复杂的概念（胡裕树，237 页），这就差错得更远了。

法的意义不是“它的几个构成成分(字面)的意义的简单总和”(高名凯、石安石,106页),例如:“白事”是“白”和“事”这两个字的有机结合,产生出了新的意义,不是“白”和“事”的“机械和”(陈光磊,10页)。“有机结合”(或“化学结合”)和“机械和”这些提法即使不算错,也差不多只是把所要解释的疑问换了个名称,而没有提供真正的解释。怎样算含义的简单相加,怎样算有机结合?约定用法中的成分在结合中真改变了含义吗?在“鞠躬”这个词里,看不出“鞠”和“躬”哪一个改变了含义。

所谓“约定”,就是不能根据规则类推得出。所以,叶斯柏森视自由组合法等于类推法(叶斯柏森,11页)。次级约定是有些道理可循的约定,就是说,并非完全不守规则,但又不可能尽由规则推导出来① ——若能完全推导出来,那就不是“约定”了。这也等于说:没有一定的规则告诉我们类推到哪一点就失效了,为什么就失效了。我听到“口译”“口信”,但我想说“口头提供的证据”时能说“口证”吗?我们常说“一等”“头等”“一流”,可偏偏不说“头流”。有时候意义相加虽等于相加的意义,可是不知道加起来成不成这个语言所用的词(赵元任,一,43页),赵元任喜欢举的一个例子是法语用“四个二十,十九”来表示“九十九”。

有一些单词,粗粗一看像是自由的组合,但若细加考察,便会发现其中含有某种约定的因素。

强大。固然“强”和“大”都是自由形式,而且“强大”也就意味着“强而大”,但“强而大”或“大而强”却不能说成“大强”。此亦约定不可类推之一例。

高山。我们说高山、高峰、高楼、高空,但不说“高树”“高柜”(比较“矮柜”)、“高杯”(比较“高脚杯”)。形名组合不是很自由的,特别是单音形容词。“组合不自由,就是有熟语性,这是复合词的特点。短语② 的组成,原则上应该是自由的,应该是除意义之外没有任何限制的”(吕叔湘,23页)。吕叔湘在这里所说的“熟语性”,就是我们所说的“约定”。

我们须注意,这里的关键不在于是否有时能够类推,而在于有时不能够类推。让我们再用一个例子来说明这一点。“窗子”和“桌子”“椅子”的构成是有规则的,我知道“窗”和“窗子”是一个意思,第一次听到“桌子椅子”,可以类推它和“桌椅”是一个意思。然而不知何时,类推终止了。我们说“桌子”“窗子”,却不说“门子”,说“新娘子”却不说“新郎子”,说“儿子”“孙子”“侄子”“外孙子”“小舅子”,却不说“哥子”“姐子”,说“狮子”却不说“熊子”。

① 叶斯柏森说:“约定用法的构成可以是有规则的,也可以是无规则的”(叶斯柏森,13页),意思相近,但说得不大确切。

② 吕叔湘所说的“短语”,相对于我所说的“自由词组”。

需要全本请访问 www.ertongbook.com