

The Study of Relations among Chinese Characters and Words

Xiujuan Han

Beijing International Studies University, Beijing
Email: hxiujuan@163.com

Received: Mar. 18th, 2020; accepted: Apr. 20th, 2020; published: Apr. 27th, 2020

Abstract

“Characters” and “words” are important linguistic units of Chinese. The relationship between characters and words not only affects the foundation of language system, but also the language research and language teaching based on it. To better understand the relationship between characters and words, this paper is written from three aspects: linguistic units relationship between characters and words, statistics for characters and words and teaching cooperated with characters and words.

Keywords

Relationship between Characters and Words, Chinese Teaching, Statistics for Characters and Words, Teaching Cooperated with Characters and Words

关于汉语字词关系的研究与分析

韩秀娟

北京第二外国语学院, 北京
Email: hxiujuan@163.com

收稿日期: 2020年3月18日; 录用日期: 2020年4月20日; 发布日期: 2020年4月27日

摘要

“字”和“词”是汉语的重要语言文字单位, 能否正确处理字词关系, 不但影响到语言体系基础的建立, 也将影响到建立在此基础之上的语言研究和语言教学。本文从字词的语单位关系、字词统计和字词配合教学等三个方面对字词关系进行了梳理和分析。

关键词

字词关系, 汉语教学, 字词统计, 字词配合教学

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

“字”和“词”是汉语的重要语言文字单位, 能否正确处理字词关系, 不但影响到语言体系基础的建立, 也将影响到建立在此基础之上的语言研究和语言教学, 特别是在信息技术迅速发展的今天, 利用计算机、语料库等现代化手段, 字词的定量研究进入了大规模考察统计阶段; 汉语教学的国际化、数字化对字词的关系研究也提出更加深刻的要求, 字词关系研究成为深化语言研究的方向之一。本文将从三个角度对字词关系研究进行探讨。

2. 字词的不解之缘——字与语言单位的对应

语言基本单位的确定是整个语言体系建立和研究的基础。“字”是汉语传统语文研究中的重要术语, 一直是文字学、音韵学、训诂学的研究对象, 在汉语语法研究轰轰烈烈展开后, “字”也自然而然地与语言单位的使用结下了不解之缘, 矛盾之大甚至有时候影响到对语言术语的正确理解和运用。产生问题的主要原因, 一是自古至今, 汉字在汉语研究中的重要地位; 在以往的小学研究中, 文字单位与语言单位的矛盾尚未显现, 随着语言的基本成分——“词”的概念产生而萌生; 二是汉语具有不分词连写和文字没有时态变化等特点。根据王立(2003)研究, 学界对字词关系的认识主要分为三个阶段[1]:

1) 19世纪末至20世纪20年代, 随着西方语言学理论的引入, 汉语学界“词”观念开始萌发, 但“词”观念的朦胧认识和“字”观念根深蒂固的影响, 常导致汉语语言研究中基本概念混淆。

作为现代汉语语法研究的开山之作——《马氏文通》用了大量篇幅对充当句子成分的“字”, 根据其分布条件(指字在句中所处的位置)及其功能进行了分类。以现代的定称来看, 马氏所称的“字”即为“词”, 并是在马氏语法体系中具有重要地位的单位[2]。章士钊(1907)认为词是由字到句的中间单位, “句, 集字而成者也……一词不必为一字, 泛论之则为字, 而以文法规定之则为词。”[3]; 刘复(1920)对“字”和“句”做了较明确的界定: “字是意义的最后的独立的单位, 句是意义的独立的单位”[4]。陈承泽(1922)阐述了“字与词, 虚字与实字”的问题[5]。黎锦熙(1924)“词就是说话的时候表示思想中一个观念的语词。”黎氏进而把汉语中的“字”、“词”、“语”、“句”分别与英语中的“syllable”、“word”、“phrase”、“sentence”作了对应, 至此, 黎氏建立了一套汉语句法分析单位“字、词、短语、句子”[6]。

这一时期, 许多学者已经开始从意义上来探讨“字”和“词”, 虽然观念还有着明显的不同, 例如, 刘复的“字是意义的最后的独立的单位”与黎锦熙的“只有有意义的字才可称作词”, 在“词”概念萌芽时期, “字”“词”复杂关系难以澄清是不可避免的。

2) 20世纪30至40年代, 汉语词观念逐步形成。

何容(1942)强调“意义是语言的灵魂”, “声音有了表意的作用, 才能成为语言”[7]。吕叔湘(1942)“‘字’和‘词’是很有分别的, 每个字只代表一个音缀, 可以成一个词, 可以不成一个词。”吕先生从声音方面, 意义方面等方面分析了音节、字、词之间的关系, 同时也表明了一个十分重要的思想: “词

有意义单位和表现单位之分”，这里隐含着词与构词成分(语素)的区别。吕先生的这一思想对其后研究有重要影响，启示我们可以从联结字、词的纽带——语素入手来分析二者的关系[8]。王力(1944)“在中国语法里，咱们有字和词的区别。词就等于英文的 word，字则大致可译为 a syllable represented by a character。”他主张从意义的角度来定义词，“我们以‘语言的最小意义单位’为词的定义，虽不敢说是完全的，至少，对中国语而论，它该是较好的”[9]。高名凯(1948)认为一个“字”并不一定能代表一个“词”：“‘词’一般中国学者叫做‘字’。‘词’就是英语的 word，法语的 mot，德语的 wort。一般人把它译成‘字’确是有毛病的”[10]。

字和词的关系是否清晰，直接关系到“词”观念的确立，本时期的学者开始从声音、意义等更多的方面来关注字和词。逐渐形成的共识是：词是有意义的音节，词必须是音义结合的整体。字是与音节相互对应的，当然，只有字记录的是有义音节才有可能成为词(此时还没有明确提出字与构词成分——语素的对应)，记录无义音节的字不是词。

3) 20世纪50至60年代，关于字词的关系语言学界基本形成了共识：明确了字和词是两种不同性质的单位，“词”这一级语言单位得到普遍认同，正式进入语法、词汇研究领域。

吕叔湘(1953)：“语言的最小的独立运用的单位是词，而文字的最小的独立运用的单位是字。一个词可能只有一个字，也可能不止一个字”[11]。张志公(1953)“代表一个概念的，或者是具有一种语法作用的一组声音，就是一个词。把这些音用符号记下来就是字”[12]。陆志韦(1956)“一个字是中国人造的一个具有单个形象并且连带单个读音的一个符号。”至于“字跟词的分别”，他说：“字是写来读的，词是用来说话的，可是也可以勉强用字写下来。字全是单音的，词可不一定”[13]。胡附、文炼(1955)“一般语法书谈‘字和词’，总是从字谈到词，以字为基础来谈词的。其实这是颠倒的说法，我们应当以词为中心来谈字。说语言是怎样用文字记录下来的，一个词写下来用了几个字。如果本末倒置，无疑会得出‘词是字构成’的结论了”[14]。

从“字”与“词”的“语言成分”之争，到从声音、意义上的区分，再到明确“字”是文字的单位，“词”是语言的单位，在经历了诸多讨论后，不但“词”观念日渐清晰，得到确立，而且提出审视字词关系的新视角——扭转过去由字到词的观察，而是“以词为中心来谈字”，从这个过程中，我们可以看到字词关系的复杂性以及科学对待这一问题的重要性，这是关系到整个语言学研究的基本理论问题，将影响到整个学科系统的建立。

胡附、文炼两位先生的观点很鲜明，也很有启发性，字、词关系的出发点是以词为中心，汉字如何书写和记录了词，而不是汉字如何组成词。因此，本文认为字与构词单位的相混是造成字词不清的主要原因。所以我们要弄清楚字词关系，首先要从词的构成单位入手，到底什么单位构成了词。其实对于“词”的构词成分，命名也不统一。陈望道(1940)首次提出“辞构成或语构成的成分，就是‘辞素’或‘语素’”。之后，他更倾向于专用“语素”这一名称[15]。

对于“语素”，目前大家的共识是：语素是语言中最小的音义结合体。这个概念也是受西方结构主义语言学理论的影响而引入的。语素是比词语低一级的语言单位。吕叔湘(1979)第一次把五级单位制引进了汉语语法研究，即“语素-词-短语-小句-句子”，并特别强调语素的重要性[16]。

古汉语中，单音节语素占多数，所以一个字是一个单音节语素，而古汉语中成词的语素多数是单音节语素，所以，字、词关系常常能一一对应起来，汉字的音义就源自于其所记录的语言单位——单音词(成词的单音节语素)，这也为后来字词的语言单位之争埋下伏笔。只是随着语言的演化，有很多单音词后来变成了不能独立活动的语素(不成词语素)，字词的对应关系出现了以下几种情况，显得复杂起来：

一个汉字仍代表一个单音词(成词的单音语素)：如，人、水、火。

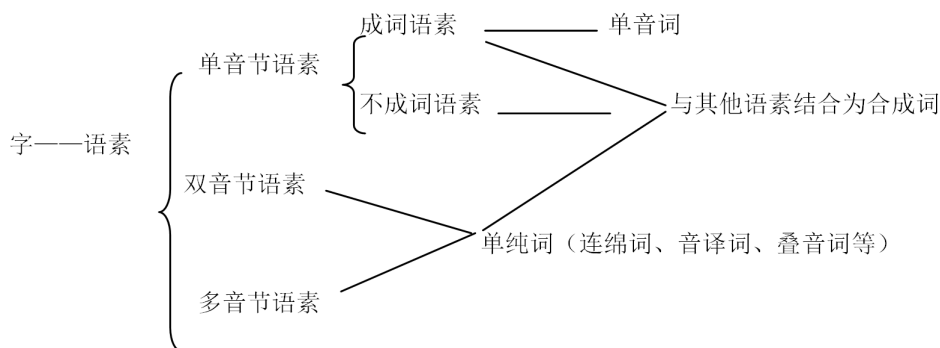
一个汉字代表一个不成词的单音语素：如，民、击、危。

一个汉字代表不同的词：如，花1(鲜花)、花2(花费)。

两个或多个汉字代表一个词：如，巧克力、沙发、蜻蜓。

汉语从古至今，由单音词为主变为双音词占优势，汉字与词语变得不一一对应，语素的地位就日渐突出出来，语素在汉语中是客观存在的，只是在汉字与单音词对应着的时候显现不出来，当双音节词语越来越多时，语素的作用就显得重要了，它成为连接字与词语的纽带。汉字进入语言中，直接记录成词语素或者不成词语素，然后以其形参与构词，由此以单字为单位，从词语中分析出来的应该是“语素义”，而不应说是“字义”，只是大家已经习惯称之为“字义”了。王宁(2002)“文字的性质首先取决于这种文字的形体与语言如何联系。汉字构形的最大特点是它要根据汉语中与之相应的某一个词的意义来构形”[17]。也就是说，汉字的本体是形，它本身并没有意义，它根据要记录的词语的意义来构建了自己的形体，其意义还是来自语言的。

综上，字、语素、词的关系可以简述为：语素是现代汉语最小的有意义的语言成分，是最小的语音语义结合体；比语素高一级的单位是词，词是最小的能够独立活动的有意义的语言成分；语素可以独立成词，也可以语素之间组合为合成词，词和词再组合为短语。汉字是记录汉语里语素的符号，汉字可以只对应一个单音词，可以既对应单音词又对应合成词的构词语素，也可以只对应不成词语素。也就是说，字和词是以语素为连接纽带的，我们可以把三者的关系用下图表示：



通过上文阐述，我们已经知道字以语素为纽带与汉语的语言单位对应着，因此，考察词语关系，弄清单字在语言环境中是如何记录汉语的，首先应该知道单字与何种语言单位对应。我们可以先简单地分出以下几类：

- 1) 记录只能单用的成词语素的单字
- 2) 记录不成词语素的单字
- 3) 记录既可单用又可与其他语素结合构词的成词语素的单字

以上三种分类，可以帮助我们明了单字与词这一级别的语言单位的对应情况。当单字记录只能单用的成词语素时，即单字与单字词的关系，进一步考察此时的单字在形音义与单字词的关系，对单字词的教学、理解和记忆都能提供很多信息；当单字记录不成词语素时，此单字常与何种语素构词，其“字义”（语素义）在合成词中有何变化；当单字记录既可单用又可与其他语素结合构词的成词语素时，可考察的信息除了以上两种情况之外，还有很多，例如，单字在何种语素义上可以单用，也可以与其他语素结合，哪些单字常单用，哪些单字常与其他语素结合构词.....

3. 字与词的统计

汉语字词的量化测定是汉语定量分析工作的重要组成部分，也是目前开展最多，成果最多的方面。统计方法从最初的频率统计，到进一步的使用度统计和通用度统计，再到近年提出的流通度统计；统计

结果从分散的数量统计，到系统的专用字表、词表，再到国家标准的《现代汉语常用字表》、《通用规范字表》等。社会是发展的，语言是变化的，我们的研究也应该是不断进步和深化的。

我们知道，字频的统计和分析，是汉语定量分析最早开始的部分，并成为现代汉字研究中的重要部分，它可以为解决汉字应用中的许多问题提供帮助，例如，教学常用字和常用词的选取。曾经，社会上有很多观点对于字频统计的分析并不全面，如：

实例一：“武汉大学研制成功的语言自动处理系统，运用 RD-11 微电脑对《骆驼祥子》进行自动处理，结论是：学会 957 个字就可以阅读全书的 95%，一字不拉地看完也只需学会 2314 个字。总之，汉字数量多，而且难认、难写、难记，的确是严重的缺点；但是由于常用字集中，掌握 1000 个常用字，就能把文字作为工具(一般小学三年级就能做到)，掌握三四千字，就能顺利阅读一般书籍报刊和书写信函文章”[18]。

实例二：“经过统计，人们发现，同是现代汉字，字与字的使用频率却天差地别。70 年代，我国用手工统计了 2165 万字的语料，共得 5991 个汉字。使用频率最高的是‘的’字，而有 345 个汉字在总字数高达 2165 万字的语料中都只出现一次。它们跟‘的’字的比差是 1:830322。汉字的这个特点，使得人们只要掌握了二三千个字就具备了基本的读写能力，从而在一定程度上缓解了汉字符号多，结构繁这个难点”[19]。

以前类似这种把识字直接等同于认读文章的论述不在少数，这种观点对于一个母语为汉语的学习者来说，也许是不错的，因为他有语言的背景知识，理解加上揣测，文章内容自然能明白个八九不离十；但是，对于一个第二语言学习者来说，在毫无语言背景知识的情况下，仅学习几千个汉字，是否可以说具备了一定的读写能力？词语所用的字如果不认识，则无法顺利认读整个词语，更进一步讲，即使认识了词语的全部用字，也不一定就能理解词语的意思，例如，认识了“开”和“关”，不代表知道名词义的“开关”；认识了“老”也认识了“板”，也不一定明白“老板”的意思。字义不能直接代表词义和语义。

其实，严格说来，对于母语学习者说，这种说法也不甚科学。我们看实例二中说：使用频率最高的是“的”字，覆盖率接近 4%，那是否能说认识了“的”字，就读懂了 4% 的文章呢？当然，这是一个我们作为引用的很极端的例子，我们在此只想说明，不能把读写能力简单地与汉字识读直接等同起来，学生学习汉字要有量的原则，更要有质的原则，到底学哪些汉字更有利于学生积累、扩展、引申、联系？学哪些字可谓是科学的汉字学习方法，起到事半功倍的效果呢？这些都是以正确处理汉字与汉语关系为前提的。

汉字和汉语的词不是一个系统，所以汉字与词语的统计要区别开来，但同时，汉字的统计也不能离开词的统计考察，更科学的方式是把从字到词的考察和从词到字的考察结合起来。单纯的汉字统计带给我们的是汉字使用频率即使用次数的数据，而词语的用字统计反映的则是汉字在汉语中的使用特征。对于研究和教学来说，汉字都必须结合具体字符使用的环境和所记录的词语才能起到全面的效果。

因此，考察汉字使用特征应把字与其所记录词语的使用情况结合起来。可是长久以来，字、词的统计分析大都是分别进行的。虽然统计汉字的指标随着频度、使用度、通用度等指标越来越科学，但还是很少从词语用字的角度考虑字的使用情况。我们并不是指“词语完全决定汉字”、“词语牵着汉字走”，而是把汉字所记录词语的情况也作为考察汉字属性的参考指标。在汉字等级确定中发挥一定的作用。

所以，从学习汉语的角度，字表的确定一定要考虑到词语用字，不能仅从字频统计的角度出发，更不能认为学习多少个汉字就万事大吉了。

笔者曾对中国主流报纸动态流通语料库(DCC¹)语料库中 5 年总字数达 11 亿 1 千 3 百 3 拾余万字的语料做过统计：词语表中覆盖率达 90% 的 22969 词，共用字 3635 字，而字频并集表中覆盖率达 90% 的汉字

¹“中国主流报纸动态流通语料库”由北京语言大学 DCC 博士研究室创建，始建于 2001 年。每年把十几种中国主流报纸的文本汇集进数据库，经过多年的努力，已经建立起一个动态大规模真实文本的书面语语料库。

只有 938 字。见表 1。

Table 1. Statistics for characters and characters list of words
表 1. 字频统计与词语用字统计

字表	交集(个)	差集(个)
90%词语用 3635 字		2698
90%字频 938 字	937	1

从这个统计数据来看，覆盖率 90% 的字频统计满足不了覆盖率 90% 的词语用字的需要。词语用字和字频统计不一致。首先我们可以知道，来自字频统计所取的单字不一定参与词语构词；其次我们可以看到，字频统计的单字与其所涉及到的词语实际用字有差额，还需另外补充汉字才能满足通用词语的学习。这是因为从汉字构词的角度看，只要一个词语中包含了此字，就算作单字“涉及”了一个词条；而从词语用字的角度看，必然包含词语所用到的全部字，例如，“嫌疑人”一词是包含在单字“人”所涉及的词条中的，但如果字表中不包含“嫌”和“疑”，那学生就无法顺利学习“嫌疑人”这个词了。因此我们这里用了“涉及”而不是“覆盖”，如果用“覆盖”来表述，则必须包含词语所用的全部单字。

对教学来说，字词等级相应、字量和词量的确定是教学大纲和教材编写面对的最重要问题。字词的教学都应当是逐步、有序进行的，我们必须了解每个阶段所应学习哪些词语和汉字，也必须科学地安排好进入下一阶段学习所要增加的内容。《汉语水平词汇和汉字等级大纲》规定了甲乙丙丁四级 8822 个词语和 2905 个汉字，2010 年的《汉语国际教育用音节汉字词汇等级划分》对 11,092 个词语和 3000 个汉字进行了三个等级的划分。这都属于较宏观总量上的指导，对于教材编写和教学方法的具体实施，需要更详尽、明确的字词量，例如可以兼类词的词性等级划分等。因此，如何根据字词等级安排教学，需要对字词关系作出更精准的量化研究后给予指导。

4. 建立字词关系数据库，探讨字词配合教学

我们所有研究的出发点和落脚点，归根结底都是为了汉语教学。近些年，“汉语热”激发了全球汉语学习的热情，但同时“汉语难学”的“定论”又打击了学习者的热情，所谓“汉语难学”，主要难在字词，但学好汉语的关键也在掌握字词。词语和汉字是对外汉语教学的重点和难点部分之一。

据吴晓春(2000)的调查,美国外交学院的学生在汉语学习过程中普遍存在着识词不识字的现象。比如,“继续”的“续”认识,但“持续”的“续”不认识;“准备”的“准”认识,“准则”的“准”不认识;“损失”的“损”认识,“亏损”的“损”不认识;“建设”的“设”认识,“设施”的“设”不认识等等[20]。造成这种问题的主要原因是以词句为基本教学单位的,字词之间缺乏有机的联系,因此,教学也必须注重字词关系研究。

在汉语教学方面,针对汉语教学的特性,很多专家提出必须关注“字词关系”,处理好汉字与汉语(词语)的关系。“在汉语研究和教学上,要正确处理汉字和语言的特有关系”(法国学者白乐桑,1996;李泉,2012);“汉字能力是汉语(词语)学习的基础条件和关键因素”;(崔永华,2008;李泉,2009)“汉字具有社会性,单字的处理要考虑到它在词中的应用”(李芳杰,1998;程荣,2003;王宁,2004;李运富,2003;张普,2005);“深化字词关系,字词需要更精准的量化研究”(张普,2005);“字词存在于一个大系统中,把汉字教学和汉语词汇教学有机结合起来”(周健,2007)。他们都强调在汉语教学和研究中要关注“字词配合教学”和“字词关系”。

考察字词关系不能再仅仅依靠小规模语料收集、手工的操作和语感的分析,必须借助计算机的语言信息处理技术,来考察大规模真实文本中的语言现实。本文第二节的字词统计已经从计量的角度给出

证据, 识字与识词既各自独立又密不可分。字词配合教学才是科学符合语言事实的方式。那如何较好地实现字词配合教学呢?

为此, 我们考虑可建立基于大规模文本考察的“字词关系数据库”, 把能反映字词关系的各种信息都综合进这一数据库中。汉语教学中, “字词关系”研究涉及很多方面, 如: 字词的统计、字词的分级、汉字与语言单位(单字词、合成词和不成词语素等)的对应、汉字的实际使用情况与词语的使用情况、汉字与词语词类的关联、汉字构词位置的方式乃至更深一层的字义与词义的关联等。科学的统计技术、丰富的信息资源和便利快捷的查询是考察字词关系的最佳方式, 也是能在教学中推而广之的首选模式。

笔者已经尝试建立“通用字词关系数据库”, 作为深化字词关系研究的“测试集”, 并已基于本文的综述和分析, 对字与词语的对应情况和多种关系展开考察, 考察结果也已另文撰述, 希望能用语言现状的事实来帮助大家认识得更清晰, 研究得更深入。

基金项目

本文为北京市教委科研计划面上项目“2017 纵向 - 科技创新服务能力建设 - 科研计划一般项目——基于数据库的汉语字词关系研究及教学应用”(SM201710031001)论文成果之一。

参考文献

- [1] 王立. 汉语词的社会语言学[M]. 北京: 商务印书馆, 2003.
- [2] 马建忠. 马氏文通[M]. 北京: 商务印书馆, 1898/1983.
- [3] 章士钊. 中等国文典[M]. 北京: 商务印书馆, 1907/1935.
- [4] 刘复. 中国文法通论[M]. 北京: 中华书局, 1920/1939.
- [5] 陈承泽. 国文法草创[M]. 北京: 商务印书馆, 1922/1982.
- [6] 黎锦熙. 新著国语文法[M]. 北京: 商务印书馆, 1924/1992.
- [7] 何容. 中国文法论[M]. 北京: 商务印书馆, 1942/1985.
- [8] 吕叔湘. 中国文法要略[M]. 北京: 商务印书馆, 1942/1982.
- [9] 王力. 中国语法理论[M]//王力文集第一卷. 济南: 山东教育出版社, 1944/1984.
- [10] 高名凯. 汉语语法论[M]. 北京: 商务印书馆, 1948/1986.
- [11] 吕叔湘. 语法讲义[M]. 北京: 中国青年出版社, 1953.
- [12] 张志公. 汉语语法常识[M]. 北京: 中国青年出版社, 1953.
- [13] 陆志韦. 北京话单音词词汇[M]. 北京: 科学出版社, 1951/1956.
- [14] 胡附, 文炼. 现代汉语语法探索[M]. 北京: 东方书店, 1955.
- [15] 陈望道. 从“词儿连写”说到语文深入研究[M]//陈望道语文论集. 上海: 上海教育出版社, 1980.
- [16] 吕叔湘. 汉语语法分析问题[M]. 北京: 商务印书馆, 1979.
- [17] 王宁. 汉字构形学讲座[M]. 上海: 上海教育出版社, 2002.
- [18] 吴峤. 汉语对汉字的影响和制约[J]. 语文建设, 1987(4): 49.
- [19] 费锦昌. 现代汉字的性质和特点[J]. 语文建设, 1990(4): 35.
- [20] 吴晓春. FSI 学生和 CET 学生认字识词考察[J]. 首都师范大学学报, 2000(z1): 133.