

西周金文字频特点成因初探^{*}

刘志基

华东师范大学中国文字研究与应用中心 上海 200062

提要 文章立足于西周金文字频研究的既有成果,分别统计了西周、春秋金文的字频,并完成了对其分布状况的断代分析。以此为基础,进行西周字频高低端集中的原因分析:字频高端集中与铭文体式、暇辞、义项数量及是否记录基本词等因素相关,而其中以涉及青铜器铭文献特点者为主导因素;低端集中首先与专名类用法字相联系,此外,又与文献用字数量的限制和文献内容对文字单位的选择性限制,乃至相关字的特定断代用法相联系。

关键词 西周金文 春秋金文 字频

中图分类号 H121 **文献标识码** A **文章编号** 1671-9484(2010)01-0080-11

1 引言

字频,是文字研究的一个独特视角,具有不可替代的研究价值。随着中文信息化的进展,汉字的字频研究近年来已呈现长足进展态势,材料对象已涉及历史文献,研究的触角也深入到了字频与词性、语义、字音及字形的关系,以及字频与文献内容、时代的关系探究等层面。(郭小武 2001)由于电脑通用字符集尚未覆盖到汉字古文字,出土古文字材料的字频研究相对薄弱,但近年来的有限研究成果,却已触及到传统古文字研究未曾涉及的研究层次和方面,同时也显示了此类研究未来发展的极大空间。

张再兴(2004)在对西周金文的文字频率进行了统计的基础上,论述了西周金文用字相对现代汉字的集中性特点,探讨了西周金文的高频字与传承率、可释读性的关系。这些发现对于我们重新认识西周金文字形系统,具有重要启发意义。但由于是西周金文研究的筌路蓝缕之作,张文也留有一些需要进一步研究的空间:首先,西周金文的字频状况,如果能与相邻时段古文字材料字频状况进行对比,当可进一步凸现其断代特点;其次,如果能够探求西周金文字频特点的成因,当可对这种字频特点获得更深入的认识。鉴于上述认识,立足张文的研究基础,本文拟作下述进一步研究:将字频统计的对象扩展到春秋金文,同时,基于原始材料的最新公布及考释研究的最新发展,重新统计西周金文字频;在此基础上着重探讨西周金文字频与青铜器铭文特点的内在联系。

[收稿日期] 2009年7月11日 [定稿日期] 2009年11月24日

^{*} 本项研究得到国家社科基金项目“基于标准数字化平台的古文字发展史定量研究”(项目号:05BYY023)和上海市重点学科华东师范大学“汉语言文字学”(项目号:B403)的资助,谨此致谢。同时感谢《语言科学》匿名审稿专家提出的宝贵修改意见。

2 西周春秋金文字频统计及断代意义分析

对古文字而言,“字频”的“字”也是需要界定的。首先,古文字尤多异体,一字而多形比比皆是。本文所言“字频”,凡不加特别说明者,均属于“字”之频率,而非字形的频率。同理,凡同形而异字者,则归为不同的字;其次,古文字中多见合文、重文(字符基础上加重文符号),系特殊文字类型,虽然其所对应的语言单位不止一个,但根据“字”的确认原则,依然视为一字。

依据严格反映原始材料实际的原则,对于作为字频研究依据材料中的已无法辨认的残泐字,以及释文中因种种原因而拟补的字亦一概予以剔除。

作为一种定量研究,字频研究所需的验证资料往往要达到海量的程度。于是便引出了一个验证研究结果的海量研究资料如何呈现的问题。毫无疑问,以传统论文形式的纸张载体是难以胜任承载这种海量资料的。笔者注意到,此前的字频论文多已暴露了这一问题。那么,如何解决这个难题呢?本文的对策是,将相关验证数据挂在网上,作为本文的网络数字附录,以供读者查询。^{〔1〕}下文所言附录,均属此种网络数字附录,不再一一说明。

字频统计,需要尽可能完整的相关文献用字为统计对象。为此我们全面收集了迄今已经发表的青铜器铭文资料,为系统掌握这数量庞大的一手材料,我们运用数字化手段,分步进行了如下工作:首先,以各种材料整理者的释文为基础,综合后出的相关考释研究成果,以确定所有文献用字的释读。其次,按照数字化处理的要求,完成所有文献用字的唯一性确定,进而进行集外字造字,生成字体,实现所有释文的全文检索等数字化处理功能。同时,释文中的每个字形都可与原拓字形及其详细出处信息对应,进而形成了一个完整的商周金文数字化平台——《商周金文数字化处理系统》(升级版)。^{〔2〕}依据该平台的统计,我们得到迄今公布的商周金文的系统文字数据。其中,字头有 5502 个,用字总量为 129397 个,器数为 13178 个。

2.1 关于西周金文字频

基于这个数字平台,剔除残泐和摹本后,共得 75166 个西周金文铭文用字,不重复单字 2869 个。字频统计的结果见表 1,限于篇幅,仅给出频率 100 以上共 136 字的具体信息:

乍_3685^{〔3〕}/寶_2967/用_2607/子_2499/父_2048/孫_1908/彝_1513/尊_1496/其_1435/永_1421/王_1392/年_1073/萬_914/白_910/毀_749/于_656/考_608/令_575/旅_561/隹_547/易_489/公_484/月_483/中_475/丁_461/休_452/乙_451/且_447/大_435/鼎_428/弔_419/享_400/皇_378/卣_372/朕_372/天_366/才_363/文_349/辛_344/女_343/對_330/揚_326/曰_318/敢_308/周_304/不_290/師_289/鬲_280/又_272/吉_267/壽_260/癸_256/事_253/史_244/冊_240/侯_231/宮_229/初_226/己_222/姬_215/右_213/既_211/人_208/母_208/鬯_206/首_198/臣_194/立_189/生_185/頤_183/季_182/亞_182/

〔1〕 查询方式为:登陆“<http://www.wenzi.cn>”,在“中国文字数字化资源库”栏点击刘志基的“重大科研项目成果验证资料”。

〔2〕 《商周文字数字化处理系统》(光盘),刘志基、张再兴等主持开发,该成果的升级版在 2003 年版的基础上,增收了 2003 年以后新公布金文材料,并吸取新的考释研究成果完善了释文,其部分内容已挂在华东师范大学中国文字研究与应用中心网站——文字网上,以供全球使用者检索查询。

〔3〕 该数字为“_”前字的频率数,各频次的铭文出处详见附录:《西周金文 GBK 字频率出处表》、《西周金文 CJK 扩展字频率出处表》、《西周金文新造隶定字频率出处表》和《西周金文原形字频率出处表》。以下同类数据仿此,不再一一说明。

乃_181/無_181/拜_177/宗_177/五_177/十_175/余_172/庚_163/鬲_163/顯_154/金_153/各_152/室_147/成_147/康_146/氏_146/克_145/貝_145/彊_141/戈_140/三_139/孝_138/馬_138/乎_137/尹_137/從_136/命_136/多_135/内_135/號_134/魯_134/正_132/先_130/屯_127/日_125/戊_125/井_122/霸_121/四_121/甲_121/壺_120/自_120/即_120/唯_118/入_118/以_117/赤_114/田_114/姜_113/黃_112/追_112/卿_111/累_109/友_109/鐘_108/𠂔_106/一_106/市_105/受_105/𡗗_105/二_104/頌_102/保_101/福_101

以上西周金文字频数据相对殷商甲骨文字频^{〔4〕},除了最高端的 401—500 频率字群及 1 频字以外,各频次文字的使用总量均有较大幅度提高,其总合数达到 44360,占总字量的 56%,这比甲骨文的同口径数据上升了近 13 个百分点。这一差异大致勾勒出了西周字频的断代特点:相对殷商时段而言,西周金文少数极高频字用量过大的状况有了明显改变,这意味着极高频字的部分记录语言单位的职能已转移到其他频次字上,而其他频次字会获得更多的被使用机会。然而,由于西周金文最低频度字与殷商文字的被使用量大致相当,这就意味着极高频字让出的被使用机会在西周时代基本上都转移到了中间频次字,这就导致了中间频次字在西周金文中的地位上升。

2.2 关于春秋金文字频

同样基于《商周金文数字化处理系统》(升级版),剔除残泐和摹本后,共得 25083 个春秋金文用字,不重复单字 1456 字。字频统计所得数据详见表 1^{〔5〕},其中 100 次以上字为:

子_1045/用_876/孫_816/之_763/其_653/乍_636/永_556/寶_439/吉_393/王_309/壽_306/自_278/萬_271/年_262/無_258/隹_256/余_242/公_241/月_234/于_225/初_211/正_199/金_191/鐘_185/保_183/不_172/大_167/亥_162/侯_153/享_152/皇_152/中_150/丁_145/白_144/擇_136/父_136/𩇛_130/弔_130/鼎_120/爲_118/簠_117/鑄_115/者_105/元_105/我_105

字频分级	字数	总使用量	占不重复单字 总数比重%	占铭文用字 总量比重%	单字平均使用量
500 以上	20/7	29564/5345	0.70/0.48	39.33/21.31	1478.2/763.57
401—500	11/1	5024/439	0.38/0.06	6.68/1.75	456.73/439
301—400	14/3	4873/4008	0.49/0.2	6.48/4.02	348.07/336
201—300	20/11	4820/2728	0.70/0.76	6.41/10.88	241/248
101—200	71/25	9768/3643	2.47/1.72	13.00/14.52	137.58/145.72
11—100	494/303	15797/8903	17.22/20.81	21.02/35.50	31.978/29.38
6—10	255/163	1956/1253	8.89/11.20	2.60/5	7.67/7.69
2—5	741/441	2122/1262	25.83/30.29	2.82/5.03	2.86/2.86
1	1242/502	1242/502	43.29/34.48	1.65/2	1/1

表 1 西周金文与春秋金文字频数据比较表

〔4〕 甲骨文字频具有“两端集中”特点,即少数高频字占文献用字总量的高比重(高端集中)和占文献用字总量极低比重的低频字占不重复单字总数的极高比重(低端集中),详见拙文《简论甲骨文字频的两端集中现象》(《语言研究》待刊),相关数据详见附录《〈今译〉〈小屯〉〈花东〉用字频率表》。

〔5〕 本表所列各项数据中“/”前为西周金文数据,“/”后为春秋金文数据,涉及文字、频次及各频次出处。凡属西周金文的见附录:《西周金文 GBK 字频率出处表》、《西周金文 CJK 扩展字频率出处表》、《西周金文新造隶定字频率出处表》和《西周金文原形字频率出处表》;凡属春秋金文的见附录:《春秋金文 GBK 字频率出处表》、《春秋金文 CJK 扩展字频率出处表》、《春秋金文新造隶定字频率出处表》和《春秋金文原形字频率出处表》。

纵观春秋金文与西周金文的字频数据比较,可以获得这样一个认识:因为春秋金文总字量大大少于西周金文,凡属两者之间绝对值的比较都不具有实质性意义,只有比重类的比较数据值得关注。就“占总字数比重”列的数据来说,需要注意如下指标:1)300次以上的春秋金文数值均低于西周金文;2)2至300次的中间段频次字春秋金文数值均高于西周金文值端;3)一见字的春秋金文数值低于西周金文。就“占总字量比重”列的数据来说,除了一见字以外,情况与“占总字数比重”列完全类同,而一见字的数值则春秋金文略高于西周金文,恰与“占总字数比重”的情况相反。

如何评估这个字频比较的实际意义呢?延续着殷商到西周的字频状况演变趋向,春秋金文较之西周金文,高端集中度继续有着明显的下降,而2至300次的中间段频次字群在字集系统中的比重继续有明显提升,这无疑体现了文字系统成熟度的提高。而一见字的字数比重和字量比重的反比例现象,实际上反映的是这一最低端频级字并未因时代推移而有明显变化:西周金文一见字占总字数比重虽然高于春秋金文,但差距并不很大;而春秋金文一见字占总字量的比重略高于西周金文,则与春秋金文总字量低于西周金文有关。所以,总体而言,从西周金文到春秋金文,字频的最低端高度集中状况得到了延续。当然,这是一个值得进一步关注的现象。

前文的统计和初步分析表明,春秋时段的字频的高低两端高度集中的状况虽较之殷商甲骨文有所变化,但相对后世文字特别是传世文献而言,依然构成一种明显的断代特征,据统计,《左传》、《论语》、《诗经》等文献字频最高前10个字的频率数合计占文献总字数的19.276%(郭小武2001),而西周金文同口径数据为28.7%。同时,作为这种特征得以体现的文献类型又与殷商甲骨文具有很大的不同,因此,进一步就两端集中的形成因由进行探讨有助于我们更加深透地把握汉字字频的发展演变规律。鉴于春秋金文用字总量相对偏少,同时考虑与甲骨文时间跨度的对应性,这一探讨仅以西周金文为对象。

3 高频集中于少数字的原因分析

西周金文高频字集中于少数字的原因,可以从这些字的用法特点去探求,由此得到如下几种内在联系。

3.1 铭文程式与高频集中

西周金文的高频集中与此时段铭文的套路有着内在联系。西周时期,青铜器铭文发展到了它的成熟期,而这种成熟的重要表现,就是表达的程式化程度提升。因而涉及这种程式化表达的文字也就获得了更多在铭文中露脸的几率。这种内在联系,可以从居于前位的最高频字的程式作用分析窥见一斑。以下以频率数的高低为序,解析西周金文前10个高频字与铭文程式的关系。

乍:铭文通常要说明作器因由及作器者,故表达铜器铸造之词总不可免,而记录这个词的正是“乍”。

宝:铭文通常有形容铜器珍贵之词以及希望人们珍视铜器之词,而这个表示“珍贵”的形容词和表示“珍重”、“宝爱”的动词都由这个“宝”来记录。

用:凡铭文通常要交待作器因由,而表示这种因果关联的连词即是这个“用”字。

子:铭文一般需要首先交待作器,或作为作器因由的事件发生日期,而干支组合正是铭文表达这个日子的基本符号。作为地支第一位的“子”,自然会在铭文中出现频率较高。另外,“子”在铭文中常常义同“子孙”,而铭文多有祈福子孙之词,这当然又会进一步提升“子”的出现次数。

父:铭文一般需要记述所作之器祭祀的是哪一辈分的先人,而作为祭祀对象,父亲当然是最主要的,故“父”字多见于铭文就是非常自然的事情。

孙:祈福子孙,是铭文程序的一个重要环节,而“孙”不仅总在这种语境中与“子”连缀,又每每单独可

以与“子孙”同意。

彝:作为青铜祭器的共名,“彝”多须在铭文的最末处出现。

尊:“尊”与“彝”同意,也表示祭祀用器,且每每与“彝”同意连缀。

其:铭文总不免祈福之词,而作为情态副词,“其”既表示希望、祈使,又表示意志,故每每出现于祈福之语词中。

永:在铭文程序性的祈福用词中,“永”的地位重要,表示福禄永久,长远,并成为重要构词成分,如“永世”,表世代相传,绵延久远;“永令(命)”,表长命,长寿;“永寿”,表示永远,世世代代。

综上,我们可以获得这样的认识:西周金文的前 10 位高频字之所以获得其高频的地位,很大程度上得益于青铜器铭文程式因素所提供的出场机会。当然,这 10 个字与程式的联系程度也是有差异的,这一点,只要分析一下后文给出的它们在铭文中的具体用法就不难作出精确的评估。

毫无疑问,程式因素不会仅仅惠顾前 10 位字,10 位以降的高频字,也同样会不同程度获得程式的提携而提升其实际出现频次。为进一步揭示其间的内在联系,我们不妨再聚焦于西周金文中程式因素的重要组成部分——嘏辞。

3.2 嘏辞与高频集中

所谓嘏辞,即青铜器铭文中的祝福词语,从程式化因素的考虑,我们将其限定在铭文内容中与作器直接相关的祝福语段。因为它们通常是作为铭文的一个必要组成部分而出现在铭文中的,所以嘏辞中的常见用字也将因此而提升其出现频率,当然,这种较高的频率,并不反映真实的断代字频。

为证明这种联系,我们选择马承源(1990)所收的西周器的铭文用字为对象作相关统计。《商周青铜器铭文选》(马承源 1990,以下简称《铭文选》)收录西周金文 494 篇,铭文字数达到 22246 个。作为选本,其规模可谓无有出其右者,完全可以满足字频研究所需要的文字数量要求。更为重要的是,《铭文选》在几乎囊括西周铭文重要篇目的同时,又剔出了同铭同文等因无价值重复而可能对字频研究造成消极影响的材料,进而有效地提升其作为字频研究材料的资料价值。

为揭示嘏辞对西周铭文字频的实际影响,我们首先需要统计《铭文选》所收西周铭文字频状况。统计结果表明,《铭文选》所收西周铭文共计 1437 个不重复单字,计重复出现则为 22246 字,以字频从高到低为序分列如下:

300 次以上 8 字,计 3860 次;200—300 次 10 字,计 2457 次;100—200 次 25 字,计 3572 次;50—100 次 49 字,计 3474 次;11—50 次 243 字,计 3474 次;5—10 次 189 字,计 1448 次;2—5 次 449 字,计 1368 次;1 次字 463 个。^{〔6〕}

为客观而有效地考察嘏辞对字频的实际影响,我们以随机抽样方式来限定材料范围,设定两个筛选标准,一是单纯祈福者,二是复音者。如此限定后所得嘏辞 24 个:

純魯/純祿/純祐/綽綽/大福/爾鬯/祓祿/黃耆/吉康/康娛/需終/魯福/魯壽/魯休/眉壽/彌生/日用享/壽考/通祿/萬年/無疆/永福/永命/永饗

以上嘏辞涉及不重复单字 33 个,统计这 33 字的在西周铭文中的总出现次数则为 2273,平均每字频率为 68.88,这比同统计范围的西周铭文单字平均频次 15.48 高出 4 倍以上。当然,这种较高频次的出现,不仅仅是嘏辞的原因,但嘏辞在其中的作用也是不可忽视的,仅以上举嘏辞“萬年”中的“萬”为例,

〔6〕 相关字及其频次的具体出处见附录《〈商周青铜器铭文选〉西周器铭字频表》。

其所涉及的两个字仅因“萬年”这个赅辞的使用就分别达到 61 次,而“萬”所有的出现频次实际只有 75 次,前者占后者的比重达 81.33%。

3.3 铭文用字义项状况与高频集中的关系

毫无疑问,高频字集中于少数字并非都是程式和赅辞因素作用的结果。因此,撇开程式和赅辞的进一步调查无疑是非常必要的。而直接聚焦相关频次字的用法当可揭示其特定出现频次的底层因由。

为提高调查的科学性,我们将分别以前文涉及的两种字频调查材料为对象来完成相关统计分析,以求获得多个层次的证明数据。

前文言及,通过对一手金文著录的直接整理,我们形成了包括迄今已公布青铜器铭文材料的《商周金文数字化处理系统》,而前文的字频研究首先是基于这一材料进行的。这一材料,因为更能保持铭文的原始面貌,因此应该成为首先聚焦的调查对象。

从用法的角度来窥探高频字的成因,下文将完成对前 100 个高频字在青铜器铭文中的用法及其数量这样一项调查。关于用法及其数量的确定,我们依据在金文单字义项汇集方面比较完备的《金文形义通解》(张世超等 1996,以下简称《通解》)来进行。《通解》所汇集金文用法虽然包括殷周两代,但其主体却在西周和春秋,尽管有少量用法属于殷商和战国,但毕竟也是青铜器铭文字用,作为同一文献系统,殷商和战国铭文的字用与西周春秋铭文的字用也是具有内在传承发展联系的,因而前者对于揭示后者字频状况的成因同样也具有意义。因此,我们在统计中并不删除用法中的属于殷商和战国铭文的部分。统计结果如下:^{〔7〕}

乍 3685_17/寶 2967_2/用 2607_6/子 2499_10/父 2048_6/孫 1908_3/彝 1513_3/尊 1496_3/其 1435_11/永 1421_5/王 1392_6/年 1073_1/萬 914_5/白 910_7/于 656_10/考 608_7/令 575_11/旅 561_10/隹 547_12/易 489_7/公 484_11/月 483_1/中 475_6/丁 461_2/休 452_7/乙 451_3/且 447_4/大 435_38/鼎 428_3/弔 419_2/享 400_8/皇 378_4/朕 372_5/卣 372_12/天 366_5/才 363_7/文 349_18/辛 344_3/女 343_10/對 330_2/揚 326_2/曰 318_4/敢 308_5/周 304_8/不 290_24/師 289_5/又 272_10/吉 267_6/壽 260_7/癸 256_2/事 253_12/史 244_4/冊 240_4/侯 231_6/宮 229_7/初 226_4/己 222_3/姬 215_1/右 213_10/既 211_12/母 208_5/人 208_9/首 198_2/臣 194_7/立 189_9/生 185_7/季 182_3/亞 182_5/乃 181_9/無 181_9/宗 177_14/五 177_4/拜 177_3/十 175_1/余 172_5/庚 163_3/鬲 163_5/顯 154_3/金 153_7/各 152_6/室 147_4/成 147_10/氏 146_7/康 146_9/克 145_7/貝 145_2/彊 141_3/戈 140_3/三 139_6/馬 138_2/孝 138_8/乎 137_2/尹 137_6/從 136_11/命 136_10/多 135_10/內 135_12/魯 134_10/號 134_4/正 132_12

综合上面的统计结果所呈现的信息,不难取得如下两个方面的认识:

1)用法的数量与字频数成正比的,这 100 字的用法数达到 683 个,每字的平均用法将近 7 个。为确定这个数值的意义,我们又统计了频率仅次于这 100 字的西周金文频次数前 101 至 200 位的 100 字的《通解》汇录用法^{〔8〕},其总用法数为 501 个,其平均单字用法 5.01 个,比频次数前 100 字的同口径数据少了将近 2 个;而《通解》共收录 2448 个铭文用字,其用法总数为 5740 个,平均每字 2.3 个用法,这个数

〔7〕 以下数据以“/”分隔每个字的字频信息,每字之后的数据为其频次数,“-”后的数字为该字《通解》所录用法数,限于篇幅,将具体用法略去。

〔8〕 详见附录《西周金文前 101—200 位高频字用法》。

字只有前 100 个高频字用法数的约三分之一。值得注意的是,前 100 个高频字中的少数的用法少的字,其用法本身都具有不得不频繁出现的理由,如只有一种用法的 4 字:“年”、“月”、“十”、“姬”,前两个属于计时的基本词汇之列,而标明年月又是铭文书式;第三个属于计数的基本词汇之列;而第四个作为姓氏用字本属于专名类,但却是周天子之姓。而两种用法的“馬貝寶揚弔對首癸乎丁”,其实也都有相同或相类的用法特点,不烦一一说解。

2) 高频字倾向于与普通词汇,特别是基本词汇对应。应该说,以今天的眼光来看,青铜器铭文是一种非常特殊的文献,这种特殊性决定了它会大量使用一些相对特殊的专名性词语,诸如职官名、祭祀名、赏赐物、氏族名等等,但是在这前 100 个高频字中,却几乎找不出此类专名,例外的除了上文说到的属于周天子之姓的“姬”以外,似乎只有一个“亞”字属于职官,但是它除了有职官意义以外,还有类似今日“次于第一”的意义,即所谓“亞祖”之“亞”。由此来看“亞”也具有普通词汇身份。

以苛刻的眼光来审视,以上高频字的字用分析尚有这样一个缺陷:虽然逐字列出了《通解》所收的所有用法,但却没有各个用法出现次数的信息,这就无法获知这些用法之间的数量差异。当然,这对高频字的高频成因探求无疑是个不利因素。为此,我们再次聚焦于《铭文选》,在对其所收 22246 个西周金文用字逐一确定用法并进行释义的基础上,具体统计该书所收西周铭文用字频率过百的字的用法及其出现频次,以进一步从字的用法的角度剖析的高频促发因素。

《铭文选》100 频次以上的字共 43 个,为节省篇幅,用简略表达方式给出各字频率及其各用法的频次信息。首先以“用”字为例解说这种简略表达方式:

用 789_6//介词,相当于“以”291/连词,表相承关系 255/使用 239/连词,用于复句中,表因果关系 2/珍重,宝用 1/永远,长远 1

“//”之前为需给出频率信息的字(即上例的“用”,字下加点)及其字频数(上例的“789”)和用法数(上例的“6”),字频数和用法数之间用“_”分隔。“//”后用“/”分隔每个用法的信息。用法信息包括用法的具体描述(上例最后一个用法信息中的“永远,长远”)及该用法的频次数^[9](上例最后一个用法信息中的“1”)。其余 42 字同类信息如下:

王 720_8//殷、周时期的最高统治者,天子 244/周开国君主姬发,或写作“珽王”2/诸侯王 10/西周国君 1/人名,乖伯之父 1/读为皇,大也 1/鼻王,周穆王之子,典籍作共王、恭王 1/周文王 3;乍 503_6//犹行也,事也。即从事,作某事 6/为也。作为,成为 12/制作,特指铸造 457/匿行也,做事不光明磊落 1/建立 6/福也,典籍作祚,与后福字同义连用,作赐福义 1;寶 502_10//永远,长远 1/珍重,宝用 213/珍贵的 208/通保,保守 2/宝用,珍重 1/保有,宝贵 50/宝贵的,珍贵的 1/宝贝,珍惜之物 12/珍重,宝爱 1/保佑 1;其 350_12//语助词 37/子孙,后代 1/人名 6/其中 1/句中语气词 6/句中语气词 1/句首语气词 2/假为基,基业 3/第三人称代词,做定语 5/第三人称代词 6/表示希望,祈使 186/表示祈使,希望 49;于 350_6//介词,引进行为动作之受事对象 101/介词,引进行为动作之处所,方位 113/介词,引进时间范围 1/动词,往 4/介词,引进行为动作之受事对象 1/介词,引进行为动作之施事者,则构成被动句式 1;永 324_6//人名 1/句中语气词 1/长命,长寿 23/长久,使动。1/永远,长远 271/永久,长远 6;子 322_22//贵族子弟 2/官名 4/儿子 25/人名 3/对男子的敬称 1/干支词,多用于纪年、月、日或表示序列 6/后代 121/即“太子”,王之子 4/帝王的称呼 92/君长 1/犹后生 1/世代 1/四夷的国君被称为子 1/孙子 1/幼畜 2/子之长者,别于次子 1/族名 1/爵名,实即国君之称 4/嫡子 1/代词,代人 1/子孙 1/弟子,门徒 1;令 292_10//

[9] 该频率的出处详见附录《〈商周青铜器铭文选〉100 频次以上字用法统计表》。

任用 4/赏赐 3/寿命 1/天命 4/天命,即寿命 25/以册书布命 26/命令、使命 1/命令 141/介词,相当于“以”1/人名 5;年 283_3//岁,载 134/人名 1/虚指时间久远 123;隹 258_3//句首语气词 204/句中语气词 20/介词,相当于“以”1;易 252_2//赏赐 182/受,受赐予 19;孫 249_4//子孙,后代 122/孙子 26/即孙 1/后代 90;尊 240_4//祭祀用器 226/同“遵”,按照,遵循 4/彝尊,彝器。4/尊,祭祀用器。4;休 231_4//训为“赐”14/息止 1/人名 1/美也 187;月 220_1//记历单位词,一年十二月 210;父 217_6//父亲 7/人名,即伯雍父 4/人名,卫康叔之子康伯 7/宗族亲戚 58/男子之美称 9/人名 77;公 215_10//祖先名 1/祖先尊称 31/周公子明保 3/周公旦,武王之弟 9/征伐东尸的军事首长之一 2/文王之子毕公高 2/国君通称 47/人名 27/庙号 1/对贵族中人之敬称 11;才 192_5//语气词,表示感叹 1/规定,约制 1/感叹词 1/裁决 1/在 165;考 190_9//对亲长的尊称 1/孝道 1/享献、进奉 4/考寿,老寿,长寿也 2/敬也,顺也 6/父亲 127/长寿也 2/长寿 1/祭祀,祭享 1;女 187_5//第二人称代词 71/读若如,往也 1/女性,妇女,与士、男相对 2/用为“嘏”,福也 1/女儿 1;曰 186_4//言,说 99/句首叹词 3/读为谓,告诉 2/为,是 8;白 175_6//白色 3/爵名 57/君长 30/用为伯 3/人名 32/奴隶头领 2;揚 170_6//称扬,颂扬 134/告退的一种姿态 1/人名 2/称扬 25/宣扬 1/颂扬 1;彝 165_5//宗庙祭祀所用之彝器,多是为宗族或某人所作 4/常 1/鬯彝,多种彝器之共名 15/彝器 105/彝,青铜祭器之共名 39;對 165_3//报答,称扬 157/人名 1/配合,应和 2;敢 159_4//犹言“岂敢”1/无畏 37/人名 1/谦敬副词,冒昧、斗胆 98;不 154_4//大也 12/否定副词 38/大显,大明,显赫之极 63/语词 2;朕 153_6//第一人称代词,作定语 3/读为“骏”,长久 1/读为“盈”,多,满 1/陪嫁 9/第一人称代词 105/人名 1;𠄎 149_6//即“厥”,义为其 57/句首或句中语气助词 4/第三人称代词 19/统兵之官 1/人称代词,指发言者自身,作定语 4/指事代词,用于近指 5;又 143_11//拥有 3/方位名词,右侧 1/存在动词 26/词头 3/举行 1/连词,用于数字间 72/引道 1/佑助 1/再次 2/相当于“以”“用”,使用 1/官名 2;周 132_8//雕饰也,典籍多作“雕”2/周王室 10/国名 1/西周的政治、军事中心 30/西周国都镐京 24/邑名,西周旧邑 1/周王朝都城 46/西周王朝 9;𠄎 129_3//祭祀用簋,特加“尝”字以名之 1/假为“饔”1/礼器,以盛黍稷 117;且 128_5//远祖 1/祖父 65/人名 1/先祖 22/与“高祖”相对 3;皇 124_4//羽饰 1/大也,美也 79/即“辉煌”1/借为光,赞誉 1;事 119_19//事件 1/任用,服事 15/治事者,官吏 1/指六卿、狱官和诸侯 2/职事 22/执掌宣读王之策命的史官 1/有关国家政治之事 1/使令,典籍作“使”13/事奉,对待 30/遣使 6/祭祀 2/官名,执事 1/官名,卜官,卜官之长 1/官名 3/官吏 4/辅助,辅佐 1/出使 1/谒见天子,听命受事 1/官,治人者 3;首 118_2//头 37/颀首,上古跪拜礼 69;文 114_5//“文王”合文,周文王名称的专用字 3/专指周文王,乃“文王”之省称 3/周文王 4/人名 1/褒词,美也,与皇用法相似 82;乃 107_4//语气助词 10/意为“若”,用于假设复句 1/第二人称代词,作定语 41/指示代词,作定语,犹是、此 10;余 106_5//某种赏赐的物品,或说是笏 1/良马也 1/第一人称代词 51/第二人称代词 2/地名 1;侯 103_4//地名 3/受封为某地诸侯 6/王畿内的诸侯 5/诸侯国君 57;拜 103_2//行拜礼,动词 25/拜礼,名词 74;師 101_4//官名 12/职官名 44/军队 2/官名,官位不详 1。

综观上表所呈现的信息,我们首先可以看到可以与前文对西周金文频高前 100 字的《通解》用法统计分析相印证的情况:首先,频高与用法数量同比,43 字的用法数量达到 262,平均每字 6.09 次。其次,高频字对应基本语言单位的倾向。这里所说的基本语言单位大致相当于“基本词汇”,具有全民性、稳定性、能产性的特点。基本语言单位本身在语言交际中就有很高的出现频率,所以记录这种语言单位的文字也必然会有相应的被使用几率。如上表中的“月”,就其用法来说只有一次,但因其所记的是个基本时间词,所以频率也达到了 210。

当然,由于用法频率这个新数据的出现,我们还可以发现一些更具本质意义的信息。比如,在用法

数量和对应基本语言单位这两个因素中,后者似乎具有更加主导的作用。具体表现为:在高频字一般具有的较多用法中一定具有若干基本语言单位,而且通常是在诸多用法中使用频率较高的。如以下 35 字选自《铭文选》100 次以上的 43 字,并只列出了它们对应基本语言单位的用法及其频率:

殷：礼器，以盛黍稷 117/白：爵名 57/白：君长 30/寶：保有，宝重 50/寶：珍贵的 208/寶：珍重，宝用 213/不：大显，大明，显赫之极 63/不：否定副词 38/才：在 165/對：报答，称扬 157/父：人名 77/敢：谦敬副词，冒昧、斗胆 98/公：国君通称 47/侯：诸侯国君 57/皇：大也，美也 79/毕：即“厥”，义为其 57/考：父亲 127/令：命令 141/乃：第二人称代词，作定语 41/年：岁，载 134/年：虚指时间久远 123/女：第二人称代词 71/其：表示祈使，希望 49/其：表示希望，祈使 186/且：先祖 22/且：祖父 65/師：职官名 44/事：事奉，对待 30/事：职事 22/首：头 37/首：颔首，上古跪拜礼 69/孫：子孙，后代 122/王：殷、周时期的最高统治者，天子 244/文：褒词，美也，与“皇”用法相似 82/頤：行拜礼，动词 25/休：美也 187/揚：称扬，颂扬 134/彝：彝器 105/易：赏赐 182/永：长命，长寿 23/永：永远，长远 271/乍：制作，特指铸造 457/朕：第一人称代词 105/佳：句首语气词 204/子：后代 121/尊：祭祀用器 226

显而易见的是,它们所对应的基本语言单位一般都较少,通常是一个,一个以上的只是少数,所以总数只有 46 个,这个数字只占这 35 字的实际用法(219 个)的 21%。而这 21% 的用法涉及铭文中的实际频次达到 5162,占了这些字总字频数(7860)的 66%。少数与基本语言单位对应的用法能够左右实际字频的绝大多数,这充分表明了与基本语言单位的对应对字频状况的主导性作用。

值得注意的是,在以上所论的导致高频集中的三个因素中,第一、二个与青铜器铭文用字特点相联系,只有第三个与一般文献用字规律相联系,因此,我们可以认为,青铜器铭文的特点才是导致西周金文高频的主导因素。而这一点,即使从第三因素,即铭文义项状况的角度进行分析,也同样可以得到印证。从上述 43 字中可以筛选出以下 8 个字的 8 种最具青铜器铭文文献特色的用法:

對:报答,称扬 157/敢:谦敬副词 98/考:父亲 127/揚:称扬,颂扬 134/彝:彝器 105/易:赏赐 182/乍:制作,特指铸造 457/尊:祭祀用器 226

显而易见,这8个用法的共同特点就是都具有很高的频次,其频次合计达到1486,平均频率185.75。然而,这8个字并不是所有用法都体现文献特点的,前文所列数据表明,它们总计用法达39个,也就是说,这8个用法在总用法数中只占20.5%,但其涉及的实际字频数却占到这8个字的39种用法的合计频次(1692)的87.8%。青铜器铭文特点在字频中的作用由此可见一斑。

4 低频字多字数集中的原因分析

低频字为什么只能有较低的出现频次?从其用法,即所记录语言单位性质入手,不失为一种最为实在有效的探究途径。为此,我们依然将探究的目光聚焦于《铭文选》中的低频字的用法。为了更加清晰地揭示低频成因,我们仅以最低频的一见字为材料对象。

经释读标注,可以确认《铭文选》一见字为 463 个,其中有三字尚有释读障碍而用法不明,故实际进入分析范围的是 460 字。

与甲骨文一见字多专名的情况相类似,这 460 个金文中一见字也有半数是专名类的用法。其中,地名 40 个:

富涪嬰罈陟叙冀甯蔽_不上陞蜀還亭_數甞_釐刑邇隊陵埜洩跣涇鼎泉_和愆_雷鼓早_并虞陶晃_戒休_日

人名 61 个：送魑耶榭厖謐聃詵卓陪赫龠同昇𪔐^𪔐𪔐奄攷宋尅罍絳奚息歎^𪔐眉句親賢晤^𪔐^𪔐^𪔐^𪔐

趁衤豈杙欒爰留髡旛頽媼冉葵巛^𪔐彖散遘淪妬誠^𪔐嫗丰

国名类 23 个: 晋 腺 過 蔞 彭 密 弈 筭 鄭 潘 寺 溝 桐 尿 鄒 蝟 杞 薛 办 罽 峯 邪

邑名类 8 个：盤頰楸𣎵罽𣎵言瀘

族名类 15 个：瘡 疔 吳重 疔 疔 疔 疔 疔 疔 疔 疔 疔 疔 疔 疔 疔

其他专名 77 个：皇隣禦杓歟粥粿油諒廚甕吳𠵼穀奎粒璜鞞嶇鑄敝鐔帑革蔗斂狃轄渙鎗岢陁疊倅
茜宥櫛蒿姚姦臣輶鴈目糴綈褌襍軫榜禽轉旆旻毳刻誰戴牙鏤鋪塵縷厖韐鉭經單翼卑環河丹鬩祭

此外,还有合文 8 个:𠂔𠂔𠂔𠂔𠂔𠂔𠂔𠂔。合文虽然未必算得上特殊用法,但这种文字形式本身就是古文字之中特殊的。

以上专名及特殊字符类共计 231 个,超过一见字总数的半数。前文言及,专名类文字的低频出现,是与早期文字的图画因素遗存相联系的,而西周金文此类一见字的半数比例,虽然较之甲骨文略低,但总体上还是显现了这种遗存的影响依然颇有力度。然而,另外还有 227 个非专名类一见字:

哀敗麻倅陂自妣分瀕兵飽裁宍側儕履竅寵吹詞酖僭各狄電席都厥截而繁放泐韃焚鳳制賦攷羔莠割
功攻局蠱貫載鯨海閤竊吳宏窠祜辭寃潢灰惠諱獲刀擠儂軌集隼劑加姦緘間劍漕峯鐘經竟開罔綱上
判質鯨款羸旻遵牢饅迨緝琤音梁敵稟靈廬鹵數律廡弘灤匡矛陟昧逾牡木幕睦匿宁毆芾申撲宙棄前
擗戟驛任伏桑審盛溼時变筮咎圻厖謁宿肅遂所它吳芄虞錫韃強囹安維珥泚巫析徙閑白憲肖鵠凶項替
龜鸛句下邛甌央就錫喙匄嬰遺改因覲縈憂尤尤由囀於圉下儼鬱困帀下孟擇賊酬倂貌昭醜遺戢直陲屍廡
晝逐奔猷墜酌字總皐

这又当如何去认识呢？从理论上说，以上非专名类一见字的出现大致有两种理由：文献用字数量的限制和文献内容对文字单位的选择性限制。具体到《铭文选》的西周铭文用字，这两种限制都有很强的作用力。就其总字数而言，仅 22246；就其内容而言，只是纪念、颂扬性质的文字，这两者作用力的叠加不可能保证字集中的所有单位都有公平的出现机会。如“囧”和“旬”字，在甲骨文中都属最高频级，很难想象它们在西周时段就真的会突然沦落到只有难得露一下脸的地步。具体分析一下，很容易发现个中缘由。甲骨卜辞“旬亡囧”作为吉凶占问的熟语几乎每辞必见，而西周金文虽然也需要记时，但却很少用这个“旬”字；而义涉灾祸的“囧”当然更不会成为作为纪念、颂扬性文字的铭文热门词语，而它在铭文中的用法也是与灾祸无关的“太”义。

由此,我们不难发现,在这些非专名一见字中,有些只是由于《铭文选》西周铭文用字这个特定文献范围没有给予其公平的出现机会才沦落为一见字的,如:賊(祸害)、憂(忧虑)、隄(视,见)、遺(给予)、擇(择取)、帀(军队)、晝(白昼,与夜相对)、所(处所)、木(树)、集(成就)、獲(获得)等字都很可能属于此类。

当然,也不能一概将看似常用的上述一见字都归为“不公平待遇”的受害者,因为其低频地位正反映了它们在当时在一般文字使用中的实际状况。如“於”,后世作为引进行为动作处所的介词,当然是很多见的,但在殷商和西周,记录该语言单位的却主要是“于”,“於”却是极为罕见的;“因”、“而”作为一个连词,后世也属于基本词之列,但从西周铜器铭文来看,承担后来“因”的职责的却主要是“用”字,承担后来“而”的职责的却主要是“以”字。由此可知,某些字看似反常的低频,乃是与其特定断代用法相联系的。

参考文献

- 郭小武 2001 古代汉语极高频字探索,《语言研究》第3期,69—84页。
- 刘志基等 2003 《商周文字数字化处理系统》,南宁:广西金海湾电子音像出版社、广西教育出版社。
- 马承源 主编 1990 《商周青铜器铭文选》,北京:文物出版社。

张世超等 1996 《金文形义通解》，京都：中文出版社。

张再兴 2004 从字频看西周金文文字系统的特点，《语言研究》第 3 期，84—88 页。

作者简介

刘志基，男，1955 年 4 月生，山东荣成人。华东师范大学中国文字研究与应用中心教授。主要从事文字学研究。

Preliminary Study of the Causes of the Character Frequency on Bronze Inscriptions of the Western Zhou Dynasty

Liu Zhiji

Center for the Studies on Sinogram, East China Normal University, Shanghai 200062

Abstract This paper conducts statistical study on the character frequency of the Bronze Inscription of the Western Zhou Dynasty. It further reveals that high frequency is related with inscription format, semantic entries, and basic and core vocabulary. Low frequency is related with proper nouns and their usages, along with the limitations on the amounts of characters allowed in works, etc

Keywords the bronze inscriptions of the Western Zhou Dynasty the bronze inscriptions of the Spring and Autumn Period character frequency

—— 致 谢 ——

《语言科学》编辑部感谢匿名审稿专家在过去一年里对本刊的支持和帮助，正是在大家的共同努力下，本刊的学术质量得以保证，在此对各位专家学者的辛勤劳动，谨致以诚挚的谢忱和崇高的敬意！

以下是 2009 年为《语言科学》匿名评审稿件的专家学者（按音序排列）：

鲍厚星 蔡 冰 常 辉 陈 剑 曹广顺 曹志耘 陈昌来 陈广忠 陈前瑞 陈小荷 陈振宇 储泽祥 邓晓华
丁邦新 董秀芳 方 梅 方光柱 方环海 方绪军 方一新 冯春田 冯胜利 顾 黔 耿振生 古敬恒 郭小武
洪 波 胡 方 黄 征 何九盈 何亚南 何元建 何自然 黑维强 胡安顺 胡松柏 华学诚 黄尚军 黄天树
黄笑山 江 荻 蒋绍愚 李 蓝 黎新第 李爱军 李锦芳 李晋霞 李如龙 李艳惠 李宗江 梁银峰 林华勇
林连通 林伦伦 刘 利 刘 钊 刘大为 刘丹青 刘广和 刘俐李 刘纶鑫 刘淑学 刘晓南 刘秀莹 陆丙甫
麦 耘 马贝加 马鹏举 马清华 马秋武 满在江 宁继福 彭 静 彭玉海 齐沪扬 石 锋 孙 蓝 邵敬敏
沈家煊 施春宏 石定栩 石毓智 史金波 束定芳 税昌锡 宋文辉 苏宝荣 苏晓青 孙道功 孙良明 孙宜志
孙玉文 谭景春 唐正大 陶红印 汪 平 王 楠 王 宁 汪维辉 王冬梅 王建军 王培松 王为民 吴福祥
邢向东 徐 杰 徐大明 徐赳赳 徐阳春 徐以中 严修鸿 杨宝忠 杨荣贤 杨荣祥 杨通银 杨永龙 姚振武
殷国光 游汝杰 于 亮 袁 宾 余光武 俞理明 袁毓林 岳中奇 张 洁 周 韧 曾晓渝 詹卫东 张安生
张伯江 张成材 张洪明 张珊珊 张树铮 张文轩 张亚军 张谊生 张玉金 张豫峰 赵长才 郑张尚芳 周长银
周国光 周统权 朱习文 朱晓农 竺家宁 庄初生

（《语言科学》编辑部）