DOI：10.15916/j.issn1674-327x.2024.01.008

试论UNICODE全汉字构件拆分的原则与流程

郑瑶瑶
(渤海大学 文学院，辽宁 锦州 121013)

摘 要：汉字构件拆分是汉字构件研究的重要一环，确定拆分原则是构件拆分研究的首要工作。UNICODE字符集所收汉字数量众多，来源广泛，汉字变异情况纷繁复杂。《信息处理用13000.1字符集汉字部件规范》的出台，为汉字构件拆分研究提供了统一标准。然而，随着字符集不断扩展，逐渐产生了游离于拆分原则外的汉字。因此，确定普适性的汉字构件拆分原则及流程至关重要。论文以扩展后的字符集为研究对象，在学界已拟定的多种拆分原则和国家相关规范的基础上，建立了一套相对完善的UNICODE全汉字构件拆分原则与流程，以期在其指导下科学高效地完成UNICODE全汉字构件拆分工作。

关键词：汉字构件；UNICODE；拆分原则；拆分流程
中图分类号：H122 **文献标识码**：A **文章编号**：1674-327X (2024)01-0038-06

汉字的本体是字形，对汉字形体构件的研究，尤其是对汉字构件的拆分研究，一直是汉字研究工作中的重要一环。近年来，学界对拆分原则的研究情况纷繁复杂，在不同原则指导下，拆分实践结果也大相径庭。仅就“现代汉语3 500常用字”的拆分结果而言，晓东[1]得到474个最小部件，其中有195个成字部件；费锦昌[2]6得到基础部件290个和常用复合部件94个；韩秀娟[3]得到546个构字部件；柳建钰等[4]拆出基础构件517个，并计算平均参构能度为6.77。此外，受拆分对象及目的影响，各种汉字拆分原则之间存在歧异，对当前的汉字拆分实践产生了一些负面影响。

鉴于此，我们结合前贤的研究成果，在对UNICODE字符集扩展A区（CJKA）6 592个汉字（此处考察的“汉字”是一个广义的概念，既包括记录汉语的汉字，也包括日本和字、朝鲜汉字、越南喃字以及我国壮族所使用的古壮字等汉字文化圈的表意文字。）进行拆分的基础上，尝试制定了一套针对UNICODE全汉字比较科学完善的汉字拆分原则与基本流程，这项工作有助于进一步提高汉字拆分原则的普适性，从而保证UNICODE字符集汉字拆分工作的科学性和高效性。本文将对UNICODE全汉字构件拆分的原则与流程进行阐释，以就教于方家。

一、汉字构件拆分原则

汉字数量庞大结构复杂，因此，汉字构件拆分存在较大难度。制定系统有效的构件拆分原则是解决这个难题的关键一环。我们以UNICODE字符集所收录的九万多个汉字为对象，在总结前人研究成果的基础上，秉承“从形出发、尊重理据、立足现代、参考历史”[5]的汉字构件拆分原则，参考柳建钰[6]为字书字料库构形属性界面制定的明确拆分规则，制定出了一套适用于扩展后大字符集且更为具体可行的汉字构件拆分原则。该原则从拆分对象、拆分下限、依理拆分、依形拆分和拆分处理五方面内容入手，旨在从多层面为构件拆分提供原则性规范，从而避免构件拆分的随意性。

（一）拆分对象

汉字构件拆分的对象为合体字，即由两个及两个以上基础构件组构而成的汉字。基础构件也叫形素，是汉字的基础构形元素，这一概念由王宁[7]首次提出。按此原则，独体象形字与独体指事字不可拆分。需要强调的是，汉字拆分对象除较为明显的形声字、会意字外，还包括合体指事字和含记号构件的字。合体指事字是由象形字和指示符号组成

收稿日期：2023-05-01
基金项目：国家社会科学基金重点项目(20AYY018)；国家社会科学基金重大项目(21&ZD296)(22&ZD294)；辽宁省社会科学规划基金项目(L21CYY002)
作者简介：郑瑶瑶(1998-)，女，河北唐山人，硕士生。

的，如“刃”“凶”“亦”“本”“末”“耒”等字。这些汉字中，不仅象形构件能参与构形、体现构意，而且指事符号构件也具有同等效力。从汉字构形学角度，这类由象形字和指示符号组成的合体指事字也应该被拆分。含记号构件的字是指组构该汉字的构件因汉字演变和书写等因素，造成其构形理据丢失，无从追溯，但从形体来看确是可独立的构件形体，只是充当记号构件使用，这类汉字也应拆分。虽然以上的指事构件和记号构件没有实质表示形、音、义的属性，但其构形功能不容忽视。要而言之，汉字构件拆分的对象是由两个及两个以上基础构件组构而成的汉字，包括形声字、会意字、合体指事字和含记号构件的合体字。

（二）拆分下限

构件的拆分下限是基础构件，此原则关乎汉字拆分到何处为止。基础构件能作为汉字拆分的下限，要同时满足“最小单位”和“参与构形”两个条件。从生成角度讲，笔画是书写汉字字形时的最小单位，但考虑到要保证汉字拆分出的单位具有特定的构意，所以应该选择比笔画更大一级的构件更为恰当。能参与构形是指某字中拆分出的单位具有构意。构意反映汉字形体所携带的意义信息，这是分析汉字构形不可缺少的部分。按照笔画多少，可以将基础构件分为单笔画基础构件和多笔画基础构件两类。

1. 单笔画基础构件

有构形功能的单笔画构件可作为汉字拆分的下限。基础构件作为汉字构形的基本单位在组构汉字时能体现构意，即有构形功能，这是构件与笔画的本质区别。费锦昌首次提出“让单笔画部件取得合法地位”[2]1，打破了傅永和“构件下限大于基本笔画，小于复合偏旁”[8]观点的长期局限。一般来讲，当笔画小于构件时汉字构件拆分到基础构件就会停止，如组构汉字的“亻”“忄”“刂”“扌”“礻”等构件被拆分出后，不再继续拆分成笔画；当笔画等于构件时，此时的笔画具有构形功能，可以将其拆分出来，如“一”在“正”“怨”“亡”“弓”等字中具有记号功能，“丶”在“刃”“太”中具有标示功能。上述“一”“丶”并非单纯的笔画，而是具有构形功能的单笔画基础构件。这种情况的单笔画单位可作为汉字拆分的下限。

2. 多笔画基础构件

多笔画基础构件也可作为汉字拆分的下限。多笔画基础构件与复合构件容易产生混淆，复合构件是从构件的可拆分性角度划分出来的构件类别。李运富[9]141 认为，如果某个构件还能够进行下位拆分，就意味着它是由两个或两个以上的下位构件组合而成的，这样的构件叫复合构件。我们认为：多笔画基础构件是拆开后各部分均为非字构件且均不再构成其他汉字的构件。非字构件与成字构件相对，非字构件又称作非成字部件、不成字部件，是依附于其他构件来体现构意的构件，这种构件本身不能独立存在，无法与语言中的词对应[10]101。换言之，非字构件不能独立成字。但非字构件也具有构形功能，能够与其他构件组构汉字，如“惺”同“狂”，对“惺”判断是否拆分，可以看拆分后各构件是否成字且是否再构成其他汉字。首次拆分“惺”为“忄”和“星”，二者均为非字构件，但“忄”“星”仍可构字，如“忄”参与构字的“快”“情”等，“星”参与构字的“馑”“歘”等。所以，多笔画基础构件应该同时满足不可拆分和拆开后不可构字两个条件，如“朩”“戋”“才”字理据丧失的字，没有继续拆分的依据。以“朩”为例，理据记录仅有“韩国拼音‘chon’”，若不考虑理据强行将其拆分为“朩”“丶”，那么二者的构形功能也无从知晓，更不能参与其他汉字的组构，所以此类汉字只由一个多笔画基础构件构成。

（三）依理拆分

依理拆分是处理 UNICODE 全汉字拆分的首要原则。采用此原则的前提是：UNICODE 全汉字的字形及其理据有所保留或能追溯。对于理据留存的汉字，要严格依据构字理据进行拆分。另外，也存在部分构件经过变形出现丧失理据的情况。但王宁认为：“现代汉字理据大量保存是历史事实。”[11]2 李大遂[12]统计对外汉语教学用字，得出约 90%的常用汉字仍具有理据性。从多个方面来观察汉字构形理据的变化可以发现，汉字的理据是历史的承袭且依赖总体的构形系统，当然也存在因汉字理据丢失游离于构形系统之外的现象。汉字演变造成构字理据存在三种结果：理据留存、理据丧失和理据重构。在汉字构件拆分时，要充分尊重汉字构形理据，做到有据可依。依理拆分时，可分为有古文字形体和无古文字形体两种情况。

1. 有古文字形体

在依理拆分时，有古文字形体的分析理据主要参考《说文解字》（以下简称《说文》）中隶变以前的篆书，其次可以参考其他古文字形体或学界成果判定。王宁指出：“大多数人都承认，隶变以前的古文字，是存在构形理据的。”[11]2 加之相比于图画性极强的甲骨文、金文和异写形体丰富的战国文字，篆书既

保留了原始构意又得到了专家规范，能够为汉字拆分提供较为客观的形体与构意标准。如“風”的小篆字形作“[illegible]”。《说文·虫部》：“風，八風也。風动虫生，故虫八日而化。从虫，凡声。”[13]284 故可将其拆分为构件“凡”和“虫”。类推之，简化字“风”也仿照“風”拆分为基础构件“几”和“乂”。当然，经过演变，也产生了一些现存形体与其古文字形体、字理的记录情况不对应的汉字，如后文的“曹”“为”等字，处理这类有古文字形体的现代汉字要采取依形拆分的原则。

2. 无古文字形体

依理拆分时，无古文字形体的汉字要根据现代汉字的形、音、义要素进行拆分。这就需要借助《玉篇》《汉语大字典》《汉字海》等国内外现存各类字、韵书以及汉字构件的形、音、义对应信息。如“乶”“[illegible]”“⿱加ㅇ”一类韩国汉字，并无古文字形体可供参考，只能根据字义和读音进行判断。以韩语音译字“⿱加ㅇ”为例，其韩语拼音为“gang”或“deong”。从读音上判定，构件“加”在“⿱加ㅇ”中起提示读音的作用，“○”与“⿱加ㅇ”的形、音、义皆无直接关联，但“○”又出现在多个韩语音译字的构形中，如“⿱丷上ㅇ”“⿱仒ㅇ”等字，因此认为“○”在以上字中充当记号。

（四）依形拆分

依形拆分原则适用于拆分理据丧失或留存理据与现代字形不对应的汉字。此时拆分汉字并无确切的理据可依，只能采用从形出发的原则。苏培成提出确定汉字的组合层次还包括“从形”的原则，即“从形的原则，应该叫单纯字形原则，就是说只考虑字形，不考虑与字形相关的字音和字义”[14]。受其启发，我们在处理无理据留存或留存的理据与现代字形并不对应的汉字时，不必受制于无音、义内容的局限，可采用“从形”原则对汉字拆分。相比于依理拆分，依形拆分具有很强的主观性，因此为避免拆分混乱，建立统一的依形拆分原则至关重要。通过拆分结果可将拆分原则归纳为依形拆开和依形不拆两种情况。

1. 依形拆开

依形拆开主要适用于构件相离、相接和搭挂的情况。“分隔沟”[15]86 是处理相离和相接构件的重要标志。根据构件间距离的长短，应采用先长后短的原则拆分相离构件。如“召”之金文隶定字“⿰酉⿱召儿”。理据内容贫乏，且现存理据与汉字形、音、义内容无法对应。根据其构成单位间的距离的长短，可将其拆分为“[丁/酉/凶]”和“⿱召儿”两个直接构件，而后再依次拆分成各级间接构件。处理相接的构件，存在分隔沟的单位应从接点处拆开。这意味着，并非所有相接的部分都拆开。如“⺊”，按“⺊、丬，二音补录。”并无拆开相接之处的必要。再如“⿱丿虫”，音“chóng”同“虫”。从“⿱丿虫”之形体出发，“丿”“虫”存在明显的分隔沟，所以将其从接点处拆开，得到“丿”“虫”两个构件。极少数不影响结构的搭挂情况按照相接处理。如韩国音译字“⿰都乙”，韩语拼音 dot，据其读音，可拆分出提示读音的构件“都”。处理“都”的构形单位“者”时，存在明显的分隔搭挂情况，所以可将其拆分为“耂”“日”两个基础构件。

2. 依形不拆

一般认为依形不拆的情况是交重。若从形出发，交重笔画组成的是基础构件，不应该进行拆分。苏培成认为：“这是从形出发得到的重要原则，叫作‘交重不拆’，这是完全正确的。”[15]87 对于无理据和字形字理不对应的交重，我们一律采用依形不拆的原则。如无理据的“乂”字，仅留存“同‘五’”的信息，若强行将其拆分为“丿”和“㇏”，不仅无法促进分析“乂”之构形，而且也会增加多余无用的“构件”，同属此类的还有“屮”“[illegible]”“夹”等。再如理据和字形字理不对应的“为”字，“为”是“爲”的简化字，根据古文字字形，“爲”是由两个表义构件“爫”“[illegible]”组构而成的合体字，现代字形“为”与之不对应，所以不将“为”的相交部分拆开。

（五）拆分处理

拆分处理原则是对以上拆分原则的补充，是对拆出构件的处理规范，其中包括构件注明功能、部首变体保留、相离构件复形和高频异写字形拆分。

1. 构件注明功能

拆分汉字得到的直接构件必须具有明确的功能。前文提到，分析汉字构形的两个重要方面是构形和构意。构件注明功能是归纳构件构意范畴、外化构件构意的重要手段。换言之，构件功能是构件构意的外在形式，即使汉字形体因演变发生了或大或小的变化，其携带的构意信息无论保留还是丧失，构件功能都应该予以说明。王宁将构件功能总结为象形功能、表义功能、示音功能和标示功能等四种，对应组合成 11 种构形模式，简称“十一书”[10]138。李运富认为构件具有象形功能、表义功能、示音功能、标示功能和代号功能五种，最终形成“二十书”[9]145。综合运用王宁和李运富提出的构件功能概念和类别，字书字料库中将构件的功能分为表形功能、表义功

能、示音功能、标示功能和记号功能五种，并为所有拆分出的直接构件确定了构件功能，为汉字构形系统的构件功能与平均参构能度数据分析提供了方便。以 CJKA 中 6 592 个汉字为例，依理拆分去重后直接构件共 2 833 个，其中表形构件数量为 0；表义构件 425 个，占比 15.00%，平均参构能度为 15.68；示音构件 2 166 个，占比 76.46%，平均参构能度为 2.85；标示构件 6 个，占比 0.21%，平均参构能度为 1.17；记号构件 236 个，占比 8.33%，平均参构能度为 1.47。以上数据更新截至 2023 年 11 月 1 日，数据内容仅供参考。

2. 部首变体保留

部首变体拆分后不需要恢复原形。一是因为部首变体不恢复原形并不影响对其功能的判断和对整字的认识。汉字部首变体种类较多且数量庞大，在进行构件拆分时，这类部首变体应该保持拆分前的形态。比如“ 刂”“亻”“礻”等部件，被拆分后不用恢复原形为“刀”“人”“衣”等。二是有助于统计部首的形变情况。例如“心”字，在充当部首时，存在“忄”“⺗”两种变体形态，为方便对两种构件进行构字统计，可以不恢复二者的“心”字原形。需特别说明的是，书写避让是为使汉字结构整齐严谨，部分构件产生细微形变的现象，因此严格意义上不同于部首变体，需要在拆出后直接恢复其原形。如“圴”的表义构件“土”，“炒”的表义构件“火”，“鵷”的示音构件“元”等。

3. 相离构件复形

因构字需要造成构件相离的，拆分后仍将相离部分组合，保留构件原形。这一原则是为了不影响对构件的分析。比如“囂”第一次拆分为直接构件“㗊”和“页”，保留了构件“㗊”的原形，《说文 • 口部》：“㗊，众口也。从四口。读若戢，又读若呶。”[13]48 随后再将其拆分为四个基础构件“口”。在 CJKA 中也有大量同类例子，例如独体字“衣”参与“衰”和“褒”的构形，二字分别拆分为直接构件“衣”和“谷”，“衣”和“執”；构件“辩”组构的“瓣”“辫”“辦”，三字分别拆分为直接构件“辩”和“心”，“辩”和“刂”，“辩”和“片”；“行”组构的“衕”和“銜”分别拆分为直接构件“行”和“亢”，“行”和“缶”；“弱”字，《说文 • 弓部》：“弱，强也。从二弓。”[13]270 “弱”作为构件组构的“溺”“鰯”“爇”分别拆分为直接构件“弱”和“耳”，“弱”和“育”，“弱”和“鬲”。

4. 高频异写字形拆分

由楷书字形异写产生的构件，即使按照古文字不应拆分，也要进行拆分，以便字形异写状况的统计研究。汉字是表意文字，因此汉字构形有理据可循，现代汉字经过隶变、楷化和简化后，形体上产生了较大变化，出现了大量异写情况。通过溯源，我们可以沟通异写字形和原初字形的关系，但是为方便异写状况的调查，异写字字形的拆分可以不考虑汉字原初字形的构字理据。例如“鹿”，在古文字中作“  ”“  ”，独体表形，不应拆分。但在楷书形制下，“鹿”字经常发生异写。例如“廘”“庶”“庶”等。为了方便对异写部位的调查描写，故将“鹿”拆分为“声”和“比”。适用于此原则的还有“鼠”“虎”等字。

二、汉字构件拆分流程

汉字构件拆分原则确定后，拆分实践工作才能顺利展开。本文的拆分实践工作是先采用字料库自动拆分，而后进行人工检验，对自动拆分中少数不合理之处进行人工修改，最后形成一套普适性流程（见图 1）。从而充分保证对 UNICODE 全汉字拆分实践的高效性、科学性和准确性。

拆分流程图中共包括六个判断框和六个指示框。流程的运行机制是将汉字及其构件置于判断框，根据判断结果及指示方向对汉字构件拆分进行引导。具体判断框对不同情况的汉字都有区分的作用。以下结合具体汉字实例对不同判断框内容进行介绍。

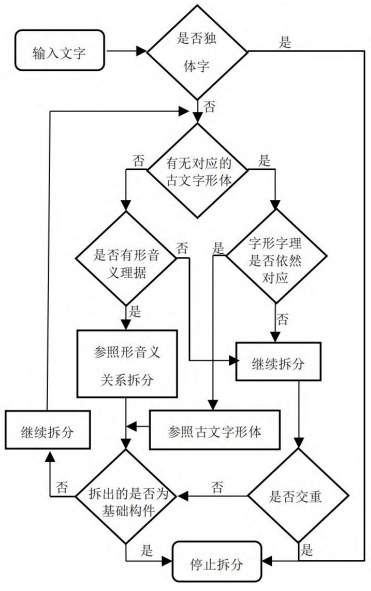

图 1　汉字构件拆分流程图

（一）确定拆分对象

此流程图中只对输入的汉字本身进行判断，如对“丛”进行拆分时，便不会受制于其繁体字“叢”的拆分。确定拆分对象程序依赖于“是否为独体字”的判断结果。独体字是由一个构件组构而成的，如

“㔾”“羊”“戈”等。按照通例，独体字不作为拆分对象进入流程，若强行将独体字拆分，不仅会得到大量无用构件，而且将会干扰其他汉字的拆分结果，造成汉字构件拆分与研究的混乱。由两个或两个以上基础构件构成的合体字可以作为拆分对象，顺利进入拆分流程，如“獷”“犨”“鬮”等字。《玉篇·犬部》：“獷，犬黄色也。”[16]本义是黄色的狗，首次拆分出表义构件的“犭”和示音构件的“矍”。“矍”作为成字复合构件，继续拆分，可得到“眢”和“頁”两个直接构件。如此循环，最终可以得到“口”“百”“八”等基础构件。

（二）对照古今形体

汉字拆分过程中，存在共性问题：现代汉字字形经过历史演变，部分构意在演变后的现代字形中无从体现。因此，追溯汉字古文字形体，使古文字形体与构意相联系，进而指导汉字构件拆分是保证汉字构件拆分结果科学性的重要一步。目前现代汉字存在有古文字形体和无古文字形体两种情况。

有古文字形体的汉字，拆分时要充分参考其形体。对照古文字形体与字理，现代汉字存在字形字理对应和字形字理不对应两种情况。字形字理对应的如“叟”字，首先考虑其古文字形体“  ”，且《说文·攴部》：“叟，㑒也。从攴，丙声”[13]68，因此可将“叟”拆分为表义构件“攴”和示音构件“丙”。字形字理不对应的如“更”字，从小篆字形上来看，“更”属于形声字，应拆分为构件“攴”和“丙”。但从现代字形来看，字形字理并不相合，因此从形出发判断“更”的拆分情况。“更”符合依形不拆的原则，可以认定“更”为独体记号字。

汉字形体不断产生，加上汉字简化的影响，如今无古文字形体的汉字占有一定比例。此类汉字要参考现代汉字形、音、义理据的保留情况拆分，如越南字“碗”“働”。“碗”越南语拼音为“phả”，根据读音“phả”与“pò”的相近性，且“乁”未见其他职能，因此将“碗”拆分为示音构件“破”和记号构件“乁”。“働”，越南字释义在汉语中同“住”，对其进行形、音、义关系对比，“働”被拆分为表义构件“亻”和示音构件“助”。

（三）输出拆分结果

输出拆分结果包括两个，一是拆分汉字时得到的基础构件，二是待整个汉字拆分完成后得到的构件拆分示意图。基础构件是汉字构件拆分到何处为止的参照对象。判断框“拆出的是否为基础构件”是判断拆分是否停止的最后一步。基础构件应满足此构件拆开后的部分均为非字构件且不再参与构形的条件。构件拆分示意图是对汉字各层次拆分的完整展示。汉字的拆分层次存在三种结构：一是汉字由构件一次性组构而成的平面结构，如“㭎”“猋”；二是汉字由构件按层次组构而成的层次结构，如“灚”“曦”；三是综合平面结构和层次结构的综合结构，如“靐”“垚”。

三、汉字构件拆分例析

本部分选取典型的字形字理对应的“晉”和字形字理不对应的“曹”进行拆分演示。在此选取的演示字例，其汉字本身及各级构件拆分满足普遍汉字拆分情况，因此对其他通用汉字拆分流程不做重复演示。

（一）晉

“晉”为“晋”之古字。《说文·日部》：“晋，进也，日出而万物进。从日，从臸。”[13]138其甲骨文字形“  ”和小篆字形“  ”相同且与《说文》对应，故首次拆分“晉”为表义构件“臸”和“日”。尹黎云在《汉字字源系统研究》中提道：“徐锴本云‘臸声’，可从。臸和晋乃屑、先对转音，臸训‘到’，而到正是进的结果，故晋从臸得义，也从臸得声。”[17]因此，首次拆分的构件“臸”除具有表义功能外还具有示音功能。按《说文·至部》“臸，到也。从二至。”[13]137认为构件“臸”为复合构件，实则是“至”的复体形式，故将其拆分为两表义构件“至”。“至”的字理内容可参考说文《说文·至部》：“鸟飞从高下至地也。从一，一犹地也。象形，不上去而至下来也。”[13]247且“至”与其古文字形体“  ”并不对应，故采用依形拆分的原则，将构件“至”拆分为记号构件“厶”和“土”。最终得到字形字理对应的“晉”的拆分结果（见图2）。

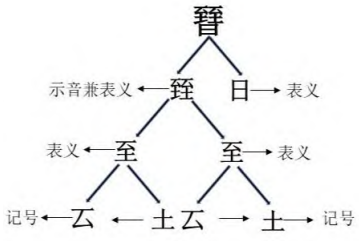

图2 “晉”构件拆分示意图

（二）曹

“曹”《说文·曰部》：“曹，狱之两曹也，在廷东。从棘，治事者；从曰。”[13]100其本义是诉讼的原告和被告。而香港中文大学人文电算研究中心《汉语多功能字库》认为，甲骨文从二“东”，后加从“口”。“东”像袋子，二“东”像一对袋子之

形。“口”是增繁符号，本义是一双、成对（丁山、林澐），引申为曹偶、同辈。“曹”为合体字，其古文字形体为“  ”。构件“棘”的字形以二“东”相对，字形演变经历了多次变化：“”“  ”“  ”“  ”“曹”，经过字形演变，现代字形中的构件“  ”是“棘”的变形构件，已经不能直接体现“棘”的字理。换言之，“曹”字的字形和字理并不对应，因此将“曹”拆分为构件“”和“曰”。按照依形不拆的原则，不对“”进行拆分，将其定为理据丧失的记号非字构件。“曰”甲骨文字形为“  ”。《说文·曰部》：“曰，词也。从口，乙声。亦象口气出也。”[13]100以此为据，将“曰”拆分为两个基础构件“口”和“一”。

为方便对比，现对“”字进行拆分。“”完全可以按照“”形体和字理进行拆分，根据《说文·曰部》“从棘从曰”的分析，可将其拆分为直接构件“棘”和“曰”。“棘”甲骨文和小篆字形相似，罗振玉据其二东并列的字形，认为其字义为双方均等，故“棘”又可拆分两个相同直接构件“朿”。以此类推，拆分“朿”得到基础构件“日”“木”。最终可得到分别拆分“曹”“”的拆分结果（见图3）。

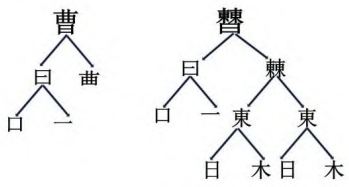

图 3　“曹”“”构件拆分对比图

四、结语

综上，制定科学的汉字拆分原则是汉字拆分工作的首要任务和良好开端。本文的构件拆分原则及流程具有四个突出特点：①普适性，此拆分原则以UNICODE九万多汉字为研究对象，其中容纳CJKA-CJKI多个扩展字符集。相较于前贤的拆分原则，此原则的适用范围更广。②理据性，依理拆分作为汉字拆分的首要原则，尊重汉字的构形理据，充分对比现代汉字的各时期字形、字理，使汉字拆分做到有理可依、有理必依。③客观性，严格规定汉字拆分的对象与下限。对无理据或理据与现代字形不对应的汉字，采用严格的依形拆分原则，从而避免因依形拆分的主观性造成的汉字拆分混乱现象。④高效性，此拆分原则和流程以字书字料库为依托产生，采用计算机与人工相结合的方式展开汉字拆分实践。前期计算机自动拆分，极大地提高了汉字构件拆分的效率，后期人工校对修改部分误差，充分保证汉字构件拆分的准确性。汉字构件拆分仍然是需要我们深入研究的长期课题，本文仅就UNICODE全汉字构件拆分原则与流程的制定提出了一些想法，不妥之处，还望方家批评指正。

参考文献：

[1] 晓东. 现代汉字部件分析的规范化[J]. 语言文字应用, 1995(3): 56-59.

[2] 费锦昌. 现代汉字部件探究[J]. 语言文字应用, 1996(2): 20-27.

[3] 韩秀娟. 现代汉字部件规范和 HSK 汉字等级大纲部件的属性调查[D]. 北京: 北京语言文化大学, 2003.

[4] 柳建钰, 王晓旭. 基于字料库的通用规范汉字构形属性调查研究[J]. 渤海大学学报(哲学社会科学版), 2019, 41(5): 104-111.

[5] 国家语言文字工作委员会. 信息处理用 13000.1 字符集汉字部件规范(GF3001-1997)[S]. 北京: 语文出版社, 1998: 1-5.

[6] 柳建钰. 字书字料库的理论、应用与实践[M]. 北京: 中华书局, 2021: 203-204.

[7] 王宁. 汉字学概要[M]. 北京: 北京师范大学出版社, 2001: 6.

[8] 傅永和. 汉字的部件[J]. 语文建设, 1991(12): 3-6.

[9] 李运富. 汉字学新论[M]. 北京: 北京师范大学出版社, 2012.

[10] 王宁. 汉字构形学导论[M]. 北京: 商务印书馆, 2015.

[11] 王宁. 汉字构形理据与现代汉字部件拆分[J]. 语文建设, 1997(3): 4-9.

[12] 李大遂. 形声字声符表义问题的探索[J]. 语文建设, 1990(6): 19-26.

[13] 许慎. 说文解字[M]. 北京: 中华书局, 1963.

[14] 苏培成. 汉字的部件拆分[J]. 语文建设, 1997(3): 10-13.

[15] 苏培成. 现代汉字学纲要[M]. 北京: 商务印书馆, 2014.

[16] 顾野王. 大广益会玉篇[M]. 北京: 中华书局, 2019: 110.

[17] 尹黎云. 汉字字源系统研究[M]. 北京: 中国人民大学出版社, 1998: 203.

(责任编辑：叶　禾)