

标题：叶典第三次会议——有关新两分规则问题的收尾

日期/时间：2021年4月29日 19:30~23:00

与会人员：白易、李佳宸、沈天珩、林成世、姜兆勤、葛炳仑

(排名不分先后，按姓氏笔画排序)

撰稿人：沈天珩

页数：5

目录

- 1.会前讨论
- 2.迭代原则的细化
- 3.包围结构再议
- 4.“屎女”结构的容许种类
- 5.常用字构成的不常用部件问题
- 6.字元少交叉原则
- 7.会后讨论

1.会前讨论

在第二次会议中，与会人员一致通过了将独体字和非独体字的拆分方法分别规定的方案；但对于独体字的认定却是一个潜在的问题。纵然可以通过扩展“不可拆部件表”来定义一些常用的部件是否为独体，但对于一些生僻部件是否应该打断却难以判断。**林**提出了“亘”不是独体字，而“互”却是独体字，其存在相似的字形结构，应如何解释这种分歧的问题。**沈**表示可以通过笔画是否接触来进行判断，当且仅当不在“不可拆部件表”且部件间笔画无接触时视为非独体字，如“亘”中“一、日、一”三个部件分别没有接触且不在表中，则不是独体字；“互”中“一、互、一”全部接触，则是独体字。

此外，独体字的认定还可通过相似字形结构进行类推。**白**首先提出是否应该将“至”拆为“至，一”，理由是如此拆分可以无残留使用。但**沈**当即表示反对。一来，“至”下半部分的“丰”笔画相连，为独体部件，不可破坏；二来，如第一次会议的会前讨论中所述，“藁”字拆作“草，八”同样能够无残留使用，但显然不破坏独体结构的规则要优先于无残留使用的存在。然而，这两条反对意见存在一个大前提，即“至”字不是独体字。该字和“至”的字形结构相似，由于“至”不是独体字，因此“至”也不是独体字，因此该前提条件成立。

对于首尾取大规则和ABA特殊结构的规则的优先级顺序，在会前讨论时存在过分歧，但会议决定维持原判，后文不再作过多赘述。如此则“亘”拆作“一，旦”，而非“二，日”，后者可作为容错；“器”保留“哭，口”的拆分，“口，犬”作为容错。

独体字的拆分细则与迭代原则的进一步细化在会议中被详细讨论，见后文。

2.迭代原则的细化

前一次会议的会后讨论中，**沈**提出迭代原则中的“尾部的尾部”可能存在一定的问题，并给出了可能情况的构字式，**林**当即找出“焯”字。该字首先拆作“火，泉”，由于“泉”并非常用字，因此需要进行迭代，但若将“泉”正常拆作“台，木”，则“尾部的尾部”原则上应该是“木”，换言之“焯”应拆作“火，木”。但对迭代原则如此描述显然违反了原本的“尾部逆向取大”的原则。因此**林**提出在迭代时不应遵守正常拆分时的规则，本着取大的原则应将尾部“泉”的尾部取“呆”。

会前，**沈**提出“啻”应当如何拆分的问题。**沈**认为既然迭代时不应遵守正常拆分时的规则，那么将尾部“衞”的尾部“逆向取大”时应先碰到“亍”，从而尾部变成了取“行”还是取“亍”的问题。但**李**表示仍然应当取“缶”。对此，会议时，**林**提出在迭代时只是不考虑首尾取大的原则，但特殊包围及ABA结构的拆分仍应保留，“衞”作为特殊包围结构应当将其尾部取“缶”。

3.包围结构再议

一般情况下，由于写字均遵从自左往右、从上到下的原则，因此向下或向右的包围结构取包围部分作为首部、被包围部分作为尾部并不存在问题；向右下包围同理；

向右上包围时，如“辶”系列的字，将含“辶”的部分作为首部亦符合常识；然而，向左、向左上、向上包围却是一些较为不常见的情况。

对于向左包围的字，如“冫”字，尽管拆字不考虑笔顺，但由于直觉上第一笔一定会写包围部分最上的横画，因此沈提出仍然将包围部分作为首部，其余与会人员无异议。

对于向左上包围的字，在会前讨论的容错可能途径中由林提出，如“斗”、“头”字，存在分歧的可能性较大，因此两种取码方式必须并行，问题转变为何为主码何为容错。考虑到与其余大多数包围的情况的一致性，决定将包围部分作为首部、被包围部分作为尾部作为主码，被包围部分作为首部、包围部分作为尾部作为容错码。

而对于向上包围的字则说法不一，因为被包围部分可能完全被包围，如“凶”字；也可能只有一半被包围，如“函”字(日韩朝字形)，与被包围部分相连的横的宽度甚至比包围部分更宽。前者如同向左上包围的字一样只是何为主码何为容错的问题；但后者显然将包围部分作为首部及其反直觉。因此对于向上包围的字的拆分统一为包围部分作为尾部，同时容错将包围部分作为首部。如此则“函”字的主码为“了，凵”；容错码为“凵，水”。

同时，对于向左包围的三种可能结构“冂”、“冂”、“工”需要定名。“工”由于是常用字可直接保留。“冂”先被放弃作为部件直接取“乙”，但李给出了许多存在“冂”部件的字，证明该部件并非仅出现在包围结构的罕见部件，同时其还作为“冂”的异体部件大量存在，有保留的必要性。沈提出可以认同至“匚”，与镜像字的处理方法一致，但该提议被否定。林提出命名为“反匚”，沈表示二字名称通常是后一字作为对前一字的描述，因此应倒转为“匚反”，最终得到一致通过。同样地，“冂”也必须保留，但命名似乎较为困难，若定为“侯右上”则过长。由于在常用非独体字中几乎找不到这个部件，林提出命名为“丑外”，虽较牵强但与会人员暂时都想不到更好的例子，只得暂时通过。与此同时，“冂”则最终被命名为“予上”，这是出于“令”在日韩朝字形中写作“令”，字形不一致必然会导致命名不统一的情况；“令下”则被命名为“节下”。

4.“屎女”结构的容许种类

新两分将“屮”拆作“屎，女”是出于字形和字理的平衡性的无奈之举，同时亦可对首尾取大原则进行简化，省去过多繁杂的限制。但这同时也会造成许多极其不直观的拆分。典型的如“廂”拆作“床，目”、“驢”拆作“驴，羽”等。这却引来了一个问题，什么结构允许按屎女规则进行拆分，而什么结构不允许，例如“匚”和“匚”是否应该取“匚”。与会人员一致反对该类拆分，林提出需要的是一条清晰的描述来排除将“匚”和“匚”的首部取出“匚”的可能性。沈提出向哪个方向包围就只能向哪个方向拆，如“匚”系列为左右方向的包围，从而内部的部分部件若欲与外部结合当且仅当内部结构为左右结构，林给出的“𠂇”字即符合这一描述，并且拆作“匚，酉”同时符合字理。向右下包围则既可以上下又可以左右，因此“屮”、“廂”等结构的拆分均符合要求。但对于向右上包围的情况，若按右下类推，则同样既可以上下又可以左右。左右并不违反直觉，例如

“迺”、“趙”等字的字理本身也是“迈，十”、“趙，卩”；但上下却极其反直觉，无论包围部分与被包围部分的上半部分结合或下半部分结合。典型例子如“遼”字，需要一条描述同时禁止将其拆分为“达，言”和“這，大”。这是因为“达，言”的拆分并不符合包围的方向，正如“麻”也不会拆成“床，亻”一样；而“這，大”又不符合被包围部分的正常拆分方向，无论如何“奢”字都应当“大”在前而“言”在后。白提出存在的嵌套包围的情况总数不过也就81种，并且有很多种是不存在例字的，可以列举一个表来详细规定。与此同时，将“摑”拆成“扌，各”的拆法也一并被排除。会前讨论中沈给出的理由是，通过检索数据库发现符合该结构的字如此拆分无一符合字理，并没有保留的必要性；此次表格整理后白给出的理由是，“扌”和“门”之间是左右结构，左右结构只能和左右结构的包围情况结合，如“漚”可拆作“汇，女”；但“门”和“各”之间的包围是上下结构的包围，因此不符合。总而言之，“摑”结构的“扌，各”拆分被彻底排除。

对于“兔”中“兔”和“丶”为何种结构并无最终定论，但无论是向右上包围或向右包围，均不影响“婉”、“銳”分别拆作“婉，丶”、“銳，厶”。

5.常用字构成的不常用部件问题

“土”和“士”均为常用字，但一般作为在左的偏旁时均为土字旁，士字旁的字极少存在，如“壻”、“埜”等；甚至在如此少的情况下仍然存在对立，如“壻”从士而“壻”从土。显然“土”和“士”无法完全认同，对于该类字给出的最终处理方法是将“士”作为主码而将“土”作为容错，主码照顾到字原本应有的面貌，容错码则防止用户查字是不知道是士字旁而造成搜不到结果。“日”与“曰”等字同理。

6.字元少交叉原则

由于独体字需要一套与非独体字不同的拆分方案，为确保主码的唯一性，必然需要更多的约束。以“車”字为例，存在“亘，丨”、“二，申”、“十，旦”、“丰，口”、“中，三”、“丰，日”等许多种无残留使用的拆法。如果仅考虑首部正向取大原则从而以首部笔画最多作为主码，显然对于许多字都有可能存在多种“主码”，并不能确保唯一性。沈提出将原先两分手册中已经存在的“字元少交叉”原则应用于此，林进而提出数两个字元之间接触点的个数，取其最少者为主码。会议中有人提出可能该原则亦不能完全确保唯一性，接触点也有可能存在同时最少的情况，沈提出在这种可能但不会多的情况下继续使用首部正向取大原则进行选择，更进一步地接近唯一拆分。此时葛找出了“卅”字，若拆作“土，丨”或“卅，一”均满足无残留使用、存在两个接触点、首部均为三笔。此时李提出区分“相接”与“相交”的概念，当上述情况发生时，相接数多的情况优先于相交数多的情况。此时由于“土，丨”存在一个相接点和一个相交点，而“卅，一”存在两个相接点，因此取“卅，一”为主码。

会前讨论时，林提出无残留使用与字元少交叉的优先级应该如何排放，若少交叉在前则“重”字不但没法拆作“壬，旦”，甚至连“千，二”都不行，只能取“丿，二”。沈表

示无论如何无残留使用都必须优先于字元少交叉。会议时得到一致通过，但此时沈提出是否应该将“更”拆为“二，史”。沈认为如此的拆分极其不直观，会前王荣(未与会)曾提出“一个字元不应当完全包围另一个字元的部分笔画”，能够将该拆分否定，因部件“史”完全包围了“二”的下半部分“一”。但如此势必会使得规则更为臃肿。会议时林认为如此拆分并不影响，将“屨”拆成“屎，女”的拆分亦不直观，但曾在官方交流群中进行的试验显示只要经过简单的学习都能拆对，包括但不限于“庵”拆为“庆，电”等情况。因此林认为将“更”拆为“二，史”亦可通过简单学习掌握。

7.会后讨论

会议决定在迭代时“尾部的尾部”忽略首尾取大原则，仅保留“特殊包围”与“ABA结构”的情况。ABA结构的字当AB能成可用部件(以下简称“成字”)时优先拆作“AB，A”，其次若AB不成字而BA成字则拆作“A，BA”，再其次若AA成字则拆作“AA，B”，若均不成字最后拆作“A，B”。该优先级顺序与会人员并无异议，然而拆成“AB，A”和“A，BA”的情况究竟应该算作使用了首尾取大的规则还是使用了ABA结构特殊规则的下位原则，却会影响对于迭代时取码的判断。沈给出了“𣎵”字作为例子，若上述情况算作首尾取大，则迭代时被忽略，因此其下半部分“木、古、木”应拆为“林，古”，从而尾部的尾部应为“古”；若算作ABA结构特殊规则的下位原则，则迭代时仍应考虑，下半部分当拆为“枯，木”，从而尾部的尾部应为“木”。姜先表示拆为“木”，但当林表示拆为“古”后姜同意其想法亦改作“古”。此后李亦表示应当取“古”，从此ABA结构在拆成“AB，A”和“A，BA”的情况均应作为首尾取大原则而非ABA特殊规则。

(End of document)