

标题：叶典第二次会议——有关新两分的进一步讨论

日期/时间：2021年4月1日 19:30~22:30

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.TD-3A40的拆法
- 3.ABA结构再讨论
- 4.容错在部分情况下的必要性
- 5.独体字的拆分规则
- 6.相似部件的认同
- 7.特殊常用部件的定名
- 8.其他
- 9.总结
- 10.会后讨论

1.会前讨论

前一次会议文件YEDN001公开发布后，引来用户们对部分存在两种拆法的字的“主码”与“容错码”设置的激烈讨论。有人认为将“屨”拆作“屎，女”太过于反人类，而同时将“鸿”拆作“汉，鸟”或“汙，鸡”都在一般人的认知范围之内，提出“合体估值”的概念，将得到的两种拆分赋予一个“主码与容错码的差距分”，以此“模拟常人”。但两分的规则不明晰是一直以来被用户们所抱怨的一个问题，此数次会议主旨便是确定一个明晰的规则，对于比常用字字数多得多的生僻字而言考虑字音字义并不合适，同时由于“屨”字的存在，此字按字理应拆作“麻，女”而非“广，夔”，但其结构却和“屨”字完全相同，均为左上至右下的半包围结构且被包围部分为上下结构，如此双标显然不是我们想要的结果。随后白又提出引入部首概念。如此虽能通过将“屨”、“屨”的部首定为“尸”、“麻”而避免上述“反人类拆分”，但一来部首概念亦是一个没有明晰的定义的概念，特别是“生僻字的部首”如同“生僻字的笔顺”与“生僻字的字理”一样并不应考虑，而如若按Unicode所定义的部首来则门槛过高，不适合一般用户使用；二来如“鸿”、“窓”、“巛”等字仍然无法一眼直观地看出部首。林提出部分字不应有“容错码”，可以设置两个“主码”；而沈表示即使是两个主码，在网页展示或输入法码表排序时也总会有先后顺序，且无论是按林所定义的双主码的情况或是真正的容“错”，沈认为一个字一般情况下不会存在三种“合适的拆法”，无论将名称定为“容错码”或是“副码”等，其本质都是一个字的“第二种拆分”，故反对该观点。白通过数据库检索找到一些可能存在三种拆法的字，但由于破坏了独体结构或是拆分出的字为非常用字，绝大部分均被淘汰，最后只剩下“淋”一字，该字被留至会议时所讨论。

会前，沈还提出由于“喬”非独体字，按首部正向取大原则是否应将“箐”拆为“笑，冂”的问题。李认为拆为“竹，喬”可以将整个字的字形全部使用，而“笑，冂”则少用了中间的“口”，因此推荐使用“竹，喬”作为主码，将“无残留使用”的规则置于“首部正向取大”之上。沈表示同意该观点，因为之前的两分手册中也提到了这条规则；但由于前一次会议决定独体字表中的独体字以不拆分为主码，而手册中该规则下举的例子是独体字“重”，因此沈未在上次会议提及该问题。但对于输入法码表中将“重”拆为“千，里”的同时，“熏”并未被拆为“千，黑”，因此此规则在原有的两分中并不统一。基于“箐”字的发现，决定统一该规则，将“无残留使用”的规则置于“首部正向取大”之上。

2.TD-3A40的拆法

前一次会议将ABA结构的字作为特殊拆分单列之后，将其视为一种特殊的“包围形结构”，如同“衣”、“行”一般，“褻”、“街”等字的首部应当为包围部分的“衣”、“行”；但ABA结构取A、B分别作为首尾部的前提是AB本身并不能作为两分部件使用，如“器”取“哭，口”而非“口，犬”。而TD-3A40“衞”字(目前非Unicode)却产生了问题，虽非ABA结构，但若按该前提限制类推到特殊包围形结构则会得到该字的拆分为“徐，亍”。会议决定该限制不得类推，该字的主码仍应取“行，余”，而“徐，亍”可作为容错。

3.ABA结构再讨论

由前一次会议的决定，ABA结构的字当且仅当AB或BA均无法作为两分字元，按常规的首尾取大原则将拆为“A，A”时，将其特殊规定为“A，B”并进行迭代。但当两个A部件的写法并不完全相同互为变体时，是否将其认同到ABA存在一定的分歧。典型例子如“狷”字，按原两分规则将“犴”与“犬”认同为相同部件，从而拆作“犴，讠”。**林**首先提出只需“字义一样就行”。这却引来了一个问题，原则上新两分并不应考虑“字义”，且“讠”与“水”、“扌”与“手”等常见的变体是否都应认同也成了潜在的问题。如此并不直观，而存在如此结构的字数并不多，为了规则的简洁性，决定将此类部件的认同关系全部删除。“莽”等字中的“艹”与“升”则面临着相同问题，会议投票决定将该认同关系亦废除。

而对于“鬲”一系列的字，由于带有非常规笔画的镜像字已经被完全认同至原字，因此其拆分仍应为“邑，共”。

4.容错在部分情况下的必要性

对于会前通过检索数据库得到可能有三种合法拆分的“淋”字，**沈**提出只保留纯字形拆分“淋，彡”和字理容错“彡，彬”，而废除不属于该两者的“沐，杉”，旨在解决该极少数存在三种拆法的情况。**林**提出对于部分独体字的拆分可能比较混乱而出现第三种拆法，便引出了之后独体字应当如何拆分的问题。但由于原问题并不存在合适的例子便被搁置，若之后在着手拆分时遇到相关问题将再另外讨论。

“箐”字由于无残留使用的规则被拆作“竹，喬”后，按首部正向取大原则的容错拆分“笑，冂”的去留便成了问题。**李**提出“冂”是否应该作为两分部件，**沈**和**姜**表示即使不作为两分部件拆为“笑，口”亦不影响该类问题的讨论。讨论后统一决定，由于“笑，冂”既不直观亦不符合字理，不适合作为容错码出现。

“藁”、“藁”等字的情况则比较特殊，其无论如何拆分均无法无残留使用，而按首部正向取大原则则应分别被拆为“芷，木”、“苗，木”。**沈**建议将字理拆分“艹，木”作为容错，得到一致通过。“窠”等字属同种情况，主码“穴，木”，容错码“宀，木”。

5.独体字的拆分规则

虽前一次会议决定表内独体字以不拆为主码且独体字表将会得到一系列扩充，但一来这些独体字需要拆分作为容错码，二来扩展区甚至Unicode外的字显然不拆是不合适的，因此独体字也需要一套拆分规则。

林提出“甬”字的尾部使用“用”不合适，因竖出头，若视为“用”则有打断笔画的现象。**白**提出打断笔画会导致自动化出现一系列问题，会议一致决定无论是否独体字，拆分时笔画不得打断。而对于笔画是否能够交叉，由投票通过允许交叉的方案，因此若“重”字拆分则为“千，二”。“甬”字则拆作“聿，月”。

对于非独体字而言，则仍维持原判，一般不得将其中的独体部件破坏。此时笔画

显然不得交叉。但如“庸”字等特殊情况，将其拆分为“广，甬”则出现非常用字或常用部件“甬”，此时“庸”字的拆分中不得不破坏其中含有的独体部件“甬”，根据迭代原则，则拆为“广，月”。

会议决定，日后将独体字表扩充时，将其改名为“不可拆部件表”，并附于新两分手册之后。

6.相似部件的认同

会议决定，“西”、“𠂇”与“𠂈”、“四”与“𠂉”均需认同，因其在日常生活中常常容易混淆，甚至“赛”的二简字是以“西”为声旁的字但写作“𠂈”、喃字“𠂉”是以“四”为形旁的字但写作“𠂉”。因此“董”应拆作“茜，土”。

对于“彦”字的包围部分，会议决定不将其认同至“产”，因“彦”字是字源分离原则的产物，本身并不符合中国大陆新字形的标准。因此按字形将其拆分改为“文，彡”，同时亦符合字理。

会议还决定，“𠂊”部件不认同至“𠂋”，因此“嘉”字应拆作“吉，加”而非“吉，茄”。

对于喃字中常出现的部件“𠂌”，会议决定将其认同至“𠂍”。一来越南本身并无绝对标准，二来在早期扩展E区的码表中带有该部件的喃字的字形均被错误地做成“𠂍”，因此将该二部件视为存在关联并认同。但与之相似的日本简化字“𠂎”的下半部分则不认同至“𠂏”，拆作“登上，儿”。

日本简化字“𠂐”原则上不应认同至“𠂑”，因其除了字形相似之外毫无关联，因此“𠂒”应拆作“言，儿”；但考虑到需要查找该字的用户比较容易将其误认为“𠂑”的异体字或讹字，因此也需增加“言，𠂑”作为容错。

会议决定，日后应整理“可认同部件表”，与“不可拆部件表”一同附于手册之后。

7.特殊常用部件的定名

汉字中存在部分常用部件被收录至扩展区甚至未被Unicode收录，这些特殊部件必须在能够作为两分字元的同时又不影响输入和显示的便利性，因此需要定一个双字名。但原两分中的定名规则不统一，如“丰”为“串义”、“𠂒”为“杨音”、“𠂓”为“春上”，后字的描述较为混乱。会议决定，将该类描述统一改为其在前字中的相对位置，废除“音”、“义”等定义较为模糊的描述。对于“𠂔”等部件则使用“外”作为其描述。此类部件的命名应同样列表附于手册之后，避免如将“𠂒”定为“杨右”或“扬右”甚至“场右”一类的分歧情况。其在输入法中的读音则根据双字名的前字直接确定。

8.其他

待两分字元所需取用的字大致确定后，做拆分工作时将使用特制字体帮助直接有

效排除非两分字元；在两分搜索页面则暂定为通过JavaScript实时判断，若输入框内出现非法字元则立刻出现相关提示。

9.总结

此次会议主要得出以下规则的变更或新结论：

- 1.将无残留使用原则优先于首部正向取大原则，但无论如何不得将笔画截断；
- 2.ABA结构的特殊前提不应类推至“衣”、“行”类特殊包围结构，但可作为容错；
- 3.不将存在一定字形差异的变体作为ABA中两个A的认同；
- 4.独体字的拆分与非独体字的拆分规则分列，并建立“不可拆部件表”；
- 5.同时还需建立“可认同部件表”及“特殊常用部件命名表”。

10.会后讨论

有关迭代问题，在两次会议中均未作详细说明。对于“尾部的尾部”，原则上应只考虑“尾部逆向取大”而忽略“首部正向取大”，如“焯”字，首先分为“火，泉”，由于“泉”不是常用字便需继续拆分，但“泉”本身的拆分为“台，木”，“厶，呆”为容错，迭代时若按“泉”本身拆分的尾部代入得到“火，木”，并不符合原“尾部逆向取大”的原则，此为迭代原则用于详细描述首尾取大的一个漏洞。因此需规定在取“尾部的尾部”时不能将尾部本身的拆分所得到的尾部代入，而是直接重新“尾部逆向取大”。

对于ABA结构中AB或BA均不能作为两分字元使用的情况，原则上使用“A，B”，但若AA能构成两分字元，则亦作为“衣”、“行”类特殊包围结构处理，如“棼”拆作“林，爻”而非“木，爻”。此类情况有限，亦将在两分手册中单列进行说明。

(End of document)