

L2/10-352r

2010-09-24

Title: Response to Irish NB comments N3908
Source: Van Anderson
Re: Duployan proposal N3895r
Action: For consideration by JTC1/SC2/WG2 and UTC

1. Introduction. Michael Everson's assistance, logistic support, and review of the various stages of the Duployan script proposal over the past two years have been critical to the successful completion of this project. His comments and concerns have invariably led to improvements in the proposal. He has been part of the mailing list for discussion of the Duployan proposal since its inception, and has been one of the most active and engaged members thereof.

The issues raised in document N3908 have been raised by Mr. Everson before. This previous spring, the issues he raised were considered, shared with all interested parties, and after considerable thought, debate, and weighing of alternatives, the consensus of the interested parties – including the Unicode Technical Committee – was that the current allocation order was preferable to one based on the collation specification.

In short, the current allocation order is not a mistake, and it is not an accident. It is the result of a conscious decision that weighed the pros and cons of each of the possible allocation orders. The fact is, Mr. Everson's argument did not carry the day when it was shared with all of the interested parties. It is unfortunate that he was unable to accept that there were perspectives on the issue that outweighed his concerns, and I only ask that the WG2 consider accepting the current allocation order as one of the acceptable alternatives. If the WG2 agrees that there is nothing intrinsically wrong with the current allocation, then I ask that it merely defer to the judgment of those of us who have a vested interest in the success of this proposal.

2. Proposed changes to the collation specification. Mr. Everson proposed several changes (section 7 of N3908) to the collation specification, regarding ordering within shape classes. Given the nature of the collation

algorithm, I have decided that his proposed changes are an improvement, and have updated the proposal document to reflect those changes and a few others of a similar nature.

3. The nature of the Duployan collation order. The Duployan collation, which Mr. Everson suggests should be the basis for allocation order, would not exist if encoding in Unicode did not require the specification of sort order. Duployan *does not have* a native collation, and the collation algorithm is based on the barest of premisses in the source documents. Essentially, I took something that looked like it might be an order of Duployan letters, and tried to make something functional, elegant, and logical out of it. Let me be perfectly clear, however: the Duployan sort order was invented by me, to sort Duployan characters, and could take many different forms and still be just as faithful to the source documents – including a simple binary sort of the current allocation.

4. Duployan code chart organization. Michael is correct that a simple binary sort of the current allocation will result in a different result than the proposed collation algorithm. This is due to the organizing principles of the current allocation, which result in a

“[...] current allocation [that] groups letters by usage, and orders them to invite comparison and contrast, shedding light on the proper usage of the characters. A collation-based order hides the relationships between a basic letter and its variants.” – UTC consensus as expressed by Ken Whistler.

The Duployan allocation is predicated on the reality that it is not a single self-contained script, but a collection of six related scripts that were derived from a proto-form that is no longer in use as a complete script. These six scripts are the French Duployéan, Romanian Stenographie, Chinook script, Pernin shorthand, Perrault shorthand, and Sloan-Duployan. Of these, French Duployéan is the modern descendent of the original Duployéan shorthand. Chinook, Pernin, and Sloan-Duployan are independent adaptations that started with the original core Duployan, but relied on novel adaptations and

new letters to represent Northwest native languages (Chinook) and English (Pernin and Sloan-Duployan). Perrault is heavily based on Pernin, and the Romanian Stenographie, as the only recently created orthography, seems to have been influenced by all five of the previous Duployan shorthands. Of these six, only French Duployéan has an active community still creating documents, and Chinook has considerable cultural and scholarly work associated with it.

This background informs the entire allocation order because the allocation groups characters by which of these six “orthographies” it is found in. The first two columns of the allocation are the original, basic letters of the original Duployéan. Almost all of these characters are shared among each of the six orthographies. The next three columns (U+1BC20..U+1BC4F) of the allocation complete French Duployéan. Many of these characters are used by Romanian, Pernin, Perrault, and Sloan. The next two columns (U+1BC50..U+1BC6F) are intended to complete the needs of the Chinook writing. Because the form of the Chinook characters are similar to them, the few remaining Romanian characters are included in these two columns as well. The last three columns are left to the English Duployan scripts, Pernin, Perrault, and Sloan-Duployan. Since these orthographies were in use at the same time, and among speakers of the same language, they grown to share much of their character repertoires.

This brings us to the three advantages that the current allocation order provides:

- 1) The two orthographies with current use – one scholarly, one by current writers – are completed first. This is in direct response to P&P, D.2.1: Block assignment starting on half-row boundary, which directs “For blocks slightly larger than 128 positions the highest frequency characters should all be allocated within the first 128 positions”. Mr. Everson's proposed allocation places characters necessary to all six orthographies past this 128 code point position (U+1BC7F).
- 2) The current allocation assists amateur developers of keyboards and other input devices in determining the correct characters for each of the six orthographies, especially the two most frequently used – French Duployéan

and Chinook. By helping to orient these people and reduce errors in determining which characters are used by which orthography, the current allocation helps to ensure that conformant documents are produced by end-users. Mr. Everson's proposed allocation obfuscates the classes of character variants, and jumbles the characters necessary to each orthography, inviting mistaken identities, resulting in non-conformant documents.

3) The current allocation is based on inherent properties of characters from the source documents. Mr. Everson's proposed allocation is based on an invented and imposed property with little basis in the source materials.

5. Conclusion. I have been unable to find any basis for imposing a collation based allocation order within the P&P document, or any other suggestions beyond D.2.1. This leaves two options that would continue to reflect the source documents for the Duployan script:

1 – If the WG2 agrees that it does not have an unstated policy of aligning allocation order with collation order, then the current proposal is complete, vetted, and meets all of the technical specifications of Unicode and ISO-10646. Thus proposal document N3895r should be approved as is.

2 – If the WG2 concludes that it does have an unstated policy of aligning allocation order with collation order, the proposal document N3895r should be approved with the collation specification abandoned in favor of a simple binary sort of the code chart as is.

I reiterate that imposition of the complex collation specification onto the allocation order is inappropriate and cannot be justified by the source documents on which the Duployan proposal is based.

-Van Anderson
2010-09-24