

**Leibniz Unicode Characters Project**

(LUCP)

Philiumm Project, Université de Paris VII

Leibniz-Archiv Hannover

Andreas Stötzner, Fontentwicklung

**Towards the encoding of new alchemical symbols**

Source: Kirk Miller, Andreas Stötzner

Number: L-2443

Date: 23-10.29.

Related: L-2432

Action: For consideration by Script AdHoc group and WG2

**A. Introduction**

**Kirk Miller** (KM) has thankfully provided comments on our proposal (L-2432) and discusses various options regarding the characters in question. In this document, the text of his reply is presented in full length and complemented with comments by **Andreas Stötzner** (AS) on behalf of the Philiumm/LUCP research group.

We kindly ask to feed this document into the ScrAHoc and WG2 document register.

**B. Technical Information**

KM posted his comments via email to AS. The full text has been copied into a text document (TextEdit on Mac) in order to create this PDF document. The whole body copy has been assigned the font *Times New Roman*, which is a default system font on the device. Most of the symbols in the text are, however, not in this font, so the system runs a fallback operation in the case of missing characters in the given font (more about this in section 3).

For the ease of further discussion each section has been given a reference number.

## C. Comments (KM) and replies (AS)

(1)

KM: 'Tartar salt' is a ligature of 1F714 ALCHEMICAL SYMBOL FOR SALT  $\ominus$  and 1F73F  $\boxminus$  ALCHEMICAL SYMBOL FOR TARTAR. It should presumably have its own code point, as we do for similar combinations, e.g.  $\oint$  ALCHEMICAL SYMBOL FOR SALT OF COPPER ANTIMONIATE from  $\oint$  ALCHEMICAL SYMBOL FOR COPPER ANTIMONIATE, and  $\eth$  ALCHEMICAL SYMBOL FOR SALT OF ANTIMONY from  $\oplus$  2641 EARTH (symbol for antimony).

Moon-Jupiter = silver-tin is also a ligature. In addition, there's a Moon-Mercury ligature (lunar crescent with the horns of Mercury attached) that's found in the [Newton's corpus](#) (listed under both Silver/Moon and Mercury), but for whatever reason that wasn't requested with the other Newtonian symbols.

**AS: These characters should get encoded as proposed. As has been mentioned, there is a lot of precedence in the Alchemical block which justifies the encoding.**

(2)

KM: The last two requested symbols, hora-2 and retort-2, are allographs. There are already a number of allographs in the alchemical block, and I never understood what the point was. In the figs btm p.5 and on p.7 we see a number of other allographs that might be equally requested. Btm p.8 we see a dozen more just for hora, potentially taking us up to hora-12 or so. My question is whether the multiple allographs already in the block were a mistake, or if this is something we wish to continue. Hora-2, as 'white hourglass', is also found in the [Newton Corpus](#) (btm, under 'editorial marks'), and wasn't requested separately.

**AS: It is true that the two characters are (at least somehow) allographs of already encoded characters. This is a known general issue with historic alchemical symbols which has not yet been addressed by Unicode yet, in a scientific and practical way. However, we request these encodings because**

a) "Hora-2" (the straight shape version) is of all 'hora' allographs the most common and frequently occurring. We question the fact that a fancy calligraphic version of 'hora' has been chosen instead for 1F76E. We also question the decision of changing the reference glyph in the new code chart version to another fancy variant form and leaving the well-known common glyph variant aside again. We suggest to either change the reference glyph of 1F76E to the simple shape we proposed, or to take into account that the current HORA in the USC will not be acceptable for editors who need to encode the simple straight hora symbol, for which use case the char. HORA-2 is intended. We see this suggestion in line with other, similar cases in the 1F700 block where two or three allographs have been given separate codepoints on ground of considerable glyph differences.

b) We suggest the encoding of RETORT-2 for the same reasons: strong glyphic difference and compliance with established encoding practice.

(3)

KM: The square for foil is productive, being used for any malleable metal. One possibility would be U+20DE COMBINING ENCLOSING SQUARE, for ☐ (silver foil) and ☐ (gold foil), and so on with ☐ (a variant of silver foil), ☐ (tin foil or old-style electrum foil), ☐ (old-style tin foil), ☐ (copper foil), ☐ (iron foil), and ☐ (old-style variant for gold foil). I haven't seen it with Saturn for lead foil. [...]

If you take a look at [folio 6 of Marcianus](#), you'll see multiple instances of an allograph of this convention, where the square is instead attached to the bottom of the planetary symbol to indicate a foil of the planetary metal, e.g. attached to the old symbol U+1F71A ☿ for sun/gold for gold leaf/foil (3rd symbol from top, left column), on ☾ Moon for silver foil (no. 11), on ♀ Venus for copper foil (no. 17), on an old variant of ♂ Mars for iron foil (no. 24, 3rd from the btm), and on Mercury for tin foil (no. 17 in the 2nd column), using an old identification of Mercury rather than Jupiter as the governing planet of tin.

[Note that in this ms the symbols for Mercury, Venus, Jupiter and Saturn haven't yet been Christianized by adding crosses to them -- additional allographs that we don't distinguish in Unicode.]

On the other hand, we might want to handle these with distinct characters as we do for with the small circle (not the Latin letter 'o'!) added to the planetary metal to indicate an ore of that metal, namely in:

♂ ALCHEMICAL SYMBOL FOR IRON ORE,  
♀ ALCHEMICAL SYMBOL FOR COPPER ORE,  
♂ ALCHEMICAL SYMBOL FOR TIN ORE,  
♂ ALCHEMICAL SYMBOL FOR LEAD ORE,  
♂ ALCHEMICAL SYMBOL FOR ANTIMONY ORE,  
♂ ALCHEMICAL SYMBOL FOR BISMUTH ORE.

There's also a silver ore (crescent + small circle) that's not yet in Unicode.

Note that if we assign separate code points, the planetary metals can also take slashes across them to indicate shavings or filings of that metal, and a combining Greek kappa for the calcinated metals. These can be seen just above and below the corresponding foil symbols in the Marcianus ms.

But, personally, if I were digitizing Marcianus, I wouldn't have a problem using ☐ and ☐ (or, to be pedantic, ☐) for his silver- and gold-leaf symbols, even though the square doesn't enclose the planetary symbols in that particular manuscript. The difference is trivial, and the combining square is the form we have in modern typeset sources.

**AS: The idea of using 20DE COMBINING ENCLOSING SQUARE to create the characters for 'foliated gold' and 'foliated silver' has its merits but any implementation of this approach in actual projects produces miserable display fails, for two reasons: a) 20DE is missing in most fonts; b) 20DE is behaving all the way differently depending on which fonts, applications or environment is used. A reliable portation of these**

characters is not to be expected under these circumstances. See the figures (end of this doc.) for some samples of how the combined sequences behave in reality when text is being copied from one source into another document or environment. The typical case is: a text containing these expressions gets composed in a font which doesn't contain some or all of the symbols needed. Then the system usually runs a 'fallback' mechanism and chooses randomly (!) some other font for the missing characters. Not only that, as can be seen in this very document and the figures, the system even picks *various* fonts for e.g 20DE, and on top it picks different fonts for a given single sequence, such as <2640 20DE> (for foliated copper). Since especially 20DE is designed differently in fonts – sometimes it is not even modelled as a combining char. –, this fallback *Roulette* produces text display fails of all kinds. The only sound solution for this problem is, as we propose, the encoding of separate characters and we sympathise with the idea of extending the set of squared planet signs to the full range of symbols which KM mentions.

(4)

KM: Also, if we're going to do hora-2 and retort-2, why not Mars-2 for the old variant of Mars seen in Marcianus? (Left column, 6th from the bottom: ring on an arrow, approx.  $\nearrow$ ). That's quite common in old sources. Or, for that matter, Mercury-2, Venus-2, Jupiter-2 and Saturn-2 for the pre-Christianized variants without the crossed stems.

**AS: These signs/variants are not within the scope of our proposal but we agree that relevant symbols in other sources also deserve separate encoding.**

(5)

KM: It seems that inverted- $\circ\circ$  for 'boiled oil' is an allograph, only contextually distinct from  $\circ\circ$  'oil'. In the fig. btm p.5, it's not clear that 'oil', 'distilled oil' and 'boiled oil' are distinct symbols, as the orientations of the drops don't seem fixed. In the figures above (the ms and its transcription), the difference is only contextual:  $\circ\circ$  di uliva ('olive  $\circ\circ$ ', that is, 'olive oil') and  $\circ\circ$  laurino ('laurinaceous  $\circ\circ$ ', that is, 'laurinaceous oil' -- or whatever the adjectival form of 'laurel' is in English!). Note that the inverted  $\circ\circ$  is used for olive oil, which is the default reading of the regular symbol  $\circ\circ$ , olive oil being by far the most common oil in the Mediterranean. So we don't have any document that makes a clear contrast. Inverted- $\circ\circ$  is also found in the [Newton corpus](#) (listed under 'other substances'), where it's labeled OIL [VARIANT-2], along with a 5-drop variant that's also just OIL [VARIANT], and the people at the Newton Chymistry Project evidently didn't feel a need for a separate character for either.

**AS: The sources we show demonstrate clearly that 'oil boiled' is NOT an allograph of 1F746. We demonstrate cases in which both characters are distinct by both graphic shape and meaning, cases in which the semantic differentiation of the two expressions is essential for the meaning of the text. According to established encoding principles that ought to be satisfying to agree to its encoding. The fact that in some other sources the distinction is not so strong or ambiguous, can not be held as a valid exclusion criterion. What 'the people of the Newton Chemistry Project feel' (or may have been feeling) is of no relevance here.**

(6)

KM: To summarize, I think only 2 and 3 (moon-jupiter and tartar salt) are clearly justified by Unicode standards, while for 4 and 5 we need to decide whether to suggest using the combining square [adequate in my opinion, as that follows modern usage] or to assign individual code points. If the latter, IMO the names should probably be 'gold foil' and 'silver foil', and so on for the rest if/when they're added, though we might end up with two 'tin foil' variants, based on Mercury and Jupiter.

The others (no. 1, 6 and 7) appear to be mere allographs.

**AS: We oppose the view that we have proposed “mere allographs” for seperate encoding. It shall be pointed out that**

**a) on grounds of established encoding practice, and**

**b) by the proof of the samples we give in L-2432**

**the encoding of all 7 characters proposed should be regarded as justifiable. We are, however, open to further suggestions in term of optimization of character names and/or a possible cooperation with other authors who consider encoding proposals of a similar nature. We welcome any constructive input which helps making the UCS's alchemical repertoire more useful.**

*next pages: figures*

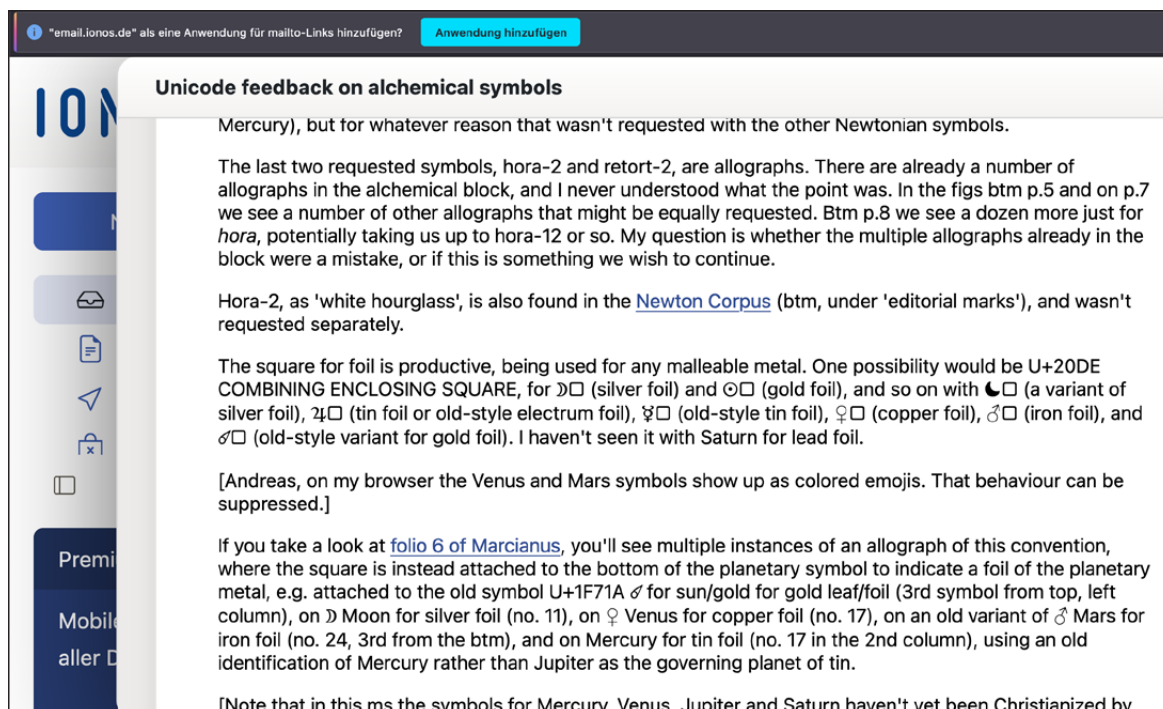


Fig. 1: A part of Kirk Miller's original explanations, viewed in a webmailer (Firefox, Mac).

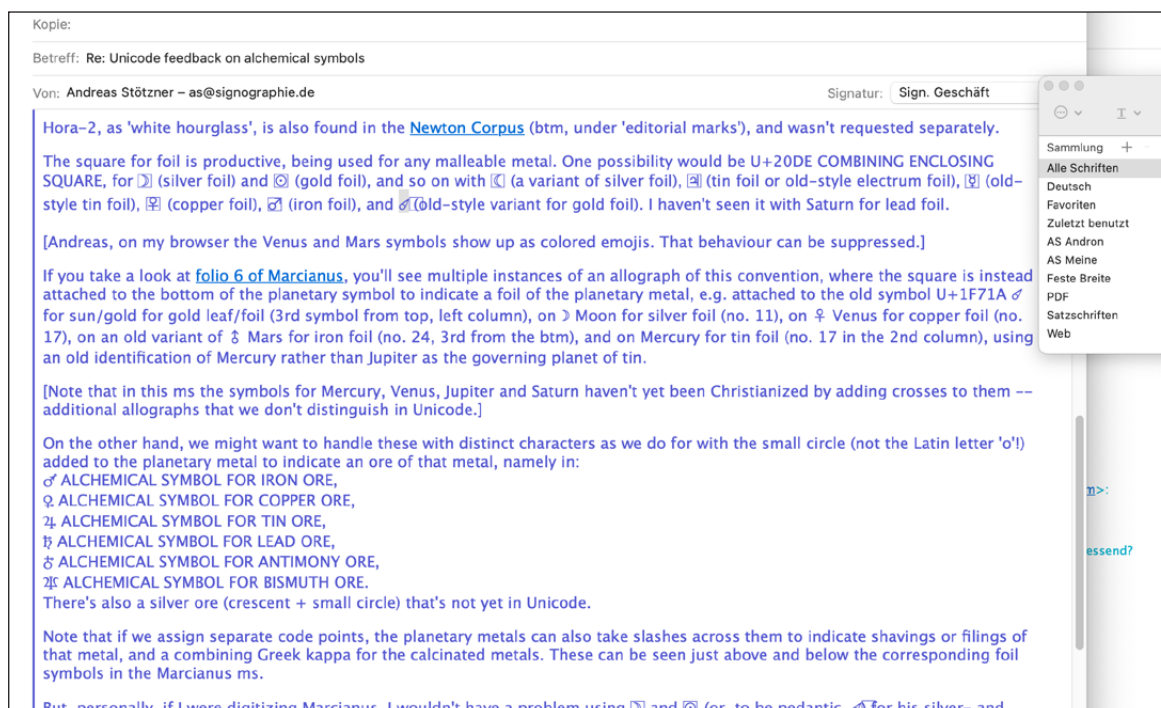


Fig. 2: A part of Kirk Miller's original explanations, viewed in Mac's Mail program. Note the different display behaviour of combined sequences.



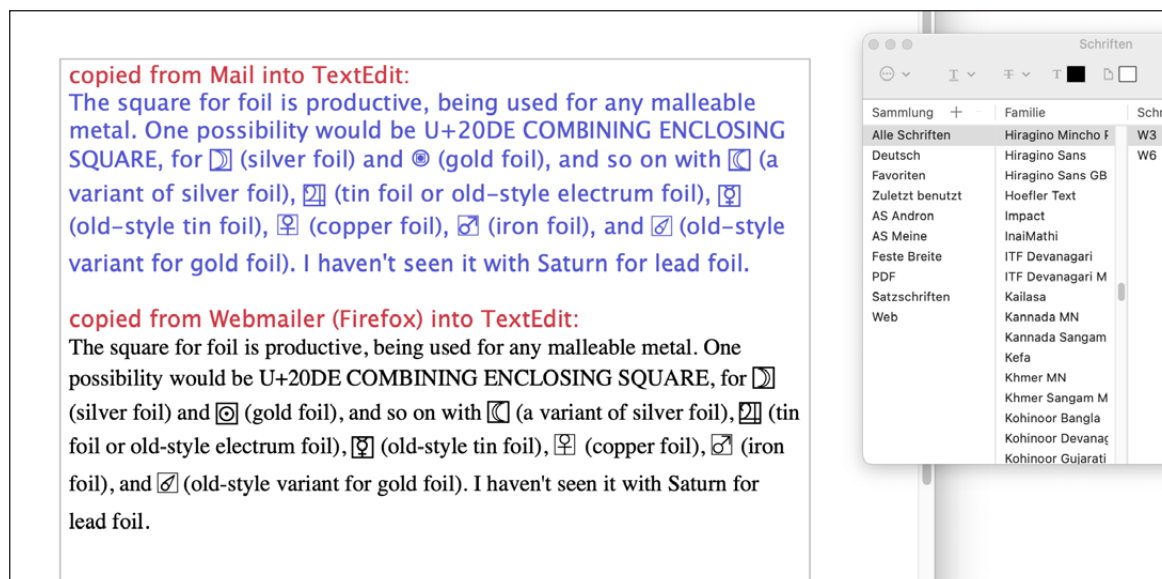


Fig. 3: Here a part of the original text has been copied into a simple text editor (TextEdit, Mac). Note the particular fail of the 'gold foil' symbol in the top section and the inconsistent display of 20DE, picked randomly from various fallback fonts. Note also the typographic insufficiency of some of the planet's glyphs clashing with the square box of 20DE.

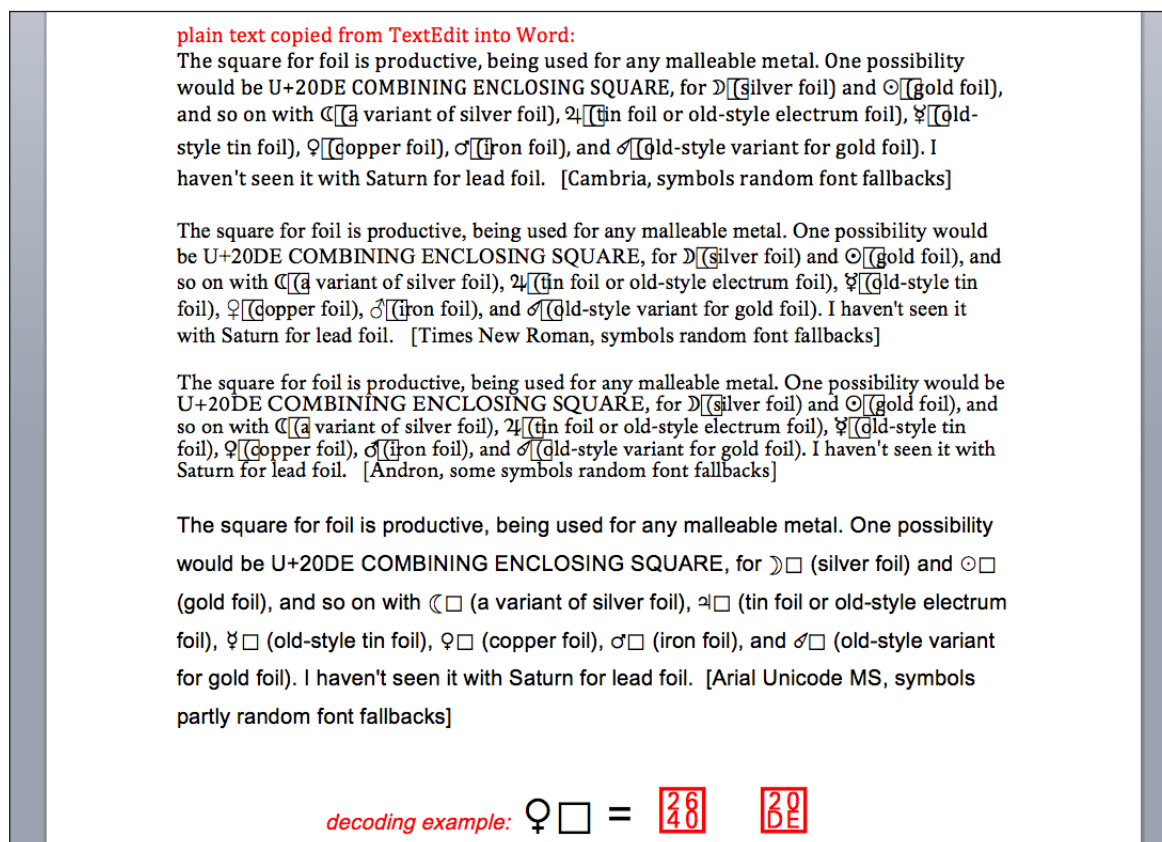


Fig. 4: A test with various fonts chosen, in MS Word. All four test bits are unusable.