

**Proposal to encode the LATIN CAPITAL LETTER AT for Koalib orthography and gender inclusive language in Portuguese and Spanish**

**Eduardo Marín Silva**

**26/12/18**

**Introduction.** This is an improved proposal, designed to address the original concerns of the letter being included in the standard. I do this by considering different encoding models to seek the middle ground between security concerns and the needs of the Koalib language community. I have also included a discussion about the use of the character on gender inclusive language on Spanish and Portuguese.

**Note on partiality.** I must address the proverbial elephant in the room, by pointing out that several pieces of feedback were based on the sheer smallness of the user community (e.g. <http://unicode.org/L2/L2004/04365-pr40-ewell.pdf>) This is a dangerous precedent, since it means that if one can neglect an orthography, because of the size of the community, then the grand majority of ancient and minority scripts would have postponed for years, for no good reason.

I am of the belief that once past a certain minimal threshold, all orthographies and scripts should be seen impartially.

I also believe that the consortium was guided more by the possible security risk, than by a disdain for the small community. Nevertheless, it is important to avoid having bad optics (doing something that seems ill intended).

**Important note.** The proposal as it is, does not propose to change the general category of the at sign to that of a letter and indeed is compatible with the at sign to keeps its current category. So it is not in contradiction with [PRI #40](#).

**EM 1. Separate pair encoding:** Encode the two letters as separate, with the correct general category.

- Pros: Parsing and interchange of Koalib text (in that particular orthography) will be made much simpler. U+0040 @ will not have to change properties.
- Issues: Having a second character that can be confused with the @ sign, would open the door for malicious attacks by spoofing email addresses (which do not have the same protection as URLs); this may have catastrophic consequences.

**EM 2. Reuse characters:** Use U+24B6 (Ⓐ) and U+24D0 (ⓐ) to act in the place of the letters

- Pros: No new characters need to be encoded. Casing relation already in place. Already recommended by SIL.
- Issues: The general category is that of a symbol (generally used as a bullet), so segmentation and spell check operations will be made more complex. The majority of text processors do not recognize case relations between non-letters (for good reasons), including the proposed characters. The glyph of these symbols, do not need to harmonize with the surrounding font, and the normative glyphs only approximate the desired glyph.

**EM 3. Single new letter with case mapping:** Only encode the uppercase letter and assign the case mapping to the existing U+0040 @.

- Pros: Only one new character added, with the correct glyph and properties. The lowercase “letter” will be accessible for the grand majority of keyboards that support ASCII anyway. Current implementations of email addresses are case insensitive, so there is no security risk.
- Issues: The general category of the COMMERCIAL AT will have to be revised from “punctuation\_other” (Po) to “symbol\_other” (So), in order to avoid breaking the assumption that punctuation characters do not have case mappings; this would spell headaches to implementations that rely on the sign to remain stable. The number of possible ways to type the same address will effectually double.

**EM 4. Single new letter without case mapping:** Encode the uppercase letter but do not assign a casing relation at all.

- Pros: Email implementations will not have to change at all. The glyph will be proper and the general category can be that of a letter. The lowercase “letter” will be accessible for the grand majority of keyboards that support ASCII anyway. The casing relation can be obtained in some ad-hoc way.
- Issues: Case folding stability would mean that once encoded, the uppercase letter will not be able to have a casing relation ever. The ad-hoc nature of the case mapping, would doom the orthography to

never be practical in digital contexts. Segmentation and spell check operations will be complicated if the COMMERCIAL AT does not change general category. Case folding of the new letter, would break the norm of all other Latin letters that expect a lowercase.

**EM 5. Single new symbol without case mapping:** The same as EM 4, but the general category will be that of a symbol.

- Pros: Same as EM 4, plus, case folding is no longer an issue.
- Issues: Same as EM 4, minus the case folding issues.

Since I have determined that the first two encoding models are inadmissible, I will only pursue the last three.

**Entry for the new letter.**

- For model 3, the entry would look like:

Ⓐ LATIN CAPITAL LETTER AT

- used for Koalib
- lowercase is @ 0040

@ COMMERCIAL AT

= at sign

- also used as a letter in Koalib
- uppercase is Ⓐ xxxx

- For model 4, the entry would look like:

Ⓐ LATIN CAPITAL LETTER AT

- used for Koalib
- this letter in particular has no formal case mapping, but it is intended to be used with @ 0040

@ COMMERCIAL AT

= at sign

- this has no formal case mapping due to security considerations, but Koalib documents may use Ⓐ xxxx as a case pair

- For model 5, the entry would look like this:

Ⓐ SYMBOL FOR LATIN CAPITAL LETTER AT

- used for Koalib
- this symbol is meant to be the uppercase counterpart to @ 0040

@ COMMERCIAL AT

= at sign

→ xxxx Ⓐ

**Gender inclusive language.** Spanish and Portuguese have two grammatical genders (feminine and masculine), but it lacks a gender neutral way of referring to substantives. So if one wants to refer to all nurses and not just male or female nurses, one has to write “*enfermeros y enfermeras*”; there is also an abbreviation that goes “*enfermeros(as)*”, however another popular (and shorter) way to write it is “*enfermer@s*”; this is due to the fact that the at sign, resembles both an “a” and an “o” at the same time, which are the letters used to indicate a different gender at the end of words. This has the added benefit of not introducing misunderstandings, when deciding which gender is in the word itself and which one is in the abbreviation.

This kind of notation is advised against by many language institutions ([Wikipedia in Spanish/Wikipedia in portuguese/Wikipedia in English](#)). However, the official suggestion is to use the male form unless gender is somehow relevant, which is ambiguous and obsolete.

This is relevant to this proposal, because if one wanted to express the same sentiment, but using all caps, the presence of this character would aid in obtaining a better looking presentation:

**ENFERMER@S vs ENFERMERⒶS**

The whole emoji submission process, is testament of the Consortium’s dedication towards better gender presentation; encoding this character, for this reason, would therefore not be unprecedented.