மூலாதாரப் பதிப்புகளைத் தயாரிப்பதில் சில பரிசோதனைகள் : செய்தனவும் செய்ய வேண்டியனவும்

முனைவர் ழான் லூய்க் செவ்வியார்
Dr Jean-Luc Chevillard (Hamburg & Paris)

https://www.tamilex.uni-hamburg.de/team/chevillard.html
https://htl.cnrs.fr/equipe/jl-chevillard/
https://orcid.org/0000-0001-7498-6587

"மயங்கா மரபின் எழுத்துமுறை காட்டி"
(தொல்காப்பியம், சிறப்புப் பாயிரம்)

Starting point: meditating on a remark by K.G. Krishnan[1984]

A possible starting point for unambiguously introducing the topic of **«The inherent ambiguity of the Tamil writing system of many historical periods»** which is what I wish to discuss in this presentation is to provide a quotation and an illustration from the 1984 ASI publication by K.G. Krishnan. The citation is as follows:

(1) *Pōttandōm* and *pōttaga* are used in lines 360-61. The former occurring in other records has been wrongly read as *pērtandōm*[1] and translated—'furnished the names'. The latter is stated to be made up of *pōga* and *taga*.[2] The correct root for both is a composite one of *pō* and *tā* (*taru*). *Pōttandōm* will mean 'we gave (i.e. appointed them)

to go with you'. *Pōttaga* is a contracted form of *pōdaruga* or *pōttaruga* (go and get). The second component in both these cases is just an auxiliary that extends the meaning of the first in accordance with the context.

[1] SII., III, p. 404, line 128, translated 'nominated'; above. XXII, p. 245, lines 48 and 50 and foot note 8. Both the reading *pērttandōm* and the correction *pēr tandōm* are untenable.
[2] Ibid., foot note 10.

(K.G. Krishnan, 1984, pp. 6-7)

The illustration is provided here as two small extracts from the photograph of plate XII, corresponding respectively to the beginning of line 360 (Figure 1) and to the end of line 361 (Figure 2), accompanied by a small extract from page 94, containing the full transcription of lines 360 & 361 (Figure 3)

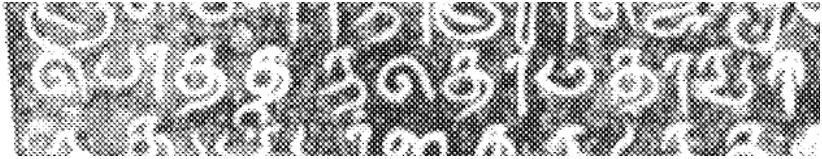

Figure 1: beginning of line 360 in plate XII

Figure 2: end of line 361 in plate XII

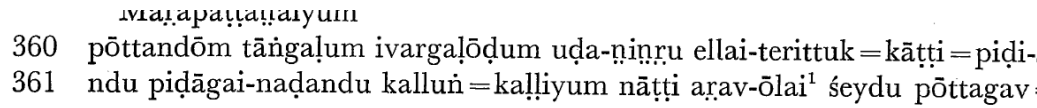

Figure 3: extract from page 94: transcription of plate XII, lines 360 and 361

Representing ambiguity. Attested reading vs. intended reading.

I shall now extract from the set of elements given in the previous section, those central elements which illustrate the point which I intend to make here, concerning the representation of inherently ambiguous written artifacts. As summarized in Table 1 below, K. G. Krishnan criticizes the

editors of SII for “both the reading pērttandōm and the correction pēr tandōm”, which he sees as untenable. His own reading is *pōttandōm*.

Attested reading (in image form)	Conflicting **intended readings**	Identification of reader
	*pērttandōm,* corrected as *pēr tandōm*	SII editors
	Pōttandōm	K.G. Krishnan

Table 1: Ambiguous text (as IMAGE)

I shall now transition

- from **(A)** the topic of the handling of ambiguous text, seen as the challenge of **finding the correct interpretation of an ambiguous written trace** (which is an **IMAGE**)
- to **(B)** the different challenge of **faithfully representing ambiguous text** (in **STRING** form).

One of the compelling reasons for making this experiment is that the challenge of performing a String-Search is much easier than the challenge of performing an image search. This will first be illustrated in Table 2 below, which exemplifies a system which I have now been using for more than twelve years and which I call the RAW format. I shall first provide the same example as in Table 1, and then continue by means of a new example, which I heard mentionned last year during a workshop in Pondicherry by Professor G. Vijayavenugopal, the well-known epigraphist. The two examples are as follows.

Possible **intended readings**	Ambiguous **attested reading** in RAW format
பேர்த்தந்தோம்	@ ப # த த ந @ த # ம
போத்தந்தோம்	
பேரருளாளன்	@ ப # ரு ள # ள ன
பொருளாளன்	

Table 2: Ambiguous text (as STRING): introducing the RAW format

Diplomatical RAW transcription of அகத்தியர் தேவாரத் திரட்டு manuscrit preserved in Paris: BnF indien 113

The example given in the previous section is of course much too short for really demonstrating to the audience the logic of the RAW system which I am presenting here. As a more elaborate example, I shall first extract a small sample from a transcription which I have made in recent years of a MS containing a subset of the *Tēvaram*, that subset being usually referred to as the அகத்தியர் தேவாரத் திரட்டு. That MS is preserved in the BnF (Bibliothèque nationale de France) and its shelf mark is “BnF indien 113”. I shall first reproduce one of its leaves (see Figure 4) and then provide a

sample RAW transcription of the first stanza in a well-known hymn, found on that leaf (on lines 5 to 7), followed by a sandhi-split version of the text in the IFP 1984 edition by T.V. Gopal Iyer, which is also available in the 2007 Digital Tevaram CD. The whole MS, as transcribed by me in RAW format is available at the following link https://gokulam.net/jlc/tEvAram/BnF_indien113_with_parallel_texts.html.

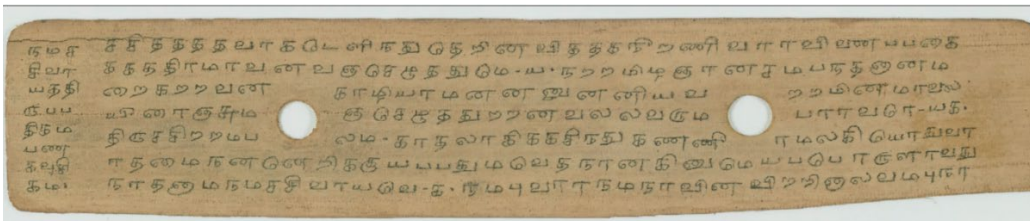

Figure 4: BnF indien 113, folio 14v (selected hymn starts on line 5 out of 7)

(2) க # த ல # கி க க சி ந து க ண ணீ # ம ல கி @ ய # து வ # # த @@ ம ந ன @ ன றி க கு ய ப ப து ம @ வ த ந # ன கி னு @ ம ய ப @ ப # ரு ள # வ து ந # த னா ம ந ம ச சி வ # ய @ வ (VERBATIM manuscript transcription in RAW format of the stanza starting on line 5 and ending on line 7)

(3) காதல் ஆகி, கசிந்து, கண்ணீர் மல்கி, // ஓதுவார்தமை நன்நெறிக்கு உய்ப்பது; // வேதம்நான்கினும் மெய்ப்பொருள் ஆவது--- // நாதன் நாமம் நமச்சிவாயவே. (Published reference text for தேவாரம் 3_49, in Gopal Iyer, T.V., 1984)

As explained in several recent publications, such as for instance Trento & Chevillard [2025: 559], the main characteristic of **the RAW format** is that it carefully tries not to remove the ambiguities inherent in the Tamil orthography which is seen in the manuscripts transcribed. That means that:

1. puḷḷi-s are not used;
2. a space is inserted after each non-breakable written symbol;
3. the symbol "@" is used for transcribing the kompu which appears in the spelling of words which nowadays would use "ொ◌" or "ோ◌";
4. the symbol "@@" is used instead of "ൌ◌";
5. the symbol "#" is used where modern spelling would distinguish between "◌ிீ", "ிஜ" and "ிஜ̇";
6. both "@" and "#" are used for transcribing syllables where -o and -ō are combined with a consonant;

7. in MSS where there is a sporadic (inconsistent) use of “ெ◌”, that symbol is transcribed as “@2”;
8. the symbol “H” in a Tamil transcription represents the position of a string hole in the MS;
9. the sequence “//” represent a change of line;
10. the sequence “////” represent a change of column or a change of leaf.

Returning now to BnF indien 113, as transcribed in (2), we can observe, if we compare it with the text provided in (3), that there is a small differences between the two, which is not simply explainable by the fact that (3) is an unambiguous sandhi-split text. The difference is seen in the presence of the particle -**உம்** at the end of the sequence “**ந ன @ ன றி க கு ய ப ப து ம**”. That means that when the T.V. Gopal Iyer 1984 edition reads

“ஓதுவார்தமை நன்நெறிக்கு உய்ப்பது”, a modern edition based on BnF indien 113 would read “ஓதுவார்தமை நன்நெறிக்கு உய்ப்பதும்”. And there are in fact many other small variant readings, which is of course to be expected, but we shall now return to the global picture.

அகத்தியர் தேவாரத்திரட்டு ms (BnF indien 113) and the corresponding statistics

I shall now provide a number of statistics concerning Indien 113, considered as one test case, before touching upon other test cases, keeping in mind another consideration, which is the possibility to produce FACSIMILE of

palm leaf manuscripts, which we shall examine in a following section. For that purpose, I shall now provide the audience with a **print-screen** (see **Figure 5**) of a window open on an XML editor and displaying the full content of the RAW transcription of a single leaf.

```
<leaf>
    <coordinates>BnF indien 113, PDF file, page 36</coordinates>
    <PDF_page>36</PDF_page>
    <incipit_found>6</incipit_found>
    <identification>தேவாரம்</identification>
    <verse_id>3_49</verse_id>
    <folio_side>v</folio_side>
    <column nature="LM">
        <line>ந ம ச</line>
        <line>சி வ #</line>
        <line>ய த தி</line>
        <line>ரு ப ப</line>
        <line>தி க ம</line>
        <line>ப ண</line>
        <line>க வு சி</line>
        <line>க ம -</line>
    </column>
    <column nature="Main">
        <line>ச சி த த த த வ # க @ ட ளி ந து @ த றி ன வி த த க நீ ற ணி வ # # வி னை ப ப @@ க</line>
        <line>க க த தி # ம # வ ன வ ஞ @ ச ழு த து @ ம - ல - <B2/> ந ற ற மி ழ ஞ # ன ச ம ப ந த னா ன ம</line>
        <line>@@ ற க ற ற வ ன HH க # ழி ய # ம ன ன னு ன னி ய வ HH ற ற மி ன ம # லை</line>
        <line>யீ @@ # ஞ சு ம HH ஞ @ ச ழு த து ற ற ன வ ல ல வ ரு ம HH ப # # வ @ # - ல க -</line>
        <line><B2/> தி ரு ச சி ற ற ம ப HH ல ம - <B1/><B2/> க # த ல # கி க க சி ந து க ண ணீ HH # ம ல கி @ ய # து வ #</line>
        <line># த @@ ம ந ன @ ன றி க கு ய ப ப து ம @ வ த ந # ன கி னு @ ம ய ப @ ப # ரு ள # வ து</line>
        <line>ந # த னா ம ந ம ச சி வ # ய @ வ - க - <B2/> ந ம பு வ # # ந ம ந # வி ன வி ற றி னா ல வ ம பு ந #</line>
    </column>
</leaf>
```

Figure 5: RAW transcription of a single leaf (view 36 in BnF Indien 113)

As can be seen, the **XML element** under consideration is called **`leaf`**, and contains **children elements**. Some of those children elements provide external information, useful for the navigation or the indentification, whereas other children elements contain the actual segmented transcription, the more significant one being **`column`**, which is further segmented into **`line`**. In order to enhance the readability, I now provide two additional print-screen images, one displaying the **LM (=Left-Margin) Column** (see Figure 6), which signals that the leaf contains the incipit of the நமச்சிவாயத்திருப்பதிகம் and that it must be sung in a பண் (or musical mode) called கவுசிகம், and the other one containing the Main Column (see Figure 7).

```
<leaf>
    <coordinates>BnF indien 113, PDF file, page 36</coordinates>
    <PDF_page>36</PDF_page>
    <incipit_found>6</incipit_found>
    <identification>தேவாரம்</identification>
    <verse_id>3_49</verse_id>
    <folio_side>v</folio_side>
    <column nature="LM">
        <line>ந ம ச</line>
        <line>சி வ #</line>
        <line>ய த தி</line>
        <line>ரு ப ப</line>
        <line>தி க ம</line>
        <line>ப ண</line>
        <line>க வ சி</line>
        <line>க ம -</line>
    </column>
```

Figure 6: RAW transcription of LM column inside view 36 in BnF Indien 113)

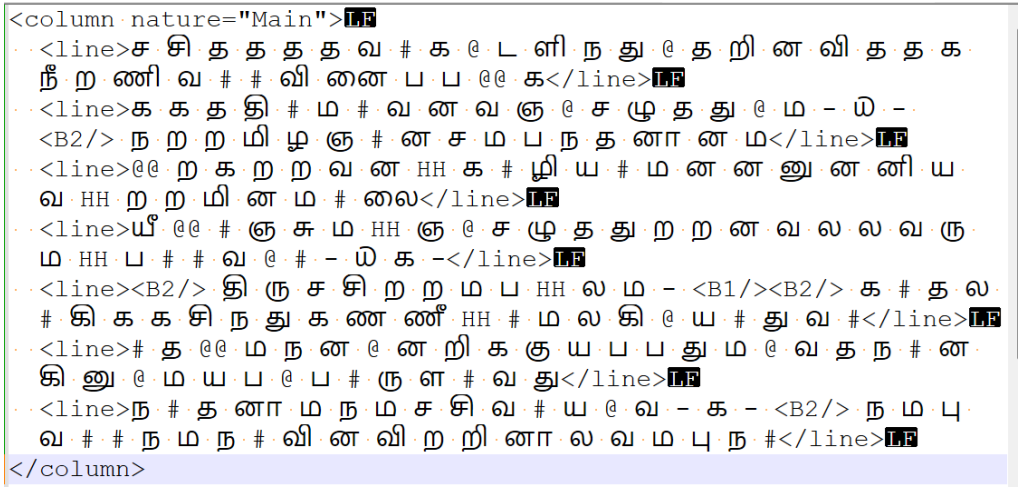

Figure 7: RAW transcription of LM column inside view 36 in BnF Indien 113)

Briefly described, the content of Figure 7 consists of the following:

- the end of stanza 10 inside Tēvāram 3-22 (on lines 1 & 2), corresponding to what T.V. Gopal Iyer prints as ((~~புத்தர், சமண் கழுக் கையர், பொய் கொளாச்~~ //)) சித்தத்தவர்கள் தெளிந்து தேறின; // வித்தக நீறு அணிவார் வினைப்பகைக்கு // அத்திரம் ஆவன---அஞ்சுஎழுத்துமே.
- Stanza 11 inside Tēvāram 3-22 (on lines 2, 3 & 4), starting with நற்றமிழ் ஞானசம்பந்த னான்மறை கற்றவன்
- Invocation திருச்சிற்றம்பலம் (on line 5)
- Stanza 1 from Tēvāram 3-49 (on lines 5, 6 & 7), already presented above

- The beginning of stanza 2 from Tēvāram 3-49 (on line 7), which **almost corresponds, but not exactly**, to what T.V. Gopal prints as நம்புவார்அவர் நாவின் நவிற்றினால், // வம்புநா [ள்மலர் வார் மது ஒப்பது;]

Inside the RAW transcription, the hymns and the stanzas inside them are separated by means of “balises”, which are seen as **empty elements** B1 and B2, which appear as **<B1/>** and **<B2/>.**

Additionally, one can see on the transcription

- numbers such as “– ൟ –“, “- ൟ க –“ on lines 2 and 4, for the last two stanzas of Tēvāram 3-22
- number “- க –“ on line 7, for the first stanza of Tēvāram 3-49
- indications such as **HH** in order to show the **String hole position**

Why is it useful to note down such information? My intention in performing those notations is to try to produce as faithful a facsimilé as possible SOLELY on the basis of the textual STRINGS which constitute the RAW format. While deciphering a text in a MS, one constantly navigates between two independant hierarchical structures:

- the **physical structure** of the manuscript, where what is visible is the succession of leaf sides, made explicit by the <leaf> elements, which are themselves divided in <column> and in <line>, with a difference of status between the main column (giving the base text) and the marginal columns (providing additional information).
- the **logical structure** of the text written on the MS, which here in the case of the அகத்தியர் தேவாரத் திரட்டு is expressed by the <B1/> and <B2/> balise inserted inside the <line> elements.

If at the beginning of the transcription of a not yet identified MS the **dominant structure** is the physical one, when the identification of the text and its structure progresses, it can become conceivable to invert the relationship between between the two structures and to make the logical one the dominant one, the physical structure being then expressed by means of balises.

How is it possible to live with an ambiguous writing system? The power of memorization.

Bibliography

Chevillard, Jean-Luc, 2003, “A proposal for the Digital Encoding of Palm-Leaf Tamil Manuscripts”, pp. 105-116, in K. Kalyanasundaram & V. Krishnamoorthy (Eds.), *Tamil Internet 2003, Conference Papers,* sponsored by The Government of Tamil Nadu, Organized by Tamil Virtual University & Anna University, Supported by International Forum for Information Technology in Tamil (INFITT) & Kanithamizh Sangam. ((Link: **“https://www.infitt.org/ti2003/papers/17_jlchevi.pdf”**))

Chevillard, Jean-Luc, 2024, « For the sake of future Tamil philologists / புத்தம்புது புலவர்களுக்காக », Live presentation in Chennai on 2024 february 10th, at the event called கணித்தமிழ்மாநாடு 24 [kaṇittamiḻ mānāṭu 24] organized by the TVA (Tamil Virtual Academy): [LINK hal-shs: **“https://shs.hal.science/halshs-04455964v1”**]

Digital Tēvāram, 2007 (online version): https://www.ifpindia.org/digitaldb/site/digital_tevaram/INDEX.HTM

Digital Tēvāram, 2007 (published CD-Rom): https://www.ifpindia.org/bookstore/ci103/

Gopal Iyer, T.V., 1984, *Mūvar Tēvāram (vol. 1, Ñāṉacampantar)*, Publications de l'Institut Français d'Indologie (PIFI) 68-1, Pondicherry.

Gopal Iyer, T.V., 1985, *Mūvar Tēvāram (vol. 2, Appar and Cuntarar)*, Publications de l'Institut Français d'Indologie (PIFI) 68-2, Pondicherry.

Gopal Iyer, T.V., 1991, தேவாரம் ஆய்வுத்துணை (Tēvāram, Etudes et glossaires tamouls), Publications de l'Institut Français d'Indologie (PIFI) 68-3, Pondicherry.

Krishnan K.G., 1984, *Karandai Tamil Sangam Plates of Rajendrachola I,* Memoirs of the Archeological Survey of India No. 79, New Delhi/Mysore.

Trento, Margherita Trento & Jean-Luc Chevillard, 2025. « Fishing for Words with Beschi in Tamil Traditional Poetical Vocabularies ». Dept. of Indian and Tibetan Studies, Universität Hamburg. *South-Indian Manuscripts in Hamburg and Paris: Paratexts and Provenance, TST*

project (ANR & DFG, 2019-2023), edited by Emmanuel Francis and Eva Wilden, **pp.501-567**, 2025. (hal-05325854) ((Book in **OPEN ACCESS** at **doi.org/10.25592/uhhfdm.18223**))