

HOW TAMIL WAS DESCRIBED ONCE AGAIN: TOWARDS AN XML-ENCODING OF THE *GRAMMATICI TAMULICI**

JEAN-LUC Chevillard

UMR 7597 Histoire des théories linguistiques
(CNRS – Paris Diderot – Paris Sorbonne nouvelle), France

Abstract

Although Tamil has an indigenous tradition of «grammatical» description — in an extended sense of «grammatical» which encompasses poetics and metrics, along with phonetics, morphology and syntax — which goes back to the first half of the first millenium AD, it was described once again from the 16th century onwards from a new angle (which included an interest for ordinary language) by Christian missionaries who brought with them a Latin model of grammatical description, which they tried to apply, by creative extension. I shall therefore refer to the corpus of their productions as the *Grammatici Tamulici*, although the earliest among them make use of Portuguese as a metalanguage for the description of Tamil. Not being in the situation of some field linguists, who have to start from scratch when confronted with a language which has never been written, those missionaries progressively discovered that the Latin alphabet, although enriched by several extensions developed for the representation of European vernaculars, was a less efficient tool than the local (Tamil) syllabary for noting down Tamil sounds and words. They also discovered that their capacity to be convincing, in front of their converts, depended in a great part on their adopting the language hierarchies which governed (and still govern) the diglossic Tamil society, as we shall see here in this preliminary

Résumé

Bien que la langue tamoule possède une longue tradition de description «grammaticale», qui remonte à la première moitié du premier millénaire de notre ère, avec une acception large de «grammatical», qui inclut la poétique et la métrique, dans un champ où sont présentes la phonétique, la morphologie et la syntaxe, cette langue a été décrite, à *nouveaux frais*, et dans une nouvelle perspective (qui incluait l'enseignement de la langue courante), à partir du xvi^e siècle, par des missionnaires chrétiens, qui apportaient avec eux un modèle latin de description grammaticale, qu'ils tentaient d'adapter, de façon créative, à une réalité linguistique qui était pour eux nouvelle. Pour cette raison, il sera ici fait référence au corpus de leurs productions comme étant celui des *Grammatici Tamulici*, bien que les premiers d'entre eux aient utilisé le Portugais comme métalangue pour décrire le tamoul. Ne se trouvant pas dans la situation de certains linguistes de terrain, confrontés à une langue qui n'a jamais été écrite, ces missionnaires se rendirent progressivement compte du fait que le syllabaire tamoul était un outil beaucoup plus efficace pour noter les sons et les mots tamouls que l'alphabet latin, même enrichi par les extensions développées pour la notation de langues européennes diverses. Ils découvrirent aussi que leur capacité de convaincre et de convertir dépendait en grande partie de leur adoption des hiérarchies langagières qui gouvernaient (et

* For reading a preliminary version of this article and for making judicious suggestions, I would like to express here my thanks to Eva Wilden, Émilie Aussant, Victor D'Avella, Cristina Muru and Dominic Goodall. The responsibility for all errors is of course mine.

exploration covering five authors who were active during a period of ca. 200 years, up to the year 1739.

gouvernement toujours) la diglossie tamoule, comme nous le verrons dans cette exploration préliminaire, qui couvre cinq auteurs, actifs pendant une période qui va jusqu'à 1739.

Keywords

Tamil, Tamil grammars, *Grammatici Tamulici*, Extended Latin Grammar, Tamil syllabary, Tamil Phonetics, Tamil collation order, diglossia, ordinary language, 16th-18th centuries

Mots-clés

Tamoul, grammaires tamoules, *Grammatici Tamulici*, Grammaire Latine Étendue, syllabaire tamoul, phonétique tamoule, ordre alphabétique tamoul, diglossie, langage ordinaire, xvi^e-xviii^e siècles

C'est parce qu'il y a derrière la variété apparente des langues modernes de l'Europe un même fond latin qu'elles se laissent traduire exactement les unes dans les autres car on ne saurait traduire vraiment une langue vraiment étrangère.

Antoine Meillet [1866-1936],

Esquisse d'une Histoire de la langue latine, [réimpression]

Klincksieck 1977, p. 283.

INTRODUCTION

The five¹ documents (See [Chart 1](#) and [Chart 2](#)) which are briefly examined in this article were produced by five authors (HH, AP, BZ, CJB and CTW) using two meta-languages (Portuguese and Latin) during a time-span of almost two centuries and are the remaining traces of some “real-life experiments”, which modern linguists might want to call “field-work”, although the original intention present in the composition of those five texts, in a missionary context, is probably better captured by the mottos “SOLI DEO GLORIA” (“Glory to God alone”) and “*Omnis Lingua laudet Dominum*” (“Let every tongue praise the Lord”), which are printed next to the word “FINIS” (“end”) at the end of two of those texts (CTW and BZ).²

Those linguistic experiments took place when some Western missionaries, some of them catholic (HH, AP & CJB) and others protestant (BZ and CTW), were posted in Tamil Nadu, and tried to learn the local language and, thereafter, to teach it to their junior colleagues, who would then be preaching in front of converts, or

1 Apart from those five, other early documents produced by early scholars such as Gaspar de Aguilar (b. 1588), Balthasar da Costa (c.1610-1673), Philippus Baldæus (1632-1672) also exist but are even less easily accessible, for the time being. See for instance the 2014 volume chapter “Gaspar de Aguilar: A Banished Genius”, by C. Muru.

2 Additionally, in HH's MS, the word **ᲚᲘᲗ** [CECU] («Jesus») is written on the top of every folio, often accompanied by **ᲛᲚ᲏Თ** [MARIYA] (“Mary”) and AP's Vocabulario ends with an invocation to “God” (“Deo”) and to “Virgini Deiparæ sanctissimæ”. As for CJB, his 1738 printed title page contains, above the title, the abbreviated Jesuit Latin motto “A.M.D.G.” (Ad Maiorem Dei Gloriam).

CHART 1
Time-line for five early missionary descriptions of Tamil

Abbreviation & title	Author's name, dates of Birth and Death (and stay in India)	Work Date, type & Meta-language used	English translation
HH <i>Arte em Malauar</i> ^a	Henrique Henriques 1520-1600 (1546-1600)	16 th c. MS grammar (in Portuguese)	J. Hein & V.S. Rajam (HOS 76, 2013)
AP Vocabulario Tamilico	Antão de Proença 1625-1666 (1647-1666)	1679 Printed vocabulary (in Portuguese)	No ^b
BZ Grammatica Damulica	Bartholomæus Ziegenbalg 1682-1719 (1706-1714 & 1716-1719) ^c	1716 Printed grammar (in Latin)	Daniel Jeyaraj (English transl. 2010)
CJB <i>Grammatica Latino-Tamulica</i>	Constantius Joseph Beschii 1680- c.1746 (1710-1746)	1738 Printed grammar (in Latin)	Horst (1806 ¹ , 1813 ²) Mahon (1848)
CTW <i>Observationes Grammaticæ</i>	Christoph Theodosius Walther 1699-1741 (1724-1740)	1739 Printed grammar (in Latin)	No

^a Another title (*Arte da Lingua Malabar*) is possible (and more current). I extract this title, slightly arbitrarily, from a sentence which appears on top of folio 8 (recto) and is transcribed in Vermeer (1982, p.5, ll.6-7) as “Em nome de nosso senior Jhesu Christo começa a arte em malauar”. The English translation by Hein & Rajam [2013, p. 38] reads: “In the name of our Lord Jesus Christ the Arte in Malabar begins.”

^b Proença's dictionary has been extensively studied by G.James in several publications.

^c BZ made a two-years journey (1714-1716) to Europe. See Jeyaraj (2010).

CHART 2
Brief material description of HH, AP, BZ, CJB and CTW

Abbreviation & title	Work Date & Meta-language	Type of text & nature of source	Number of images (or pages)	Magnitude as (plain) texts
HH Arte em Malauar	16 th c. Portuguese	Grammar. Undated MS discovered by Thani Nayagam [1913-1980] and critically edited in 1982 by H.J. Vermeer [1930-2010].	159 folios (recto & verso)	ca. 160.000 signs
AP Vocabulario Tamulico	1679 Portuguese	Vocabulary (not lemmatized). Printed posthumously in Ambalacatta (1966 Facsimilé Reprint by Thani Nayagam, Kuala Lumpur).	18 + 508 pp	ca. 500.000 signs
BZ Grammatica Damulica	1716 Latin	Grammar. Printed in Halle.	15+128 pp.	ca. 93.000 signs (on the 128pp)
CJB <i>Grammatica Latino-Tamulica</i>	1738 Latin	Grammar. Composed in 1728 [cf. Preface]. Printed in 1738 in Tranquebar. Reprinted in 1813 (Fort St George) and in 1843 (Pondicherry).	175 pp.	ca. 220.000 signs
CTW <i>Observationes</i>	1739 Latin	Grammar. Printed in 1739 in Tranquebar ^a .	56 pp.	ca. 70.000 signs

^a Cambridge University Library has a handwritten copy of CTW's grammar, probably produced on the basis of a printed copy.

performing various Christian rites, and who would also try to translate texts (such as the Gospel) into Tamil. Mastering the linguistic complexity of Tamil was not an easy task, for many reasons, the first one being the absence of prolonged anterior contacts between the linguistic area of origin of those missionaries, namely Europe, and their new field of activity, which was Southern India. Besides being important as external³ sources for the History of the Tamil language, those five texts are also of interest for the History of Descriptive Linguistics⁴ and of Linguistic typology. Additionally, it should be emphasized that one of the challenges for anyone who engages in this exploration is properly to understand (and explain) the nature of the Tamil Diglossia (or of Tamil language hierarchies), but that, if that challenge is successfully confronted, it provides insights about the dynamics of standardization and its long-term effects.

1 NATURE OF THE SOURCES: THE *GRAMMATICI LATINI* CORPUS

Although I shall try to keep the technical side of this project in the background, it is necessary to state here first that the questions which are dealt with here arose in the course of a (still very incomplete)⁵ process of transforming the five ancient documents named in [chart 1](#) into XML documents. The accomplishment of this task has made it necessary for me to try to make explicit every aspect of the underlying (intended) encoding which I postulate to have been (implicitly) present in the minds of the authors and of their audience, mediated through the agency of the printing press operators, in the case of the last four texts.⁶ Although it entails many choices which may appear subjective, the explicitation⁷ process performed on the basis of a direct examination of the Portuguese and Latin originals is

3 The internal sources, which go back to the first millenium BC (3rd cent. BC), if we include epigraphy, are of course much more abundant, but will not be our topic here.

4 As explained in [Chevillard \(2015\)](#), this research has to be seen as a part of a collective research program, called “Extended Grammars” (French “Grammaires Étendues”), which has its roots in the work of Sylvain [Auroux](#), who is responsible for coining the expression “Grammaire Latine Étendue”. From my point of view, we potentially have a “grammaire étendue” event whenever someone creatively, and boldly, uses a Language A for trying to describe a Language B, especially if this is the first such (maiden) attempt at extending the virtual reference of the terminology. Such an extension always has practical consequences and may be felicitous (or unfelicitous). A long series of agnostic observations on the terminological practices of successive grammarians (hopefully trying to describe the same language, across centuries) and the manner they are received, is of course necessary.

5 At the time of this writing (on 5th september 2017), the global degree of completion for the entering of the corpus is 30%. This average is calculated on the basis of the individual degrees of completion for the FIVE texts (ponderated by their lengths). Those individual degrees are: HH (8%), AP (22%), BZ (15%), CJB (64%), CTW (100%).

6 In the case of HH, this remark may apply to the copyist, unless we want to see the available source as an autograph MS.

7 “Explicitation process” might mean for instance deciding which tags will be used when a word is in italics in a printed book.

unavoidable, and is of course rendered much easier by the fact that for several of them, I am not the first one to deal with that task, having been most of the time preceded by intermediate editors (for 4 of the 5 texts) and by translators (for 3 of them).

As already stated, the exploration of the corpus, of which these five texts are a part, is at an early stage, and the challenges are many, the first one being that the language which is used in three of those texts and which was dominant at the time, in scholarly circles, namely Latin, has long been replaced in scientific communication by other languages, among which English is nowadays clearly in a dominant position. This means that Latin is no longer the transparent instrument of knowledge which it may have been in those days and has now become itself a partly opaque object of study. The same can be said of Portuguese, which is found in the two other texts (HH and AP), as is clear from the fact that, as remarked in the second preface to [Hein-Rajam \(2013\)](#), [Vermeer \(1982\)](#) did not have many readers, because:

- (1) The original grammar was not a book in which anyone could browse. The only persons who could possibly use the grammar were those remarkable persons, hardly existent in East or West, who could read both old Tamil and old Portuguese. (Norvin Hein, 2nd Preface to Hein, Jeanne and V.S. Rajam (2013, p. vii)

To this, we could add that the same is partly true of [Hein-Rajam\[2013\]](#), which is not really a stand-alone book, because it does not contain the original Portuguese text. Anyone who really wants to make a serious study of HH is in need of at least these three items: (1) a Facsimilé of the original MSS; (2) Vermeer's edition of the original text and (3) Hein and Rajam's English translation. Those three components (of a future electronic edition) should then be supplemented by various tools (indices, intertextual extensions...). And the same statement applies, *mutatis mutandis*, to the other future components of the *Grammatici Tamulici*, of which the five texts mentioned in [Chart 1](#) are of course only a kernel.

2 ELEMENTS OF TAGGING

The essence of XML is tagging, but if I want to be more specific, concerning the present task at hand, where the preliminary target itself, as summarized in [Chart 2](#), could be described, before tagging, as a plain text file containing more than one million characters,⁸ the preliminary steps towards a usable enriched (XML) file could be described as introducing in the texts under examination a first layer of simple specific tags which will allow us to extract automatically (using XSLT scripts), for further treatment, homogeneous subsets from the global corpus, the

⁸ This estimation is based on an addition of the approximate figures contained in chart 2, which reflect variable degrees of completion for each of the texts.

most obvious candidates for that being the linguistically homogenous fragments, which can be expected to be:

- segments in Tamil.
- segments in Latin (in the case of BZ, CJB and CTW).
- segments in Portuguese (in the case of HH and AP).
- segments in other languages.⁹

However, as should become progressively clear, the situation is of course much more complex than that because

- a segment in Tamil can be written: (a1) in the Tamil script (using several possible orthographies), or (a2) in systematic transliteration, or (a3) in approximate transcription (of various types).
- a segment in Latin can be: (b1) part of the metalinguistic discourse concerning Tamil; (b2) the Latin translation of a Tamil example; (b3) an instance of a Latin sequence which is object-language because it is used as an illustration (possibly in a comparison); (b4) a mixed segment where Latin has been supplemented by Greek (see Section 8).
- Similarly, a segment in Portuguese can be (c1) part of the metalinguistic discourse concerning Tamil; (c2) the Portuguese translation of a Tamil example; (c3) an instance of a Portuguese sequence which is object-language because it is used as an illustration (possibly in a comparison); (c4) a mixed segment where Portuguese has been supplemented by Latin (see Section 8).

3 WRITING DOWN AN OBJECT LANGUAGE WHICH HAS UNFAMILIAR SOUNDS, OFTEN DIFFICULT TO PRONOUNCE

The preceding remarks were of course very general, and the easiest way to make them more clear seems to be an examination of specific examples. We shall start by a relatively simple printed example taken from the 18th century CJB, because MSS witnesses (such as HH, which we shall examine later) are of course more difficult to handle. This (and other examples) should play the role of a direct window to the past, allowing the reader to have a brief direct glance on the raw material as it appeared in the original sources, in order to make clear the nature of the task in which those early descriptors were engaged. I have chosen a passage from [Beschi's 1738](#) printed book which illustrates the (phonetic) difficulties of “pronunciation” due to the discrepancy between the oral/aural spheres and the written sphere (i.e.

⁹ Concerning that last group, see my presentation “The Early Modern European Multilingualism, as seen in the Tamil grammars composed in Latin by Beschi and by Walther” (ROLD [Revitalizing Older Linguistic Documentation], Amsterdam, nov. 2016), to be published in the future.

the orthographic conventions). This passage is extracted from the third section of the 1st chapter (parag.8, p.15), which has as its title “De variatione Pronunciationis”.

As should progressively become clear, this passage illustrates several of the linguistic topics which will be touched upon by me in this article, but it also illustrates the nature of the source and of the competences which are necessary for making use of it, namely the simultaneous command of Latin and of Tamil. Fortunately for us, we can partly rely in this case, on a 1848 English translation, due to Mahon, which reads (with the addition of a special¹⁰ transliteration, in square brackets):

(2) Section III. Of the Variations in Pronunciation.

Sometimes, the *form* of a letter being unchanged, the *sound* of the same is varied: for which the Rules are these. Rule 1. *A*, short, at the end of a word which is a polysyllable, and which after the *a*, has for its last letter one of these six consonants, ல [L^a], ழ [L^a], ள [L^a], ர [R^a], ன [N^a], ண [N^a], then the *a* is pronounced with so gentle a sound that it seems *e* soft. Thus பகல் [PAKAL] is not pronounced *pagal*, but *paguel*, a day: in the same way புகழ் [PUKAL] is sounded *puguel*, praise; அவள் [AVAL] *avel*, she; சுவர் [CUVAR] *suver*, a wall; அவன் [AVAN] *aven*, he; அரண் [ARAN] *aren*, a citadel, &c. (Mahon, 1848, p. 13)

The attentive reader will have noticed that Mahon (whose translation I have tried to reproduce *verbatim*) is not as faithful as we might hope. A number of peculiarities have disappeared in the translation process. They are:

- The modernization of the Tamil orthography,¹¹ in the replacement of பகல் [PAKAL^a] by பகல் [PAKAL], of புகழ் [PUKAL^a] by புகழ் [PUKAL], of அவள் [AVAL^a] by அவள் [AVAL], of அவன் [AVAN^a] by அவன் [AVAN] and of அரண் [ARAN^a] by அரண் [ARAN].
- The suppression of the diacritics, when *puguel* becomes *puguel* and when *avel* becomes *avel*, a topic to which we shall come back in the next section.
- The disappearance of the Greek article τὸ (seen in the clause “tunc τὸ *a* tenui adeò fono pronunciatur, ut videatur *e* lene”), which is of course not necessary

10 The transliteration system used by me here is special because it tries to preserve the ambiguity inherent in the 1738 orthography, while at the same time indicating to the reader how the word is pronounced. This is achieved by a **dual** transcription. For instance, ல is transliterated either as [LA] or as [L^a], the second possibility being chosen when modern orthography makes use of a pulli (“dot”) and writes ல் (transliterated as [L]). As a clarification, let me conclude by saying that the 1738 orthography for the famous name written today as TOLKĀPIYAM (தொல்காப்பியம்) is TOL^aKĀPIYAM^a (தொல்காப்பியம்), which an inexperienced reader might be tempted to read as TOLAKĀPIYAMA.

11 For more details on the question of spelling, and on the difference between the various authors in that respect, see Chevillard 2015. There are in fact more than two orthographic systems.

in English (“then the *a* is pronounced with so gentle a sound that it seems *e* soft”), whereas the neo-Latin of Beschi (and of Walther) was in need of it.

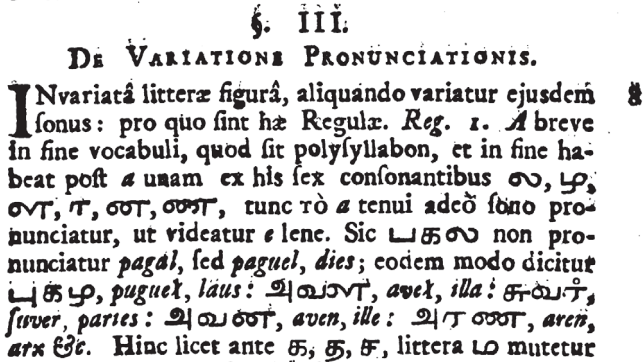


Figure 1 : CJB, Grammatica Latino-Tamulica [1738, p. 15, parag. 8]

Importantly, it should be clear from [figure 1](#) and from the remarks which I have just made, that the printer who printed [Beschi's book in 1738](#) did NOT manage to give three distinct representations to the following three consonantal items,¹² ஸ [L^a], ஸ் [L^a], ஸு [L^a], which would nowadays be transliterated as “l, l̥ and l̥̥” (as per the University of [Madras Tamil Lexicon](#) (MTL), but which appear in the passage reproduced from the 1738 book as “l” (end of “pagal”, on line 9), as “l̥” (end of “puguel”, on line 10) and as l̥̥ (end of “avel”, on line 10). In the modern system of transliteration, the three words which have those three distinct “l” as their endings would be transliterated as *pakal*, *pukal* and *aval*.

4 OPENING A SECOND WINDOW: HH'S *ARTE EM MALAUAR* AND THE “TRES MANEIRAS DE L”

In order to explain the discrepancy noted in the previous section between Mahon and Beschi concerning *puguel* (becoming *puguel*) and concerning *avel* (becoming *avel*), or, rather, in order to clarify the nature of the information which is contained in Beschi's text but which Mahon has lost, it is necessary to travel back in time. We shall have the occasion to see that opinions differ (between Vermeer and Hein-Rajam) on the question whether, in the 16th century, HH had been more efficient at noting down certain distinctions which his successors seem to neglect. In order for the readers of this article to make their own opinion, I shall now open a second window on a more ancient document, namely HH's *Arte em Malauar*, from which the following illustration ([figure 2](#)) is extracted. In this passage, HH explains, for the first time to a Western audience, his observations on ஸ, ஸ் and ஸு.

12 These three belong to the list of 18 consonants.

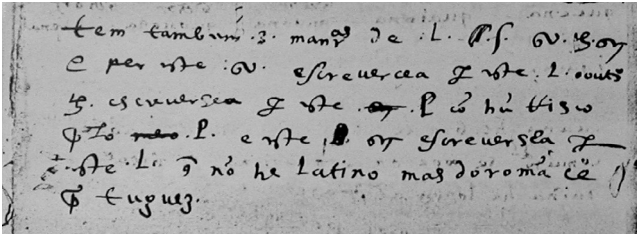


Figure 2 : (HH, folio 5v, extract)

I shall first of all reproduce below the transcription for this passage found in the 1982 critical edition by Hans J. Vermeer and the joint English translation for the same by Jeanne Hein and V.S. Rajam, which has been in the making for a very long time, but which came out only recently, in 2013. They are as follows:

- (3a) Tem tambem 3 maneiras de l, scilicet ω , μ , σ , e este ω escreuersea per este l, o outro μ escreuersea per este l, cõ hu Risco polo l, e este σ escreuersea per este l que nõ he latino, mas do romãce portuguez. (Vermeer, 1992, p. 3 [line 33] & p. 4 [lines 1-3])
- (3b) There are also three ways of writing l: ω , μ and σ . They will be written thus:
 ω l
 μ l (with a stroke through the middle of the l)
 σ L (which is the Portuguese Romance l, not the Latin). (Hein & Rajam 2013, p. 36)

We are lucky to have two publications available, trying to do justice to HH's MS, but, as will be noted by perceptive readers, Vermeer (1982) and Hein & Rajam (2013) do not agree on how to transcribe this difficult passage in the MS, because the former takes HH's notation for σ to be «l», whereas the latter pair thinks HH has written «L», using a capital letter.

The reader of this article might want to examine other pieces of evidence, such as the four final lines (i.e. lines 8 to 11) inside Figure 3, which is provided on the next page and where we have some items belonging to the declension of the Tamil equivalent of arroz "rice", which are transcribed by Vermeer and by Hein-Rajam, respectively, as

- (4a) "choRugal", "choRugalucu", "choRugalæi", "choRugalile" (Vermeer 1982, p. 18, lines 25, 26, 28 & 29)
- (4b) "choRRugaL", "choRRugaLucu", "choRRugaLæi", "choRRugaLile" (Hein & Rajam 2013, p. 58)

I tend to think that Hein & Rajam are right in their transcription and that the word "choRRugaLile" (on the last line of Fig. 3) is a clear piece of evidence that the scribe was making a conscious distinction while shaping the two types of "l", although he may not have been consistent throughout the whole MS, for every single occurrence.

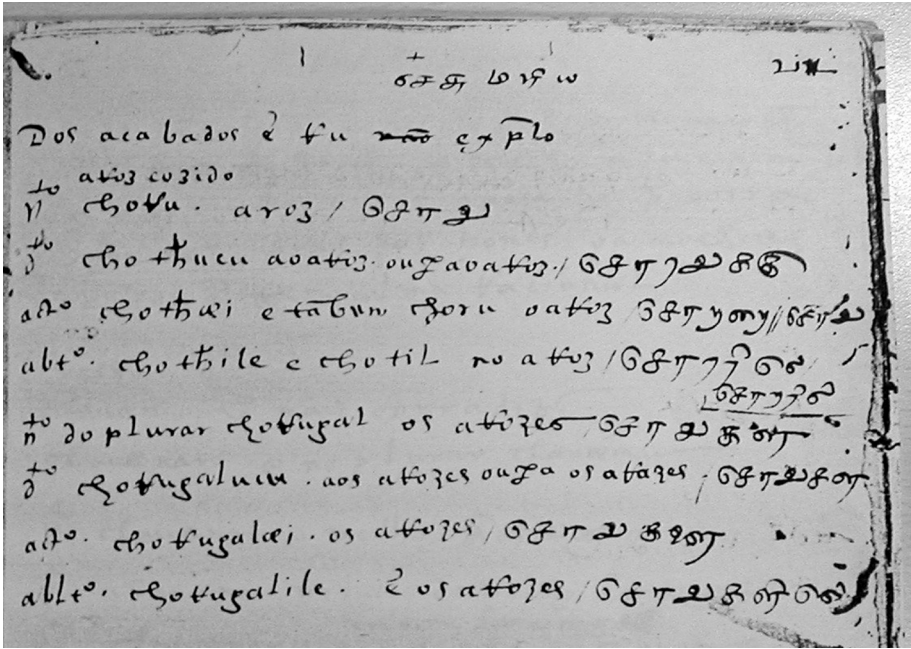
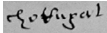


Figure 3 : HH, Arte em Malauar, Folio 21r, extract

5 EXPLAINING “TRES MANEIRAS DE R” (PARTLY ENTANGLED WITH T-S AND D-S)

While comparing (4a) and (4b), for the sake of deciding whether the 16th c. MS makes use of a capital L or not, what the reader may have FIRST noticed is in fact the difference between the (4 times repeated) single “R” (in 4a) and the (4 times repeated) double “RR” (in 4b). Both “R” (used by Vermeer) and “RR” (used by Hein & Rajam) are different conventions for reproducing the symbol which appears as  in HH’s MS as a notation for consonant *ɾ*, concerning which HH declares (on folio 5r of the MS) that its pronunciation is identical to the “r dobrado”. That “r dobrado” is a feature in the Portuguese of HH, and we see it used in the MS not only for transcribing a Tamil word such as “choRugal” (alias “choRRugaL”), which appears as , but also for writing ordinary Portuguese words, such as “Risco” in example (1a), translated as “stroke” by Hein and Rajam, and written  inside the MS. We are now entering a domain whether the task at hand consists in distinguishing

- “tres maneiras de r” (three types of r), namely rh, r and R (i.e. intervocalic $\perp$ [t], σ [r], ρ [r])
- “tres maneiras de t”, namely th, t, th (i.e. double $\perp$ [t], initial and double th [t], double ρ [r])

- three types of d, represented by dh, d and dḥ (for 𑌤[t], 𑌤[t] and 𑌤[r] in post-nasal position)

However, since the space devoted to phonetic questions in this brief presentation of the corpus of *Grammatici Tamulici* has to be limited, I shall not discuss fully the topic of the three r-s, three t-s and three d-s, limiting myself to the observation that what is normally (but not systematically) represented by “rh” in HH’s text is occasionally represented by “ḍ” in CJB’s text, as in parag. 6 (p.13) when he writes the word-form “pattirattinôḍḍē”. The corresponding sound (noted “rh” by HH and “ḍ” by CJB) would probably be described nowadays as “retroflex flap”, and the reader might also be interested in reading its description by CJB, which is:

- (5a) Scilicet 𑌤 [ṭ̣ᵃ], hæc, quando simplex est, pronunciatur hoc modo: inverfâ omninò retrorſum linguâ, adeó ut interiorem palati ſummitatem attingat, impellitur impetu, pronunciando inter da et ra. (Beschi, p. 12, par. 4)
- (5b) For instance 𑌤 [ṭ̣ᵃ]; this when single is pronounced in this way: the tongue having been turned back as far as possible, so as to touch the highest part of the interior of the palate, is impelled forward with some force, pronouncing between *da* and *ra*. (Mahon 1848, p. 10)

We can certainly conclude from the fact that there are countless transcription mistakes¹³ in the MS of HH’s *Arte* and from the fact that Beschi very rarely uses the symbol “ḍ” and never attempts to distinguish in transcription 𑌤 [Ṇ] and 𑌤 [Ṇ] (both represented by simple “n” in figure 1) that our missionary linguists must have quickly realized that the only reasonable solution for writing Tamil words was to use the Tamil script, even though a transliteration system is occasionally seen, especially in the initial part of those grammars. Nevertheless, even though he uses the Tamil script, it is clear that AP (who will be our next topic), when he compiled his *Vocabulario*, which was posthumously published in 1679, still had the Latin alphabet in mind, because the words are ordered on the basis of their pronunciation and, therefore, on the basis of the lexicographic order which they would have if written by means of the Latin alphabet, as we shall see in the next section.

6 HOW PROENÇA HANDLED THE PHONETICS OF TAMIL, WHILE ORDERING 16,208 ITEMS

In this section, I shall briefly deal with AP, who stands between HH and CJB in time and whose work represents an impressive effort at mastering the lexical wealth of Tamil. In order to explain his strategy, we must explain how he established a mapping between the Portuguese/Latin alphabet, and the Tamil syllabary. The

13 The mistakes consist for instance in using “r” where one would expect “rh” or in using “ṭ” when one would expect “th” (for a retroflex t).

easiest way to do that seems to be to give a Bird's eye-view of the 508 pages of his Vocabulario, which contain 16,208 entries in two columns,¹⁴ and which are divided into 28 sections, each one being headed by a Capital Latin letter, with a few exceptions, as will appear from [Chart 3](#).

The first comment which can be made on this chart is that for anyone used to a modern Tamil dictionary, the words are in a very unnatural order.¹⁵ In a typical Tamil dictionary, we would have an initial section containing all the words starting with one of the twelve Tamil vowels, and that Vowel-section would probably be the biggest. That section would be followed by a section containing all the words starting with the consonant க [K^a] combined with a vowel, and that section would probably be the second biggest. Eight more sections would follow, each containing the words starting with one of those eight other consonants (namely ச [C^a], ஞ [$\tilde{N}^a$], த [T^a], ந [N^a], ப [P^a], ம [M^a], ய [Y^a], வ [V^a]) which can, under normal circumstances, be found in initial position in a word. It should be added that the remaining nine consonants (out of a total of eighteen consonants) are found in other positions (medial and final), but not in initial position. If we take the example of the set of 1,091 polysemic words enumerated in the 10th section of the *Pinkalam* (a traditional Thesaurus¹⁶), we would have the proportions found in the first three columns of [Chart 4](#), which follows immediately below.

As for the content of the three other columns, it can be understood by reference to [Chart 3](#), because it indicates in which section of AP's Vocabulario a word whose spelling is known should be looked for. I should add as a clarification that the words with voiced initials, such as B, I (ச),¹⁷ D, G, which are found in some cells in the company of words with unvoiced initials are for the greater part borrowed from Sanskrit (or from some other language). Both fall under the same head because the distinction of voicing is not phonemic in Tamil in initial

14 I have provided, in [Chevillard \(2015, p. 121, Footnote 2\)](#) a coordinate system for giving precise references to any one of those 16208 entries. A reference such as 287_R_d (which designates the entry reproduced in [Figure 7, infra](#)) indicates that the entry is on the right (=R) column of page 287, and that it is the 4th item, counting from the top (because "d" is the 4th letter in the alphabet).

15 The normal order for the 12 Tamil vowels and 18 Tamil consonants is A Ā I Ī U Ū Ē Ē A I O Ō K Ṇ C Ṇ Ṭ Ṇ T N P M Y R L V Ḷ Ḷ. The words starting with the special letters which are sometimes used for writing those Sanskrit words which are not fully tamilized, such as Ṣ, KṢ, etc. are put at the end of the dictionaries.

16 NB: The reader should not believe that traditional Tamil thesauri are normally alphabetically ordered. The 10th section of the *Pinkalam* (which might be dated in the 9th century AD) is in fact the EARLIEST known example of a list of Tamil entries alphabetically ordered. That 10th section enumerates the meanings of 1091 polysemic words. The *Tivākaram*, an earlier Thesaurus (which might be dated in the 8th century AD), also contains a section dedicated to polysemic words, but that section (which contains 383 items) is not in alphabetical order.

17 See footnote a (chart 3, p. 14).

CHART 3
Table of Contents for Proença dictionary (1679)

Group rank	Page and Column where group starts	Group Heading	Entry Count inside group	+++++	Group rank	Page and Column where group starts	Group Heading	Entry Count inside group
1	1_L	A	1382	+++++	15	176_L	᳚ [Ñ]	17
2	44_R	A. ᳚ [Ā]	366		16	177_L	᳚ [O]	271
3	56_R	B	114		17	185_R	P	2128
4	60_L	Ch.	25		18	252_R	Q	2251
5	61_L	D	168		19	323_R	R	180
6	65_R	E	408		20	329_L	L.	22
7	77_R	G(a.)	70		21	329_R	S	268
8	80_L	I	552		22	337_R	T	1523
9	97_L	I. ᳚ [Ī].	62		23	383_L	V vogal.	597
10	98_R	Y ᳚ [Y].	60		24	401_R	V. ((᳚ [Ī]))	89
11	101_L	I. ᳚ [C] ^a	108		25	404_R	V. consoante.	1435
12	104_L	L	103		26	450_L	X	1843
13	107_R	M.	1273		27	507_R	᳚ [᳚].	2
14	149_R	N	870		28	507_R	᳚ [K᳚].	21
(continued on the right side)					TOTAL			16208

^a The capital I is here a “consonantal i”. The ᳚[C] is here used for the disambiguation. This corresponds to words which have a voiced palatal affricate (᳚) as initial, such as Sanskrit “JALAM” “water”, spelt “᳚᳚᳚ [CALAM^a]” (entry 101_L_d).

position.¹⁸ As for the items in the last row, they also mostly correspond to items borrowed from Sanskrit.

7 ALPHABETIZING ALSO ON THE SECOND SYLLABLE: HOW IT WAS DONE

I have explained the first step in the alphabetization of the 16,208 items enumerated by AP. But that does not tell us the whole story, because only 9 Tamil consonants (out of a total of 18) are found in word-initial position. In this section, I shall explain the strategy followed concerning the second syllable, which can start with any of the 18 consonants. At this point, it should be added that no Tamil word starts with two consonants and that, therefore, if we are looking for a pure Tamil word¹⁹ with initial consonant in AP's *Vocabulario*, it will start with C_1VC_2 where V is one among 11 possible vowels between the first consonant C_1 and the second consonant C_2 . The (collation) order²⁰ which applies to vowels in that position inside AP's *Vocabulario* is "A, Ā, AI,²¹ E/(Ē), I, Ī, O/(Ō), U, Ū" where the presence of the sequences "E/(Ē)" and "O/(Ō)" indicates that 17th-century spelling does not allow one to distinguish in writing between what is pronounced " $C_1\bar{E}$ " and what is pronounced " C_1E ", because both are written " C_1E ".²²

After these preliminary explanations, I shall now deal with the question of the lexicographical ordering for non-initial consonants, inside words starting with C_1VC_2 or with VC_2 , which is based on the collation order of [Chart 5](#).

8 FROM LATIN-ASSISTED PORTUGUESE TO GREEK-ASSISTED LATIN AS A META-LANGUAGE FOR THE DESCRIPTION OF TAMIL

In the introduction to this article, I have indicated inside [Chart 1](#) (column 4) which "main" metalanguage, Portuguese or Latin, was made use of by each author. However, to this must be added that other languages, in addition to Tamil and to the

18 Limitations of space do not allow me to say much more and I shall content myself with adding that in intervocalic position, the most significant opposition is considered to be the opposition between three degrees, PP, P and NP, where PP represents a duplicated (long) plosive consonant, P represents a single (weakened) consonant and NP represents a consonant combined with a nasal (NP).

19 There are examples of Sanskrit words starting with two consonants inside AP's *Vocabulario*: those words are written with a ligature, using the Grantha script, but dealing with them here would take us too far.

20 NB: that order is NOT the same as the traditional Tamil order which is A, Ā, I, Ī, U, Ū, E, Ē, AI, O, Ō, AU (see footnote 15).

21 The diphthong AI is a possible value for vowel V inside words starting with C_1V , unlike the case of words starting with a vowel.

22 The same applies to what is pronounced " $C_1\bar{O}$ " and what is pronounced " C_1O ", because both are written " C_1O ".

CHART 5
collation order for non-initial consonants

Rank	C ₂	Remarks	word initial position cross-reference	Chart 4 (native words)
1	ப [P]	if pronounced as Portuguese “b”	Chart 3	
2	ச [C]	if pronounced as Portuguese “ch”	group 3	
3	த [T]	if pronounced as Portuguese “d”	group 4	
4	க [K]	if pronounced as Portuguese “g”	group 5	
5	ஐ [Y]	noted as “y” in sub-headings	group 7	
6	ஜ [J] [written ʃ]	“Consonantal i”. Distinguished from the preceding item in word initial position. ^a	group 10	V8
7	ல [L]	See section 4: “tres maneiras de l”	group 11	
8	ள [L̥]		(group 12)	
			NOT FOUND	
9	ழ [L̥]		NOT FOUND	
10	ம [M]		group 13	V7
11	ந [N]		group 14	V5
12	ண [N̥]		NOT FOUND	
13	ன [N]	Called by AP “N de tres olhos” (“n of three eyes”)	NOT FOUND	
14	ஞ [Ñ]		group 15	V3
15	ள [Ń]		NOT FOUND	

Chart 5. (continued).

Rank	C ₂	Remarks	word initial position cross-reference	
			Chart 3	Chart 4 (native words)
16	ப [P]	if pronounced as Portuguese “q” ^e	group 17	V6
17	ஃ [K]		group 18	V1
18	ர [R]	See section 5: “tres maneiras de r” and “tres maneiras de t”.	(group 19)	
19	ற [R̥]		NOT FOUND	
20	஌ [Ṭ]		(group 20)	
21	ஸ [S]	(found in some Sanskrit words) ^f	(group 21)	
22	த [T]		group 22	V4
23	வ [V]		group 25	V9
24	ச [C]	if pronounced as Portuguese “x”	group 26	V2
25	ஷ [Ṣ]	(found in some Sanskrit words)	(group 27)	
26	க்ஷ [KṢ]	(found in some Sanskrit words)	(group 28)	

^a In the intervocalic position, we may have a (quasi-)minimal pair in the succession of entries 325_L_c (ஈராகாந் [RĀKAKAṆ^a] Escríuaõ da puridade). See also 325_L_1 (ஈராகாந் [RĀC^aCIYAM^a] Reino). More research is needed.

^b The *Vocabulario* normally disambiguate the sub-heading by using the Tamil character ஸ [L].

^c The English translations in this chart (“thick L”, “curved L”, “in of three eyes”) are due to Edgar C. Knowlton and Xavier S. Thani Nayagam, who provided an English translation of the original 11-page preface in the 1966 Facsimilé reprint of the *Vocabulario*.

^d The *Vocabulario* normally disambiguates the sub-heading by using the Tamil character ண [L̥].

^e In the 16th century, HH used “c” for transcribing ஃ [K^a] when followed by vowels a, o & u, and used “qu” for transcribing ஃ [K^a] followed by vowels e & i. However, in AP’s 17th c. *Vocabulario*, we find an empty section on page 60 (Left column), with capital C as a head, and explanations in Latin saying that words starting with “ca, co, cu” are to be found in the ஃ [K^a] section, which seems to be an indication to go to the page 252 (Right column), where the long section having capital Q as a head starts. That long Q section ends on page 323 (See Chart 3).

^f Especially those containing ligatures such as SN, ST, SP, etc.

main metalanguage, are also found in those documents, some of them being used as illustrations for increasing pedagogic efficiency²³ and some of them being used as metalinguistic resources, as will be shown briefly in this section.

In the case of HH, we can for instance compare the following three statements, where various components of verbal morphology are examined.

- (6a₁) O preterito se forma do presente *aquiren* mudado in *âden*: *pilaquiren* – *pi landen*.
(Vermeer 1982, p. 63, l.26)
- (6a₂) The preterite is formed from the present [tense ending] **-aquiren** changed into **-aden**: **piLaquiren**, **piLanden**. (Hein-Rajam 2013, p.124)
- (6b₁) O preterito se forma do presente *thquiren* ou *rquiren* **mutato** in *tê*: *pathquiren* – *patê*; *quorquiren* – *quorten*; (Vermeer 1982, p. 69, ll.7-8)
- (6b₂) The preterite is formed from the present by changing **-thquiren** or **-rquiren** into **-te**: ***pathquiren* – *patê*; *quorquiren* – *quorten*** (Hein-Rajam, 2013, p.135)
- (6c₁) O futuro formatur a presente *quiren* mutato jn *pen*: *pathquiren* – *pathpê*; *quorquiren* – *quorpen*. (Vermeer 1982, p. 69, ll.11-12)
- (6c₂) The future is formed from the present by changing the **-quiren** into **-pen**: ***pathquiren* – *pathpê*; *quorquiren* – *quorpen*** (Hein-Rajam 2013, p. 135)

It seems clear, from a comparison of the boldface and the underlined passages that (6c₁) and (6b₁) are composed in a hybrid language, or, rather, a mixed metalanguage combining Portuguese words and a few Latin words such as *formatur* (in 6c₁) and *mutato* (in 6b₁). Moreover, although (6a₁) seems to be a Portuguese sentence, the construction with *mudado* looks like the Latin “ablative absolute” construction used in 6c₁ and 6b₁.

I shall now provide one example (see 7a and 7b) which illustrates the use of the Greek article in the grammars composed by CJB and CTW. We have already encountered one example in Figure 1, and noted that Mahon’s translation (cited in 2), did not retain it. Before the examination of the example, I shall content myself with saying that, as per my current counts, based on a fully entered text for CTW and a partly entered text for CJB (see Footnote 8), I have collected 16 attestations of the Greek article in the 1738 text of CJB and 6 in the 1739 text by CTW.

- (7a) Quando verò τὸ *que* apud Lufitanos et Gallos vertitur latiné non per infinitivum, fed per fubjunctivum *ut*; tunc tamulicè eleganter utimur infinitivo. fic, *dic, ut veniat*, [VARAC^aCOL^aLU] &c. (Beschi, 1738, parag. 134, p.117)
- (7b) But when the *que* in Portuguese and French is rendered in Latin, not by the infinitive, but by the subjunctive *ut*, *that*; then in Tamul we elegantly use the infinitive: thus *dic, ut veniat*, *say that he may come*, [VARACCOLLU] &c. (Mahon, 1848, parag. 134, p. 96)

23 See my ROLD presentation (mentioned in footnote 9).

Unlike the medieval scholars of France, who, when writing in Latin, made use of the Old French definite article “li”²⁴, Beschi, who was Italian but who wrote primarily for the international (Jesuit) audience called *Societas Iesu* (S.J.), could not use his own mother tongue for writing and therefore had to revert to Greek. Mahon, whose mother tongue was English and who wrote for an English audience did not, of course, have the same problem, because English is well-equipped.

9 WHICH LANGUAGE TO DESCRIBE? THE PROBLEM OF DIGLOSSIA,
THE HIERARCHY OF REGISTERS, THE DYNAMICS OF DEPRECATION AND
THE FASCINATION FOR POETICAL TAMIL

I shall start this section, which touches upon the question of registers, by providing two windows on AP’s *Vocabulario* (Figures 4 & 5), accompanied by a transcription of entries 262_L_j and 263_R_k. This will be followed by a judgment made by CJB concerning those two entries. After that we shall examine a judgment passed by the 19th-century scholar Rhenius on his predecessors CJB and BZ, and the possible causes for the judgment.

Each of these windows allows us to see three entries, but those which are of direct interest to us here are the following two:

- (8a) கன்று [KAN^aRU] Quod கண்ணு [KAN^aNU]. bezerri-
nho, itê aruoesinha, ou plan-
ta tenra.
(AP 1679, Entry 262_L_j [3rd item inside Figure 4])
- (8b) கன்று [KAN^aRU]. Same as கண்ணு [KAN^aNU]. Calf. Also [means]. Sapling
of tree, or tender plant.
(My translation)
- (9a) கண்ணு [KAN^aNU]. Bezarinho nouo
(AP 1679, Entry 263_R_k [1st item inside Figure 5])
- (9b) கண்ணு [KAN^aNU]. Young calf.
(My translation)

I have chosen these entries because we have an explicit reference to them inside CJB’s 1738 grammar, where he says the following:

- (10a) Præterea eodem modo, quando littera ற [R] sequitur consonantem ன [N],
judicant aliqui posse promiscuè vel has duas litteras னற [N^aRA], vel
duplicem ன [N] scribi. Et Lexicon ipsum Tamulico-Lulitanum expresse
hoc habet, docens v.g. கன்று [KAN^aRU] posse scribi கண்ணு [KAN^aNU]
&c. Attamen quàm falsò hoc dictum sit, ex hoc ipso videri potest,
quod கன்றுக்கு [KAN^aRU^aKU] significat, *vitulo*, in dativo, et
கண்ணுக்கு [KAN^aNU^aKU] est, *oculo*.
(Beschi, 1738, parag. 12, p. 19)

24 See Lallot et Rosier 2005.

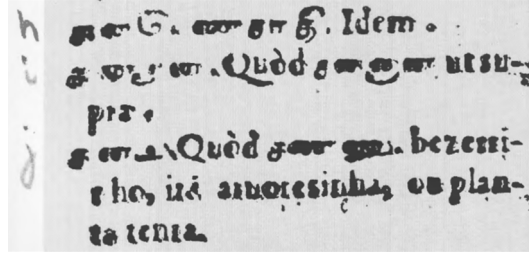


Figure 4 : AP 262_L_h to 262_L_j 1

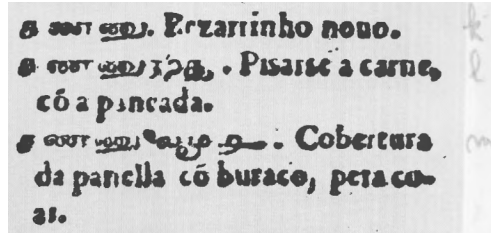


Figure 5 : AP 263_R_k to 263_R_m

- (10b) Moreover, in the same way, when the letter ற [R] follows the consonant ன [N], some decide that it may be written, indifferently, either as these two letters, ன்ற [NRA], or as a double letter, ன் [NN]. And the Tamul Portuguese Lexicon expressly has this, teaching for example that கன்று [KANRU] may be written கண்ணு [KANNU], &c. But how untruly this is stated, may appear from this very thing, that கன்றுக்கு [KANRUKKU] signifies, *to a calf*, in the dative; and கண்ணுக்கு [KANNUKKU] means, *to the eye*. (Mahon, 1848, parag. 12, p. 16)

Beschi's attitude will of course not be surprising to anyone who has been confronted directly with the Tamil diglossia and who knows that the very forms which Tamil speakers use every day are actively deprecated by those same speakers. Such an attitude generates of course a never-ending search for perfection. It is therefore interesting to note that we find inside the book *A Grammar of the Tamil Language*, published in 1836 by C.T.E. Rhenius [1790-1838], the following statement:

- (11) It is not the object of the above observation to detract any thing from the valuable works of Ziegenbalg, Beschius and others. They did in their days what they could in Tamil literature, and we are greatly indebted to them for the degree of knowledge they have given us of the Tamil language. But they have all failed in giving us pure Tamil; they have mixed vulgarisms with grammatical nicities [sic], and left us in want of a regularly digested Syntax. (Rhenius, 1836, p. ii)

The key word is of course “vulgarism” and the important thing is the attitude towards it, which Rhenius seems to have adopted from his Tamil teachers. Already in the 16th century, 300 years before Rhenius, HH had noted, while explaining an

கேள்¹-தல்[கேட்டல்] kēḷ-, 9 v. [K. M. kēḷ.] *tr.* 1. To hear, hearken, listen to; செவிக்குப் புலனுக்குதல். சொல்லுந் போலவுங் கேட்குந் போலவும் (தொல். பொ. 513). 2. To learn, be instructed in; பாடங்கேட்டல். ஒரு குறி கேட்போன் (நன். பொது. 42). 3. To ask, inquire, question, catechise; வினாதுல். என்னை நீ கேட்கையாலே (கைவல். 20). 4. To investigate; விசாரித்தல். களவு போனமை எங்குனே என்றுகேட்டு (Insc.) (சோழவ. 66). 5. To request, solicit, crave; வேண்டுதல். படிக்கப்

Figure 6 : Proença, entry 287_R_d

element of the conjugation system, namely the 1st person of the present tense of the verb கொல்லுதல் [KOLLUTAL] which means “matar” (“to kill”), and for which he has given “coliren” (“I kill”) as the ordinary form, that there is another possibility:

- (12a) Nos verbos desta conjugação os que muito sabê os pronosiaõ muitas vezes cõ gui antes do Ren: coluguiren. (Vermeer, p. 82, ll.8-9)
 (12b) Those who are more learned pronounce the verbs of the fifth conjugation [First class] with **-gui** before the **-RRen**: **coluguiren**. (Hein-Rajam, 2013, p. 165)

And since the reader may wonder what the basis may be for Rhenius’ negative judgment (mixed of course with compliments) on Ziegenbalg and on Beschi, I shall simply remark, providing a single example (but one could provide several, both for Beschi and for Ziegenbalg) that Ziegenbalg, following Proença (who is reproduced here in Figure 6, cf. infra), and followed by Walther (who does the same on page 7, line 6 of his 1739 book) seems to believe that the normal form of the verb for which I have provided an extract of the MTL entry in Figure 7 (see below) is கெழுகிறது [KEL^aK^aKIRATU], whereas every standard Tamil source considers that the root is கேள் [KĒḶ]. It is therefore not surprising that when Ziegenbalg’s 1738 grammar was translated into English in 2010, the text which was translated was a modified (or corrected) text, whereas what must have appeared as a glaring mistake was transferred to a footnote, as will be clear while comparing BZ’s text (from Figure 8) with the published translation:

- (13a) ந ன னா க க கெ ழு கி ரு து
 Nannāga kélkiradu
 Bene aufcultare
 (Ziegenbalg, 1716, p.12)
 (13b) நன்றாக கேட்பது {{Footnote: நன்னாககெழுகிறது}}
 nanṇāka kēṭpatu
 To listen well
 (Daniel Jeyaraj, 2010, p.48)

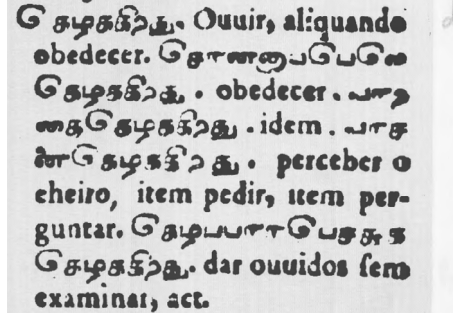


Figure 7 : MTL (p.1096), KĒḻ to hear



Figure 8 : Ziegenbalg, 1716, p.12

10 IN LIEU OF A CONCLUSION (OR AS AN “À SUIVRE”)

All this is of course inconclusive, which is not surprising because the basis is a work in progress. I hope to have made it clear that a careful reading of the texts to which I refer collectively as the GRAMMATICI TAMULICI, and which I would like to make accessible (with the collaboration of others) as faithfully as possible, as an electronic corpus, can be rewarding from many points of view, the most obvious one being the history of Descriptive Linguistics. If we stay detached from the desire to reach an impossible perfection (while recognizing its obvious strength, across history), but give equal attention to the richness of the NON-STANDARD language of which we find the (now deprecated) traces at every page, while simultaneously appreciating for itself the beauty of the ideal cultivated by Tamil poets, it seems to me that we could transmit to future generations a more faithful (or realistic) vision of a fascinating and long-lasting linguistic (soul-searching) series of adventures, from the point of view of eternity.

BIBLIOGRAPHIE

Further readings

- Beschi 1806. English translation of Beschi 1738 [ms. 1728]. See Horst, 1806.
- 1813. Republication of Beschi 1738, by the college at Fort St George. [I have not seen this book. My statement is based on Mahon (1848, p.vi)].
- 1843. *Grammatica Latino-Tamulica, in qua de Vulgari Tamulicæ Linguae Idiomatice செந்தமிழ் [KOTUNTAMIL] dicto fusius tractatur.* Auctore P. Constantio-Josepho Beschio, Societatis JESU, In Regione Madurensi, Apud Indos Orientales, Missionario. NOVA EDITIO, cum notis, et compendio grammaticæ de elegantiori dialecto

- செந்தமிழ் [CENTAMIL] dicta, ab uno missionario apostolico congregationis missionum ad exteros. PUDUCHERII, e typographio missionarium apostolicorum dictæ congregationis.
- 1848. English translation of Beschi 1738. See Mahon, 1848.
- Brentjes, Burchard & Gallus, Karl, 1985. *Grammatica Damulica von Bartholomaeus Ziegenbalg, Halle 1716*, Herausgegeben von —, Martin-Luther-Universität Halle-Wittenberg Wissenschaftliche Beiträge 1985/44 (I 32), Halle (Saale) 1985.
- Henriques, *Arte da Lingua Malabar*. See Vermeer, 1982 and See Hein Rajam, 2013.
- James, Gregory, 2000. *Col-porul. A History of Tamil Dictionaries*, Cre-A, Chennai.
- 2009. “Aspects of the structure of entries in the earliest missionary dictionary of Tamil”, pp. 273-301, Zwartjes, O. and Arzápalo, R. & T. Smith-Stark, T. (eds), Amsterdam.
- Meillet, Antoine, (1928¹, 1933³), 1976 (réimpression). *Esquisse d’une Histoire de la langue latine*, Editions Klincksieck, Paris.
- Muru, Cristina, Muru, Cristina, 2014a. “Review of Hein & Rajam [2013]”, *Histoire Épistémologie Langage* 36/2, p. 184-188.
- 2014b. “Gaspar de Aguilar: A Banished Genius”, in Amaladass, A. and Zupanov, I. G. (Ed.), *Intercultural Encounter and the Jesuit Mission in South Asia (16th-18th Centuries)*, Asian Trading Corporation, Bangalore.
- Murdoch, John, 1968 [1865]. *Classified Catalogue of Tamil Printed Books*, with introductory notices, (Reprinted with a number of Appendices and Supplement), Tamil Development and Research Council, Government of Tamil Nadu, Chennai [Original edition was printed in 1865 by The Christian Vernacular Education Society, Vepery, Madras. Murdoch [1865, p. xxxiv] wrongly supposes that Beschi’s *Grammatica Latino-Tamulica* was printed in 1739, which is in fact the date when Walther’s *Observationes* were printed and joined to Beschi’s grammar as an Appendix. That initiative made Beschi “very unhappy”, according to Jeyaraj (2010: 165).].
- Ziegenbalg, 1985. see Brentjes & Gallus 1985.
- 2010. see Jeyaraj 2010.
- Auroux, Sylvain, 1994. *La révolution technologique de la grammatisation*, Liège, Mardaga.
- Beschi, 1738 [ms. 1728]. *Grammatica Latino-Tamulica, ubi de Vulgari Tamulicæ Linguae Idiomate கொடுந்தமிழ் [KOTUNTAMIL] dicto, ad Ufum Missionariorum Soc. Iesu. Auctore P. Constantio Iosepho Beschio, Ejusdem Societ. In Regno Madurenfi Missionario. A.D. MDCCXXVIII. Trangambariæ, Typis Millionis Danicæ, MDCCXXXIIX.*
- Chevillard, Jean-Luc, 2015. “The challenge of bi-directional translation as experienced by the first European missionary grammarians and lexicographers of Tamil”, Aussant, Émilie (ed.), *La Traduction dans l’Histoire des Idées Linguistiques*, Préface de Sylvain Auroux, Librairie Orientaliste Paul Geuthner, Paris, p. 111-130.
- Hein, Jeanne (†) and V.S. Rajam, 2013. *The Earliest Missionary Grammar of Tamil. Fr. Henriques’ Arte da Lingua Malabar: Translation, History and Analysis*. Harvard Oriental Series (v. 76), Harvard University Press, Cambridge, Massachusetts and London, England.
- Horst, Christopher Henry (translator), 1806¹ (1831²). *A Grammar of the Common Dialect of the Tamulian Language, Called Kotuntamil, composed by R.F. Const. Joseph Beschi, Jesuit Missionary, after a study of Thirty years*, translated by —, Vepery Mission Press. [I have not seen the 1806 (First) edition of this book (Vepery Mission Press) but it is referenced in Google Books (<https://books.google.fr/books?id=OkfsPgAACAAJ>), although NOT currently accessible for reading, and some copies are known to exist. Christoph Heinrich Horst [1761-1810] (also known as Christopher Henry Horst) was born in Ratzeburg (Germany), came to India in 1787 and died in Thanjavur. The second edition of his translation of Beschi’s grammar was printed by the Christian Knowledge Society, according to Murdoch (1865, p. xxxiv-xxxv). A third edition was considered, but Mahon, who should have been in charge of revising Horst translation, was not at all satisfied with it and decided to make his own fresh translation, which came out in 1848 (see THIS bibliography)].

- Jeyaraj, Daniel, 2010. *Tamil Language for Europeans: Ziegenbalg's Grammatica Damulica (1716)*. Translated from Latin and Tamil. Annotated and Commented by —. Harrassowitz Verlag, Wiesbaden.
- Lallot Jean, Rosier-Catach Irène, 2005. «Le devenir d'un merveilleux outil», *Histoire Épistémologie Langage* 27/1, p. 7–10.
- Mahon, George Wiliam, 1848. *A Grammar of the common dialect of the Tamul Language, called கொடுந்தமிழ் [KOTUNTAMIL]*, composed for the use of the Missionaries of the Society of Jesus, by Constantius Joseph Beschi, Missionary of the said Society in the district of Madura, Translated from the original Latin by —A.M., Garrison Chaplain, Fort St. George, Madras, and late fellow of Pembroke College, Oxford. Madras. Madras. Printed by Reuben Twigg, at the Christian Knowledge Society's Press, Vepery.
- MTL: Madras Tamil Lexicon (1982 [reprint]). Tamil Lexicon, Published under the authority of the University of Madras, 6 volumes and 1 supplement [original publication date: 1924-1939].
- Piṅkalam: Piṅkalantai eṇṇuṁ Piṅkala Nikaṇṭu, 1968. Kaḷaka Veḷiyiṭu 1315.
- Proença, Antão de, 1679. See Thani Nayagam, 1966.
- Rhenius C.T.E., 1836. *A Grammar of the Tamil Language*, Madras, Printed at the Church Mission Press.
- Thani Nayagam, Xavier S., 1966. *Antão de Proença's Tamil-Portuguese Dictionary A.D. 1679*, Prepared for Publication by —, Department of Indian Studies, University of Malaya, Kuala Lumpur, Sale Agents: E.J. Brill, Leiden (Netherlands).
- Vermeer, Hans J., 1982. *The first European Tamil Grammar*, A Critical edition by —, English version by Angelika Morath, Julius Groos Verlag, Heidelberg.
- Walther, 1739. *Observationes Grammaticae, quibus Linguae Tamulicae idioma Vulgare, in usum operariorum in Messe Domini inter Gentes Vulgo Malabares Dictas, Illustratur a Christophoro Theodosio Walthero, Missionario Danico, Trangambariae, Typis Missiois Regiae, MDCCXXXIX*. (The book is freely available on Google books at: "http://books.google.fr/books/about/Observationes_grammaticae_quibus_linguae.html?id=IApUAAAACAAJ").
- Ziegenbalg, Bartholomaeus, 1716. *Grammatical Damulica, / quae / per varia paradigmata, regulas & necessarium vocabulorum apparatus, / viam brevissimam / monstat, / qua / lingua damulica / seu malabarica, quae inter Indos Orientales in / usu est, & hucusque in Europa incognita fuit, / facile disci possit : / in / usum eorum / qui hoc tempore gentes illas ab idolatria ad cultum veri / Dei salutemque aeternam Evangelio Christi per / ducere cupiunt : / In itinere Europaeo, seu in nave Danica, / concinnata / a / Bartholomaeo Ziegenbalg, / Serenissimi Regis Daniae Missionario inter Indos Orientales, & eccle / siae ex Indis collectae Praeposito. / Halae Saxonum / Litteris & impensis Orphanotropei M D CC XVI*.
- Zwartjes, O. and Arzápalo, R. & T. Smith-Stark, T. (Eds), 2009. *Missionary linguistics IV: Lexicography. Selected papers from the Fifth International Conference on Missionary Linguistics, Mérida, Yucatán, 14-17 March 2007*. Studies in the History of the Language Sciences, 114, John Benjamins Publishing Company, Amsterdam.