

John Benjamins Publishing Company



This is a contribution from *History of Linguistics 2021. Selected papers from the 15th International Conference on the History of the Language Sciences (ICHoLS 15), Milan, 28 August – 1 September.*

Edited by Savina Raynaud, Maria Paola Tenchini and Enrica Galazzi.

© 2024. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

How far are the horizons of descriptive linguistics?

Jean-Luc Chevillard

CNRS, Laboratoire HTL (UMR7597)

European missionaries engaged in the linguistic description of 16th–18th c. Tamil Nadu discussed two possible attitudes. Proença (1625–1666), who wrote in Portuguese, thought that the most important task was to concentrate on what is useful *pera a pratica* “for practical purposes” and wanted to master ordinary language, both (A) in its colloquial forms — including substandard and dialectal variants — and (B) in its more standardized form. Beschi (1680–1747), who wrote in Latin, thought he could become influential by mastering (C) *Centamil*, the poetic “more elegant” dialect, cultivated for many centuries by traditional grammarians and poets of Tamil Nadu. This article evokes the strategies of Proença, Beschi and others, who navigated the components of Tamil “triglossia”, in which both (A) and (C) can coexist with (B), but not simultaneously. Either (C) is ignored, being considered as “useless for practical purposes”, or (A) is shunned, being considered as “barbaric”.

Keywords: Tamil, diglossia, triglossia, Beschi, Proença, traditional language scholarship, language of poetry, ordinary language, standardized language, missionary grammars

Chaque époque rêve la suivante
Michelet

Je suis belle, ô mortels! comme un rêve de pierre,
Et mon sein, où chacun s'est meurtri tour à tour,
Est fait pour inspirer au poète un amour
Eternel et muet ainsi que la matière.
Baudelaire

Parce qu'il est limité, l'acte de savoir possède par définition
une épaisseur temporelle, un horizon de rétrospection,
aussi bien qu'un horizon de projection.
Sylvain Auroux (1989: 13)

Calculus
Leibniz

1. Accessibility of a forgotten ancient new world of linguistic descriptions

A forgotten corpus of linguistic descriptions, not composed in Modern English and dealing with many world languages as those languages existed several centuries ago, is in the process of slowly re-emerging, in front of our eyes, thanks to the concerted efforts of a number of historians of linguistics. This changing situation in the case of Tamil is exemplified by the six pairs, associating an author's name with an acronym or a title, which are visible inside Figure 1, and numbered, from ① to ⑥.

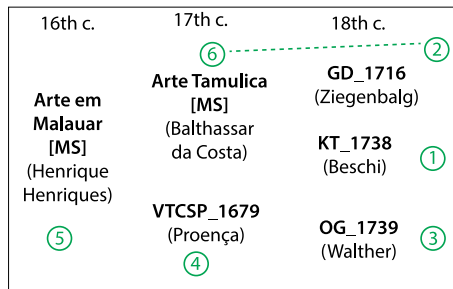


Figure 1. Foreign explorers of Tamil, in the 16th, 17th and 18th centuries, from most accessible to least accessible

The numbers which are on the sides of the six pairs attempt to rank, in a rough manner, with a bias towards our time, the **accessibility** of the corresponding texts for the interested readership during the time span separating the date of their composition from our present time. This means that the text ranked as last,

Table 1. Accessibility features of the six works appearing in Figure 1

Most to least accessible	①	②	③	④	⑤	⑥
Author	Beschi	Ziegenbalg	Walther	Proença	Henriques	Da Costa
Grammar or dictionary?	Gram.	Gram.	Gram.	Dic.	Gram.	Gram.
Metalanguage	Latin	Latin	Latin	Portuguese	Portuguese	Portuguese
Original printing or MS?	1738	1716	1739	1679	MS. (16th c.)	MS. (17th c.)
Known copies	several	several	several	one (Vatican)	one (Lisboa)	five
Subsequent reprinting	19th c.	1985	None	1966 Facsimilé		
Printing of original MS					Vermeer 1982	Muru 2022
English translation	Horst 1806 & 1831 ² , superseded by Mahon 1848	Jeyaraj 2010	None	None	Hein & Rajam 2013	Muru 2022

i.e. ⑥, was made more easily accessible to its potential readers¹ only in 2022, both in the original Portuguese and as an English translation, thanks to Cristina Muru, after being inaccessible for roughly 300 years whereas the text ranked as ① has almost always been accessible to those readers who were reasonably informed and sufficiently motivated, although we must add that the majority of those readers, since the beginning of the 19th century, most probably read it in English translation rather than in the Latin original, as we shall see later. Regarding the other texts, among which two, namely ② and ③, are in Latin and the remaining two, namely ⑤ and ④, are in Portuguese, my ranking is of course highly subjective, being based on my belief that Latin has been a scientific *Lingua Franca* among scholars for a longer period than Portuguese, and also on the idea that a manuscript such as ⑤ was less susceptible to reach an audience than printed books

1. There were, of course, also readers in earlier periods, as attested for instance by the dotted line connecting ⑥ and ② inside Figure 1, which alludes to a probable case of plagiarism, of Da Costa’s manuscript grammar by Ziegenbalg (1716), which is discussed in Jeyaraj (2010) and in Muru (2022).

such as ②, ③ and ④. I shall discuss in this presentation only a subset of the linguistic technical literature about Tamil which has been evoked inside the introductory section, including for instance the two works, to which the ranks given are ③ and ④. I shall also remark, seemingly contradicting myself, that those ranks might quickly become outdated, because the 20th century point of view is being ineluctably phased out, thanks to the pioneering works of Vermeer (1982), of Hein & Rajam (2013), of Jeyaraj (2010) and of Muru (2022), who have respectively made items ⑤, ② and ⑥ potentially as accessible today to modern linguists as Beschi's ① has been since the first half of the 19th century, thanks to his English translators (see Figure 4). On the other hand, a lot of efforts are still required if one wants the same to be true for the VTCSP_1679, alias ④, alias *Vocabulario Tamulico com a Significaçam Portuguesa* (VTCSP), composed by Proença (1625–1666) and for the OG_1739, alias ③, composed by Walther (1699–1741), and this will be part of our focus here, because I have been engaged for now ten years in the preparation of an electronic edition of the first one (i.e. the VTCSP) and of an English translation of the second one.

2. There is more than one variety of Tamil

Before going more deeply into an exposition of the specificities and the difficulties of various texts, I must first of all provide the reader with more information concerning the situation of the Tamil linguistic universe, and especially with what I have called elsewhere² the Tamil Triglossia, expanding on the more familiar diglossia. That triglossia will occasionally be represented by the diagram , for which an expanded form – completed by a European analogy, for the sake of clarification – is provided inside Figure 2.

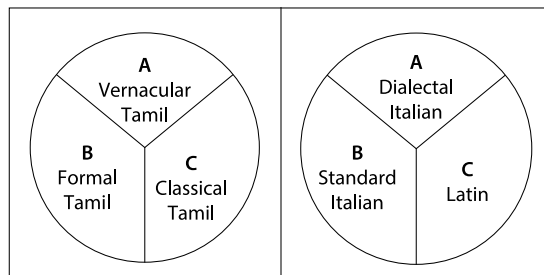


Figure 2. Comparing the linguistic situation of Tamil Nadu and the linguistic situation of Italy as two examples of “triglossia”

2. See Chevillard (2021).

Being aware of the existence of the Tamil triglossia is necessary indeed if one wants to understand the differences between what the three missionaries mentioned in Figure 3 (“Changing Horizons”) were trying to do.

As already discussed elsewhere, by others and by me, the first grammatical description of Tamil made by a foreigner, which is the item referred to until now as ⑤, composed by Henrique Henriques (1520–1600), henceforth HH, in the 16th century, rediscovered in Lisboa by Thani Nayagam in the 20th century,³ critically edited by Vermeer in 1982 and translated into English by Hein & Rajam (HOS, 2013), was primarily a description of a dialect of Tamil spoken on the “Fishery coast” (Costa de Pescaria), although it also incorporated forms taken from “the language of the educated” (Vermeer 1982:xx). To give a concrete example, HH makes the following two statements:

- (1a) *Colliren que he por matar faz col e tãbē collu* “[The verb with citation form] *Colliren* ‘I kill’, which corresponds to *matar* ‘to kill’ becomes [in the imperative] *col* ‘kill!’ or *collu* ‘kill!’ ” (Vermeer 1982: 82) [My translation]
- (1b) *Nos verbos desta conjugação os que muito sabē os pronosiaō muitas vezes cō gui antes do Ren: coluguiren* “Regarding the verbs of this conjugation, those who know much (*os que muito sabē*) frequently utter them (*os pronosiaō*) [i.e. the present forms] with a *gui* before *Ren*, as *coluguiren*.⁴ (Vermeer 1982: 82) [My translation]

In these examples, the forms *colliren* “I kill” and *collu* “kill!” belong to one dialect of what I call the A-variety of Tamil (Vernacular Tamil), whereas the forms *coluguiren* “I kill” and *col* “kill!” are their counterparts in the B-variety of Tamil, i.e. “Formal Tamil”, which was a standard used in formal speech occasions — see the use of *os pronosiaō* “utter them” by HH — and also in writing. Concerning the special writing mode which we call Printing, I shall add that the B-variety is indeed the linguistic variety seen in another work composed by Henriques, namely the *Flos Sanctorum*, alias *aṭiyār varalāru* “life of the Saints”, printed in 1586 on the Fishery Coast (Henriques 1967 [1586]). Also concerning writing and how it differs from ordinary speech, if we move in time up to the 18th century and now turn to Walther’s *Observationes Grammaticae* (i.e. OG_1739 in Figure 1), we can see him declare in 1739:

3. See Thani Nayagam (1954).

4. *Colliren* “I kill” is the citation for one verb, which is also the model chosen by HH for the conjugation of a group of verbs (*5a conjugação, 1a maneira*). This explains the use of the plural “os”, because what is said of *Colliren* is also true of the other verbs which are conjugated like it. It should also be noted that there is an inconsistency in the MS in the use of *Ren* and *ren*, in this sentence. Capital R and lower-case r are tentatively used for two distinct phonemes — written *ṛ* [ṛ] and *ṛ* [r] — but the scribe failed to maintain the distinction here.

- (2) *Loquimur cum vulgo, fed scribamus cum doctis* “we talk with the common people but with the learned we write” (Caput 3, § 7, 13). (Walther, OG_1739, p. 34)

and this will be familiar to everyone who has experienced the modern Tamil diglossia.

3. The Tamil linguistic hierarchy

Going beyond the ordinary diglossia, I shall now provide some explanations concerning the third level in what I have called the Triglossia and the central importance of level C, which I have referred to as “Classical Tamil” inside Figure 2, where I compare its role with the role of Latin in Europe.

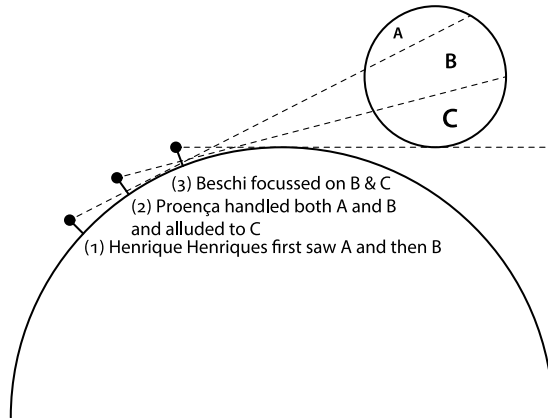


Figure 3. Changing Horizons. Three Jesuit missionaries progressively discover the Tamil triglossia, including Classical Tamil, in the course of two centuries

As a support for this statement, I shall provide the reader with an extract from the preface which appears inside a Paris BnF MS (indien 227) of Beschi’s *Caturakarāti* “Quadruple dictionary”, which is hand-written but has a title page dated 1732 which looks like a lithograph. In this passage, the 42-year old Constantius Joseph Beschi [1680–1711–1747]⁵ explains why it is extremely important to study Classical Tamil, which he refers to as the *elegantius idioma* “more elegant dialect”.

5. The middle date is the date of his arrival in India.

(3a) *quâ ipsis utilitate futurum fit, facile noverint // omnes, fi paulisper animadvert-
erint in his regionibus, // monumenta Deorum, Fabularum figmenta, scientiarũ
// præcepta, Poëtarum carmina, astronomiæ calculos // medicinæ leges,
musices, choreæq.e regulas, omnia deniq. // vel ipsa prima Grammaticæ rudi-
menta, elegantiori // hoc idiomate ab antiquis scripta fuisse*

(Beschi, *Caturakarāti*, 1732, Præfatio, BnF indien 227)

(3b) “and how useful it will be to them, everyone will easily know, if they have
observed for a moment how in these regions the scriptures of the Gods, the
fictions of their fables, the rules of science, the poems of the poets, the calcula-
tions of astronomy, the rules of medicine, the rules of music and dance, every-
thing, and even the first rudiments of grammar, were written by the ancients
in this more elegant register.”

[translation excerpted from Trento & Chevillard (forthcoming)]

I must add to this, in order to give a more precise idea of Beschi’s view of the
world, that he had become so engrossed in his exploration of the Tamil linguis-
tic universe that he wrote four distinct Tamil grammars, three in Latin and one
in Tamil, in addition to his lexicographical efforts, some of which have just been
evoked (See 3a and 3b). Among these grammars,

- the first grammar was written in Latin in 1728 and later printed in 1738 by
the protestants in Tranquebar. It is mentioned here as KT_1738 in Figure 1,
where the acronym KT stands for *Koṭun-Tamiḷ* “Rough Tamil” [lit. “Crooked
Tamil”], in order to distinguish it from the second grammar, and a small cita-
tion appears in (7a) and (7b)].
- the second grammar contains a description of *Cen-Tamiḷ* “perfect Tamil” [litt.
“Straight Tamil”].⁶ Beschi wrote it shortly afterwards, its preface being dated
1730.⁷ A small citation appears in (8a) and (8b).
- the third grammar is in Tamil and is called தொன்னூல் விளக்கம் [*tonṇūl
viḷakkam*], “Light/Lamp on the ancient treatises” (Beschi 1838).
- the fourth one, called *Clavis*, is a compendium, written in Latin, of all the five
parts of traditional Tamil grammar (including poetical conventions, metrics
and rethorical devices).⁸

6. “Perfect (or Straight) Tamil” and “Crooked Tamil” are my translations for *Centamiḷ* and *Koṭuntamiḷ*. The adjectives used by Beschi are “elegans” and “asperum”. See KT_1738, p.4: “Ac fi illud elegans Tamulicum idioma dicerent, hoc asperum”.

7. The MS of this grammar, written in Latin, was published as a book only in the 20th century, by Besse, in 1917, but an English translation, by Babington, had appeared in 1822.

8. The MS of Beschi’s *Clavis* was published thanks to A.C. Burnell in 1876.

I must add that when Beschi made use of the terms *koṭuntamiḷ* and *centamiḷ*, in the titles of his 1728 and 1730 grammars (resp. Beschi 1738 and Beschi 1974 [1822]), in order to characterize these two varieties⁹ in the preface of his KT_1738 book, he was not using ordinary terminology but borrowing two terms belonging to the poetical high language, i.e. to *Centamiḷ*.¹⁰ If one looks for those terms inside Proença's VTCSP_1679, which is the oldest extant Tamil-Portuguese dictionary, one does not find them.¹¹ Among the terms which are found inside the VTCSP, one can mention:

- (4) **Koccai.** Palaura, ruftica, baixa. “Language which is low [and used by] rustic [people]” (VTCSP, entry 297_L_j)¹²
- (5) **Koccaiyan.** H. ruftico, barbaro no falar “Man who is rustic and speaks in a barbaric manner”. (VTCSP, entry 297_L_k)

What the wording of such entries from Proença's 17th c. VTCSP reflects, which will be taken to a higher level by Beschi in the 18th century, is the growing adoption, by missionaries, of the point of view of traditional Tamil scholars. Tamil grammarians categorize the language which they describe under two heads, which they call *Ceyyul* “poetry [lit. (poetical) composition]” and *Valakku* “(ordinary) usage”. However, as stated in the last section of the *Tolkāppiyam*, the most ancient Classical Tamil grammatical treatise, probably dating back to the first half of the first millenium, not everyone's *valakku* is fit to be called *valakku*:

9. Those two terms, although they have stuck to the name of Beschi, in the course of history, because of the high visibility of book titles, were in fact **not** his preferred terms. See below (7a), (7b), (8a) and (8b) for more precision on Beschi's terminological preferences. One could also debate whether it was appropriate to make use of the expression *koṭun-tamiḷ* for referring to the B-variety, but that would take us too far. I shall simply say here that, before the missionary age, the terminological opposition between *cen-tamiḷ* and *koṭun-tamiḷ* has its roots in the attempted description of the regional variety of Tamil by traditional grammarians, concomitant with the codification of a poetical language, common to the users of all dialects, in a large area which encompasses both Tamil Nadu and modern Kerala. See Chevillard (2008).

10. The term *cen-tamiḷ* is as old as the *Tolkāppiyam* (1st half of 1st millenium). However, the oldest attestation of *koṭun-tamiḷ* seems to be found in the *Nūrkatturai* (approx. “postface”) to the *Cilappatikāram*, where it is paired with *centamiḷ*. *Koṭun-tamiḷ* is also found in medieval Tamil grammatical commentaries, as part of discussions on the concept of *tikai-c col* “regional word”. See previous footnote.

11. They are, however, found inside Beschi's *Caturakarāti*, already presented.

12. The coordinate system which I use for identifying entries inside the VTCSP is explained in Chevillard (2021).

- (6a) *valakku* enappaṭuvatu* uyarntōr mērru* -ē // nikaḷcci avarkattu* ākalāna*
(TP638i)¹³
- (6b) “That which is to be called ‘usage’ (*valakku*) rests on the Superior (i.e. learned) people, due to the fact that (significant) (language) events happen by them”
[my translation]

We shall now turn to Beschi’s 1728 preface to his first grammar, where he explains for the first time, the difference between the two language varieties which are the respective targets of this first grammar and of the second grammar which he was about to write, and would complete in 1730. Thus, he explains:

- (7a) *Duplex in hâc regione Tamulicæ Linguae idioma est: sublimem dixerim unam, communem alteram. Aliqui non satis appositè Poeticam vocant, quæ à communi recedit. Attamen, cum eâ linguâ Tamulenses non tam in his, quæ metro ligantur, quàm in cæteris omnibus, quæ ab antiquis hujus linguae peritis solutâ quoque oratione conscripta sunt, usos esse videamus, quod præcipuè in commentariis poëtarum patet: ea satius quam poëtica lingua, elegantior vel sublimis vocabitur.*
(Beschi 1728 [1738]: 3–4)
- (7b) “In this region there are two dialects of the Tamul Language: I would call one the High, the other the Common. Some not very correctly call that which differs from the Common, the Poetical dialect. But since we see the Tamulians use that dialect, not so much in those writings which have the trammels of meter, as in all others which by the old authors skilled in this tongue are composed in prose also, which is especially to be seen in the commentaries of the poets, that dialect will be better named the more elegant, or high, than the poetic.”
(translation by Mahon, Beschi 1848: 2)

It could be interesting to determine whom the 38-year old Beschi is disagreeing with when he advocates the terminological use of “*elegantior vel sublimis*”, rather than “*poëtica lingua*”. However, remaining focussed on our core topic, I shall now provide the reader with an extract from the 1730 grammar preface, where Beschi seems to have somehow changed his mind, because he declares:

- (8a) *Cum vero quæ hoc idiomate scripsere Tamulenses fere omnia metro ligata sint, ne indecorum censeatis, quæso ; si vos ad profanos poetas adducere et ad poeseos studium advocare præsumo. Plura in D. Hieronymum acrius scripserint olim criminantes, quasi adductis a Poetarum codicibus exemplis, candorem Ecclesiæ ethnicorum sordibus pollueret, quibus eruditissime respondens D. Hieronymus, manifeste ostendit et apostolum Paulum passim in suis epistolis poetarum carmina adduxisse, et [...]*
(Beschi 1917 [ms. 1730]: xi)

13. *Tolkāppiyam, Poruḷatikāram, Cūttiram* 638 (in the numbering of the *Ḥampūraṇam*). See *Tolkāppiyam, Poruḷatikāram, Ḥampūraṇam*, 2003.

- (8b) “But since almost all the Tamil works in this dialect are in verse, I trust you will not deem it improper, if I venture to draw your attention to heathen poets, and to the study of poetry. In former times, St. Jerome was severely censured for having, by the introduction of examples from the poets, sullied the purity of the church with the pollution of the heathen. St Jerome, in his learned reply, demonstrates, that the apostle Paul repeatedly cites from the poets, in his epistles, and that [...]” (translation by Babington 1822, in Beschi 1974 [1822]: viii)

To wrap up our description of Beschi’s activities in this field, I shall add that in addition to his composition of four grammars, enumerated earlier in this section, and to his compilation of dictionaries, multilingual and (Tamil) monolingual — see the *Caturakarāṭi* already mentioned in (3a) and (3b), modelled after the traditional poetical vocabularies used by traditional Tamil Scholars — he had also become himself a Tamil poet, known under the Tamil name of வீரமாமுனிவர் [*vīramāmuṇivar*]¹⁴ and had composed a long epic poem in Poetical Tamil, the *Tēmpāvaṇi*, in honour of Saint Joseph. This was the culminating point in a gradual change of horizon, visually represented in Figure 2, where the slow rising of C, i.e. of the Classical Tamil Sun is shown. That rising Sun shone on the stunning fact that a jesuit who had come from Italy to Tamil Nadu in order to be a missionary was himself converted to a form of devotion towards the high variety of the Tamil language and performed very significant தொண்டு [*toṇṭu*] “service” for it, becoming himself a Tamil Poet and having today his statue near the Marina beach in Chennai, all that happening “A. M. D. G.”¹⁵ of course.

4. From the age of AB to the age of BC, and back: “Calcuemus!”

Having finished this short evocation of one of the paths which was followed by some of the early language explorers who lived in the 16th, 17th and 18th centuries, we shall now return to the 21st century and to the **accessibility problem** (for modern researchers) which we first discussed in the introductory section. Our new window of observation, for visualising the discontinuities and continuities in transmission, is represented by Figure 4, in which three of the items previously mentioned are represented with the addition of three elements which stand as elements in a chain between the past and the present.

14. For an overview of Beschi’s career in South-India, see Ebeling & Trento (2018).

15. “Ad maiorem Dei gloriam”.

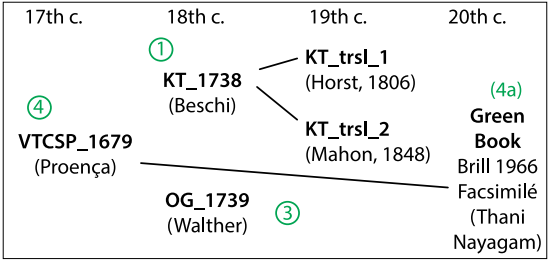


Figure 4. Crossing time, while exploring Tamil as a foreigner in the 17th, 18th and 19th centuries

Two of those items are 19th century translations of item ① from Latin into English, made respectively in 1806 and 1848, the latter one having been quoted in (7b). The third item, which appears in the diagram as (4a), alias “Green Book”,¹⁶ is the 1966 facsimilé edition of the now familiar 1679 *Vocabulario* (or VTCSP). It contains an introductory part in Portuguese — for which the modern editor has provided an English translation. That introductory part is followed by 508 pages containing 16218 entries and we shall now evoke first the difficulties inherent in moving in time

- **from point (4)**, in 1679, which is the moment when Proença’s colleagues, probably as the result of a huge effort, printed a few copies of the posthumous VTCSP_1679, of which only a single one remains in existence today, in the Vatican Library,
- **to point (4a)** which is the time when Thani Nayagam published the 1966 facsimilé.

If we think of some of the possible reasons for gaps to occur in the transmission of collective knowledge, such as for example

- (a) when there is a change of scientific lingua franca and no translation is made
- (b) when a printed work has never been reprinted and very few copies are available
- (c) when a work has been transmitted only, or partly, as a MS
- (d) when a work is available but is not read, for various reasons (difficulty, absence of desire, absence of time, ...)

16. This refers to the well-known colour of many of the books published by Brill, including this one.

We can say that the printing in 1966 of the VTCSP facsimilé eliminates only the cause listed here as (b),¹⁷ but leaves causes (a) and (d) unchallenged. Whoever wants to access the information contained in the 1966 facsimilé¹⁸ encounters obstacles which are a combination of linguistic and logistic problems. At the level of the micro-structure, the modern reader must decode individual entries,¹⁹ where Tamil words written following an archaic ambiguous spelling and 17th century Portuguese words which are not all found in Modern dictionaries are combined. At the level of the macro-structure, the reader is confronted with an incompletely lemmatized²⁰ “dictionary”, where Tamil words are ordered in the Portuguese alphabetical order, which makes it difficult to find out whether a given Tamil word is contained in the VTCSP. More precisely, the VTCSP is a list of 16,218 Tamil word-forms (with explanations in Portuguese) which is ordered following the Portuguese alphabetical order, which we can represent by the sequence of eighteen letters “A, B, C, D, E, G, I, L, M, N, O, P, Q, R, S, T, V, X”.²¹ As a consequence, the 16,218 Tamil word-forms, printed in Tamil characters, are distributed between 18 larger sections, based on the initial letter of their virtual Portuguese transliteration. Those larger sections are further subdivided into 382 subsections, based on the letter, or the pair of letters, which follows the initial letter.²² That means

17. To this can be added the fact that during the 3 centuries separating the two dates, 1679 and 1966, some authors complained that they were not able to locate a copy of the VTCSP_1679, although they tried because they had heard of its importance. We see for instance A.C. Burnell, in his 1880 catalogue of Portuguese sources pertaining to India, referring to this book and to its possible location, but declaring that the copy could not be found. Julien Vinson nevertheless explained in a 1910 article that he had successfully located the Vatican library copy of the VTCSP. This piece of information however did not reach S. Vaiyāpuri Pillai, general editor of the Madras Tamil Lexicon (MTL), who declares in 1924 inside his introductory essay “History of Tamil Lexicography”, on p. xxxvi inside the 1st volume of the MTL, that “no trace is found” of the VTCSP (Vaiyāpuri Pillai 1982 [1924]).

18. I must mention that there was also a MS transmission of the VTCSP, but it would take us too far to discuss it in detail. See Gregory James (2000) and C. Muru (2010) for more information.

19. A typical entry in the VTCSP starts with a Tamil word-form followed by one or more Portuguese glosses, depending on the number of meanings which the explorer had managed to discover. Occasionally, Tamil sentences provided as examples are included in the entry, accompanied by Portuguese translations. There are also cross-references between some of the entries.

20. See Chevillard (2021).

21. Including variant forms for I and N, and taking into account the fact that lower case “u” is the counterpart of uppercase “V”, the list of 18 can also be written as “A, B, C, D, E, G, I/Y/J, L, M, N/Nh, O, P, Q, R, S, T, U/V, X”.

22. As an example, the largest section in the VTCSP is the Q section (or 13th section) and contains 2253 entries. It is subdivided into 44 subsections. Among these, the first 14 subsections all

that, in practice, the 1966 facsimilé of the VTCSP has remained a closed book for most modern readers, unless they have undergone a special training for reading the Tamil words, and are familiar with the Portuguese language, as it existed in the 17th century. Nevertheless, the VTCSP remains visible on the horizon, as a tantalizing possible target of research for linguists, and this is what Figure 5 is trying to suggest.

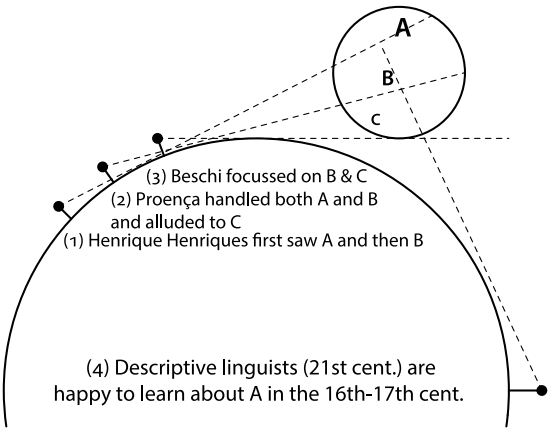


Figure 5. Doing field-work in ancient sources and rediscovering A, alias 17th century “Spoken Tamil”

One of its interesting features is that the opinions of Proença were quite different from those of Beschi, as far as poetical Tamil is concerned. This should become clear from the following quotation, to which our access has been greatly facilitated thanks to Thani Nayagam 1966 edition, and to his introduction and his translation of the preface (prepared jointly with another scholar). Taking a stand quite different from the one seen in the *Caturakarāti* preface, as shown in (3a) and (3b), the strongly worded preface of the VTCSP states that:

contain words starting with Q+A, the difference between them being the third letter. However, when the third letter is L, we are supposed to know that Tamil has “three types of L” (nowadays transliterated as l, ɭ and ʟ), and the same is true for R, of which there are “three types”, nowadays transliterated as r, ɽ and ɿ. For more details, see Chevillard (2017). Given the fact that the standard phonemic inventory of Tamil consists of 12 vowels and 18 consonants, the mental exercise required for imagining where a Tamil word, apprehended through an unprecise transliteration, will be found inside the VTCSP is rather complex, unless one has gone through the whole text, which is now the case for me, after several years of efforts, although the final form which my electronic edition of the VTCSP will take is not yet completely decided.

- (9a) ...Deixa- // =rey tambem as palauras poeticas (de que o vocabulario do P. Igna- // =cio Bruno eſtâ cheo) por que não feruem nada, pera â pratica, ov // profa. e dellas tem os poetas Tamuis ſeus vocabularios, â onde os // curioſos, que quizerem compor verſos, as podem ver...
- (Proença, 1679, https://digi.vatlib.it/view/MSS_Borg.ind.12/17)
- (9b) ... I shall leave out also poetic words (of which the vocabulary of Fr. Ignacio Bruno is full) because they are useless for practical purposes or for prose, and the Tamil poets have their own vocabularies for them, in which those who are interested and might wish to compose verses may find them...
- (translation Edgar C. Knowlton & Xavier S. Thani Nayagam, Thani Nayagam 1966: 13)

A direct examination of the VTCSP shows that Proença made nevertheless a few exceptions to his stated principles, because several of his entries contain the mention *palaura de uerſo* (“word used in poetry”).²³

5. How far is the horizon of descriptive linguistics?

We have now reached the final section of this short presentation and it is now time to explain the title chosen, which is the continuation of a conversation, started with Sylvain Auroux, about the two horizons,²⁴ and continued with Christian Puech and Valérie Raby, when HEL XXXIII-2, “Horizons de rétropection”, was under preparation (Puech & Raby 2011). I have always hesitated between two possible uses of the “horizon” metaphor, wondering whether the most important feature is to constantly tell ourselves that we do *not* know what lies behind the horizon, either because it has been forgotten or erased, or because we do not yet know what lies ahead. I was, at the time when the “Horizons de projection” workshop was planned engaged in a long-running task, which consisted in entering, in XML format, the content of the VTCSP, and I did not yet know how many entries it contained, although I now know that there are 16,218,²⁵ but am at the same time

23. See for instance the following entries: *peñfir* (217_R_m) “Molher, honori- // =fice, hê palaura de uerso.”, *pāni* (211_R_d) “Agoa, palaura de uerso.”, and other similar ones.

24. That conversation is present in the choice of a title (South-Indian Horizons) for the Felicitation Volume in honour of François Gros which I have edited in 2004, with the collaboration of Eva Wilden as associate editor (Chevallard & Wilden 2004).

25. Although this kind of information appears simple, Thani Nayagam who published the facsimilé in 1966 wrote in his “Preface” (p.9) that “There are altogether 16,456 entries separate main entries against which meanings are given in Portuguese or Latin”. The figure 16,456, which is too high by 238, is probably the result of an estimation but has been repeated by several authors. Similarly, Vinson wrote in 1910 that the main body of the VTCSP consists of “247 feuil-

aware of the fact that most readers do not enjoy “drowning by numbers”. Therefore, I shall in this brief conclusion simply call attention to the fact that much remains to be discovered inside the VTCSP, as it can progressively reveal itself to a larger audience, by being made readable, progressively overcoming the handicaps evoked in the previous section. Among the possible targets, I shall mention here

- The presence of many dialectal forms,²⁶ which the compilers of later dictionaries eliminated.
- The presence of many pairs of entries, linked by a cross-reference mechanism, one of them being the formal (B-variety) Tamil form and the other one being the spontaneous (A-variety) form, deprecated by later dictionaries, untill “Spoken Tamil” started to be reapprciated, in the 20th century, when descriptive linguists undertook to dethrone normative grammarians.
- The wealth of information on the morphology and the syntax of early modern Tamil.²⁷
- The presence of many taboo terms, which are not found in any existing dictionary of Tamil, because of the statement found in the *Tolkāppiyam*: *avaiy-al kiḷavi maraiṭṭaṇar kiḷattal* (TC442c)²⁸ “expressions which do not belong in the assembly, should be uttered in a dissimulated manner”. This probably explains why, when explaining many of the 96 such expressions found in the VTCSP, Proença switches from Portuguese to Latin,²⁹ also adding as a warning to the users of his *Vocabulario*, one label taken from the set {*modesté, modestissimé, turpe, turpissime*}.³⁰

lets numérotés au recto” (Vinson 1910). This is however incorrect because there are in fact 254 “feuillets”. The mistake is due to the fact that wrong folio numbers have been printed on most folios and that, as a consequence, only 16 folios are reliably numbered, on a total of 254. The last folio, which should have been numbered as 254, was numbered as 247.

26. See for instance entry 231_R_m: “*piṛā*. pomba, ou rola. *kāṭṭuppiṛā*. propriamente rola.” The standard form would be *puṛā* “pigeon”, as seen in the MTL. As per my XML-encoded data-basis, there are hundreds of such examples.

27. See for instance Chevillard (2021), which deals with a set of 298 entries, in which we see the earliest clear attestation of the use of vector verbs in compound verbal expressions in Tamil.

28. *Tolkāppiyam*, *Collatikāram*, *cūttiram* 442, as per the numbering in the *Cēṇāvaraiyam* (*Tolkāppiyam*, *Collatikāram*, *Cēṇāvaraiyam*, 2003).

29. Pupils in the Lycée Français de Pondichéry once told me, in 1982, that when they say: “nous parlons latin”, it means in fact “nous disons des gros mots”.

30. James (2007 and 2009) discusses some of the most extreme techniques, such as partial encryption, which are used in the VTCSP for dealing with some the topics labelled as “*turpissime*”.

Regarding this last point, I shall offer as a final illustration (see Table 3), an extract from my listing of taboo expressions, hoping that it will be sufficiently intriguing for some readers, and entice them to embark in the exploration of that most beautiful language, the triglossic Tamil, following in the footsteps of their Portuguese and Italian predecessors, Proença and Beschi.

Table 3. Extract from the listing of Taboo words in the VTCSP

TABOO: @ ஸ த து வ # # ம (102_L_b)
Voeiro, cano // de agoa, item de orina, mo- // =destè.
TABOO: @ ஸ ம (101_R_p))
Agoa, item orina, mo- // destissimé.
TABOO: @ த # வி (369_L_k))
Verenda, mulieris vetu- // =læ, roindade itē C. q̃ não presta.
TABOO: @ த # ன (369_L_n) [also categorized as {BODY_PART LITERARY}]
Hombro. {{@ த # ன @ ஸ -}} {{= ௫ கி ற து}}. habère rem. com fæ- // =mina, modeste.
TABOO: @ ந # னி (174_L_c))
Ceço, que fae pera // fora a gente grande por doen- // =ça, item fimilè in verendis fæ- // =minæ, item muito fraco e ma- // =gro.
TABOO: @ ந # ன னி (174_L_a) [also categorized as {NOT_IN_MTL}]
Quòd. In medio pro- // =minet, verenda fæminæ tur- // =piffime.
TABOO: @ ப # ஸ ஸ (237_R_p))
O traceiro; aliquantulum, modeftè.
TABOO: @ ப # ட ட (243_L_n))
M. ruim, mo- // =defètè, item apud. Br. cefto.
TABOO: @ ப # த தி (244_L_c))
Testiculi, modeftè.
TABOO: @ ப ன னு ட ம டி (217_R_l))
verenda // mulieris modestè
TABOO: @ ப லு கி ற து (217_L_o) [also categorized as {V}{META}]
Cagar, fazer came- // =ra, turpe, ன @ ற ன. லு @ வ ன.
TABOO: @ ம # கி கி ற து (135_R_d) [also categorized as {V ALIAS}{SEX}]
Quòd. {{@ ப # கி கி ற து}}. habere rem uenereā ou ter tentaçoes e eftimulos, in re uenerea; aliquando folum, pati pollutionem, // aliquando perturbarse, vacilar.

Dedication

This article is dedicated to the memory of my late colleague and friend Valérie Raby, thanks to whom it was written.

Bibliography

- Auroux, Sylvain. 1989. “Avant-propos”. *Histoire des Idées linguistiques. Tome 1. La naissance des Métalangages en Orient et en Occident* edited by Sylvain Auroux. Liège/Bruxelles: Mardaga.
- Babington, Benjamin. 1822. See Beschi (1974).
- Beschi, C. J., 1732. *Caturakarāti*. Manuscript BnF indien 227. Paris: Bibliothèque Nationale de France.
- Beschi, Constantius Josephus. 1738 [ms. 1728]. *Grammatica Latino-Tamulica, ubi de Vulgari Tamulicæ Linguæ Idiomate கொடுந்தமிழ் [KOṬUNTAMIL] dicto [...]* [See translated title in following entries dated 1831 and 1848]. Tranquebar: Typis Miſſionis Danicæ [Danish mission press].
- Beschi, Constant Joseph. 1974 [1822]. *A Grammar of the High dialect of the Tamil Language, Called செந்தமிழ் [centamil]* by Constant Joseph Beschi, translated from the original latin by Benjamin Guy Babington. Thanjavur, India: reprinted by The Tanjore Sarasvati Mahal Library.
- Beschi, Constantius Josephus. 1831² [1806¹]. *A Grammar of the Common Dialect of the Tamulian Language, Called Koṭuntamiḷ*. [translated by Christopher Henry Horst]. Madras: Vepery Mission Press.
- Beschi, 1838, [*Vīramāmuṇivar tiruvāymalarn taruḷicceyta vaintilakkaṇat*] *Toṇṇūḷ Viḷakkam [mūlamum uraiyum]*. Putuvai (Pondicherry). (IFP Library “TA GRAM 41”, Pondichéry).
- Beschi, C. J. 1848. *A Grammar of the Common Dialect of the Tamul Language, Called கொடுந்தமிழ் [KOṬUNTAMIL]*, *Composed for the Use of the Missionaries of the Society of Jesus, by, Missionary of the Said Society in the district of Madura*. Translated from the original Latin by George Wiliam Mahon. Madras, Vepery: Christian Knowledge Society’s Press.
- Beschi, C. J. 1876. *Clavis Humaniorum Literarum Sublimioris Tamulici Idiomatis. Auctore R. P. Constantio Josepho Beschio Societatis Jesu, in Madurensi Regno Missionario*. Tranquebar, India: Printed for A. Burnell by the Evangelical Lutheran Mission Press.
- Beschi, C. J. 1917 [ms. 1730]. *A Grammar of High Tamil, Latin text Published for the first time by L. Besse, s.j., with the English Translation by B.G. Babington, M.C.S., 2nd ed.* [*Grammatica Latino-Tamulica, ubi de elegantiori linguæ tamulicæ dialecto செந்தமிழ் [centamil] dicto [...]*]. Trichinopoly, India: St. Joseph Industrial School Press.
- Besse, L. 1917. see Beschi (1917).
- Burnell, Arthur Coke. 1880. *A Tentative List of Books and Some Mss. Relating to the History of the Portuguese in India Proper*. Mangalore: Printed at the Basel mission press.
- Chevillard, Jean-Luc. 2008. “The Concept of *ticaï-c-col* in Tamil Grammatical Literature and the Regional Diversity of Tamil Classical Literature”. *Streams of Language: Dialects in Tamil* edited by M. Kannan, 21–50. Pondicherry, India: French Institute of Pondicherry. (<https://shs.hal.science/halshs-00442188v1>)
-  Chevillard, Jean-Luc. 2017. “How Tamil was Described Once Again: Towards an XML-encoding of the Grammatici Tamulici”. *Histoire Epistémologie Langage* 39:2.103–127.
-  Chevillard, Jean-Luc. 2021. “From Grammar to Dictionary. The early challenge of lemmatizing Tamil verbal forms, through categories used for Latin and Portuguese”. *Journal of Portuguese Linguistics*, 20. (<https://jpl.lettras.ulisboa.pt/article/id/5689/>).

-  Chevillard, Jean-Luc, 2021. "The use of vector verbs in early modern Tamil". *Trends in South Asian Linguistics* edited by Ghanshyam Sharma and John J. Lowe, 367.261–290.
-  Chevillard, Jean-Luc & Eva Wilden, ed. 2004. *South-Indian Horizons. Felicitation volume for François Gros on the occasion of his 70th birthday*. Pondicherry: Institut Français de Pondichéry & École Française d'Extrême-Orient. [https://books.openedition.org/ifp/7481.]
-  Ebeling, Sascha & Margherita Trento. 2018. "From Jesuit Missionary to Tamil Pulavar. Costanzo Gioseffo Beschi SJ (1680–1747), the 'Great Heroic Sage'". *L'Inde et l'Italie* edited by Tiziana Leucci, Claude Markovits & Marie Fourcade, 53–89. Paris : Éditions de l'École des hautes études en sciences sociales. (https://books.openedition.org/editionsehes/23276)
- Hein, Jeanne & 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). Cambridge, Mass.: Harvard University Press.
- Henriques, Henrique. 1967 [1586]. *Aṇṭirikku aṭikaḷār iyaṛriya Flos Sanctorum enṇa aṭiyār varalāṟu*. [The *Flos Sanctorum* composed by the Venerable Henriques, also called *History of the Saints*]. Edited by Ca. Irācamāṇikkam. Tūttukkuṭi (India): Tamil Ilakkiyak Kaḷakam. [The original 1586 book is freely accessible online on the Vatican Library web site: https://digi.vatlib.it/view/MSS_Vat.ind.24]
- Horst, Christopher Henry, 1831² [1806]. See Beschi (1831² [1806])
- James, Gregory. 2000. *Colporul. A History of Tamil Dictionaries*. Cre-A, Chennai.
- James, Gregory. 2007. "The Manuscript and Printed Versions of Antão de Proença's Vocabulario tamulico (1679)". no:kku, 1.
-  James, Gregory, 2009. "Aspects of the Structure of Entries in the Earliest Missionary Dictionary of Tamil". *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) edited by Otto Zwartjes, R. Arzápalo & T. Smith-Stark, 273–301. Amsterdam: John Benjamins.
- Jeyaraj, Daniel. 2010. *Tamil Language for Europeans: Ziegenbalg's Grammatica Damulica (1716)*. Translated from Latin and Tamil, annotated and commented by Daniel Jeyaraj Wiesbaden: Harrassowitz Verlag.
- Mahon, G. 1848. See Beschi (1848).
- Muru, Cristina. 2010. *Missionari Portoghesi in India nei secoli XVI e XVII. Studio comparato di alcuni manoscritti*. Viterbo, Italy: Sette Città.
- Muru, Cristina. 2022. *The Linguistic and Historical Contribution of the Arte Tamulica by Baltasar da Costa, S.J. (c. 1610–1673)*. Annotated and commented Portuguese transcription and English translation from Portuguese and Tamil by Cristina Muru. Centro de Estudos em Letras / Universidade de Trás-os-Montes e Alto Douro, 25, LINGUÍSTICA, 978-989-704-496-0. [https://hal.science/hal-03878660]
- Proença, Antam (de). 1679. *Vocabulario Tamulico com a Significação Portuguesa (VTCSP)*. [See Thani Nayagam 1966]. [The original is available online on the Vatican Library web site at https://digi.vatlib.it/view/MSS_Borg.ind.12]
-  Puech Christian & Valérie Raby. 2011. "Présentation. Formes et enjeux de la rétrospection". *Histoire Épistémologie Langage* XXXIII:2.5–14.

- Thani Nayagam, Xavier S. 1954. "Tamil Manuscripts in European Libraries". *Tamil Culture* 3.219–228.
- Thani Nayagam, Xavier S. 1966. *Antaõ de Proença's Tamil-Portuguese Dictionary A.D. 1679*, [University of Malaya, Kuala Lumpur]. Leiden, Netherlands: E.J. Brill.
- Tolkāppiyam, *Poruḷatikāram*, *Ḫampūraṇam*. 2003. *Patippācīriyar T.i. Vē. Kōpālaiyar* (2 vol.). Chennai, India: Tamil maṇ patippakam.
- Tolkāppiyam, *Collatikāram*, *Cēṇāvaraiyam*. 2003. *Patippācīriyar T.i. Vē. Kōpālaiyar*. Chennai, India: Tamil maṇ patippakam.
- Trento, Margherita & Jean-Luc Chevillard. (Forthcoming). "Fishing for Words with Beschi in Tamil Traditional Poetical Vocabularies", *TST volume* [final outcome of the ANR-DFG project "Texts Surrounding Texts" (TST)] edited by Eva Wilden & Emmanuel Francis.
- Vaiyāpuri Piḷḷai, S. 1982 [1924¹]. "History of Tamil Lexicography". *Madras Tamil Lexicon* (MTL), Vol. 1., xxv–xliv. Madras: University of Madras.
- Vermeer, Hans J. 1982. *The First European Tamil Grammar. A critical edition*. [English version by Angelika Morath]. Heidelberg: Julius Groos Verlag.
- Vinson, Julien. 1910. "Le premier vocabulaire tamoul imprimé". *Revue de Linguistique et de Philologie Comparée* 43.74–82.
- VTCSP (= Vocabulario Tamulico com a Significação Portuguesa). See: Proença 1679.
- Walther, Christoph Theodosius. 1739. *Observationes Grammaticae, quibus Linguae Tamulicae idioma Vulgare [...]*, [Grammatical Observations Pertaining to the Ordinary Variety of Tamil]. Tranquebar: Typis Miffionis Regiæ [Royal mission press].
- Ziegenbalg, Bartholomäus. 1716. *Grammatica Damulica*. [Tamil grammar]. Halae Saxonum (Halle): Orphanotrophi [Orphanage]