



**HAL**  
open science

# From grammar to dictionary. The early challenge of lemmatizing Tamil verbal forms, through categories used for Latin and Portuguese

Jean-Luc Chevillard

► **To cite this version:**

Jean-Luc Chevillard. From grammar to dictionary. The early challenge of lemmatizing Tamil verbal forms, through categories used for Latin and Portuguese. *Journal of Portuguese Linguistics*, 2021, 20, pp.1-34. 10.5334/jpl.269 . halshs-03335445

**HAL Id: halshs-03335445**

**<https://shs.hal.science/halshs-03335445>**

Submitted on 5 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

RESEARCH PAPER

# From grammar to dictionary. The early challenge of lemmatizing Tamil verbal forms, through categories used for Latin and Portuguese

Jean-Luc Chevillard

Laboratoire d'Histoire des Théories Linguistiques (HTL, UMR 7597), CNRS – Université de Paris, FR

[jean-luc.chevillard@univ-paris-diderot.fr](mailto:jean-luc.chevillard@univ-paris-diderot.fr)

---

This contribution,<sup>1</sup> based on an examination of several Tamil dictionaries and Tamil grammars, composed in Portuguese and in Latin, by missionaries who were in Tamil Nadu during the 16th, 17th and 18th centuries, examines the lemmatization strategies which they followed, while dealing with Tamil verbal morphology. If nominal forms were not really a problem, verbal forms were more difficult to cope with. This is why for instance Proença's dictionary is very far from being completely lemmatized, and many of the forms which modern lexicographers would consider as falling under the same head, are listed as separate entries, and given separate translations. The complexity of the morphology was progressively mastered by grammarians, using labels taken from Portuguese or Latin terminology, although they did not always agree between themselves, concerning for instance what should be called *infinitivus*, some of them introducing new labels such as *infinitivus substantivus* and *infinitivus absolutus*. The most difficult nut to crack, however, was probably the existence of diathetic pairs, consisting of two paired verbs, which some modern linguists have referred to as 'affective' and 'effective', additionally accompanied by some causatives.

---

**Keywords:** Tamil; Portuguese; Tamil Verbal paradigm; Lemmatization; Tamil Lexicography; *infinitivus substantivus*; *infinitivus absolutus*; affective; effective; causative

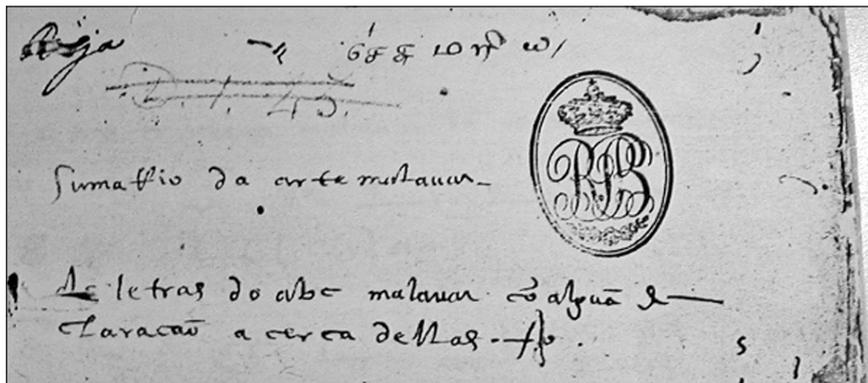
---

<sup>1</sup> This article is the final redacted form of a segment of linguistic exploration which was started at the time of the Lisbonne 2016 SALA conference, when I made a presentation for which the title was: 'On the debate between early Western descriptors of the Tamil linguistic complex concerning what should be called *infinitivus*'. I wish to express here my thanks to all those who have helped me when I was trying to reconcile several not easily compatible goals: one of them was to maintain an intermittent focus on my initial intuition while confronted with an even larger amount of linguistic data; another one was to make steady progress in the accomplishment of a long-running task which had been started in September 2013 and was finally completed in October 2020. That task consisted in entering as an XML file the text of the 1679 VTCSP, a rare printed book which is a bilingual Portuguese-Tamil *Vocabulario* and which is also one of the main theaters for the grammatical historical investigation which is conducted here. I must first acknowledge the help of Cristina Muru & Hugo Cardoso, thanks to whom I have often been in a situation of more easily deciphering, thanks to the consultation of a set of photographs of mid-17<sup>th</sup> century Goa MSS kindly provided by them, the content of many passages in the VTCSP, which is in fact a printed abridged version of the content of those earlier MSS. My thanks are also due to the NETamil program, and its principal investigator, Eva Wilden, for financing the digitizing of more manuscripts, preserved in the Paris BnF, which are 18<sup>th</sup> century copies of the VTCSP. I am also grateful to my colleague Gonçalo Fernandes for providing me with several Portuguese dictionaries, which have helped me navigating the complex universe of Portuguese glosses, and for being one of the organizers of the 'Host of Tongues' conference, in december 2018 in Lisboa, during which I had the occasion to benefit from my interaction with a number of linguists who are also native speakers of Portuguese, from Portugal and from Brasil. I also wish to express my thanks to Hugo Cardoso and to Shiv Kumar Sing for making it possible for me to give a talk at the CLUL (Centro de Linguística da Universidade de Lisboa) on 5<sup>th</sup> december 2019, and receive more feedback on my exploration of the VTCSP. I also express my thanks to the members and to the direction of my CNRS research team, the 'Laboratoire d'Histoire des Théories Linguistiques' (HTL, UMR 7597), CNRS – Université de Paris, for supporting me in my exploration of this branch of what we call 'Extended Grammars', following Sylvain Auroux, who has been an inspiration for many of us. Finally, I also express my thanks to the two anonymous reviewers of my article, when submitted to the JPL: their suggestions have helped me to clarify some aspects of this very complex long search, which I hope will be continued by others. I am also grateful to Cristina Muru for her role as efficient editor of this issue of the JPL, which would not have existed without her tireless energy.

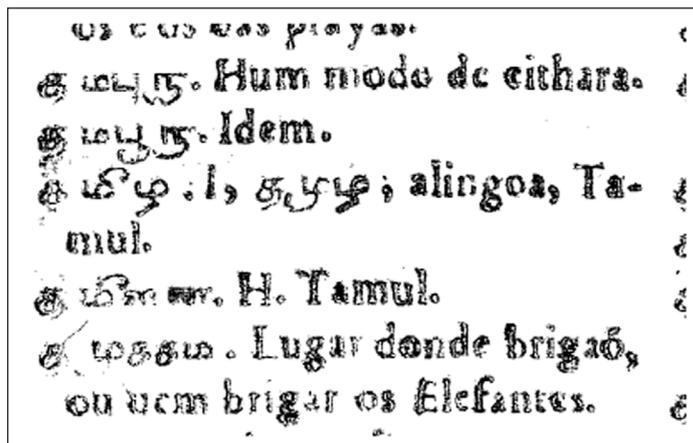
### 1. Prologue: encounter with a complex linguistic scene

This article, which falls under the larger domain called History of Descriptive Linguistics and, inside that domain, concerns what Auroux (1994) calls *grammatisation* ‘grammatization’,<sup>2</sup> concentrates mostly on some of the traces left by speakers of Portuguese who were trying in the 16<sup>th</sup> and in the 17<sup>th</sup> centuries, by writing grammars and compiling dictionaries —see the extracts from the *Arte* composed by Henrique Henriques [1520–1600] (henceforth HH) and from the *Vocabulário Tamulico Com a Significação Portuguesa* (henceforth VTCSP) inside **Figures 1** and **2**— in order to fulfill their Missionary activities, to master a South Indian Language, which they referred to as *malauar* (see **Figure 1**) or as *Tamul* (see entries 342\_L\_j & 342\_L\_k in **Figure 2**, and see the transcription-translation provided in (1) & (2) below).

As those missionaries would progressively discover, the *tamul* language which they were trying to master —usually referred to in English as ‘Tamil’— turned out to be a symbiotic combination of three languages, difficult to separate in practice, because each component of the Tamil Triglossia (see **Figure 3**) had its own role to play in the global picture of the everyday life in Tamil Nadu, linguistically punctuated by solemn occasions, often religious. In this linguistically complex environment those missionaries also wanted to play a significant role, an ambition which was, to some extent, fulfilled for some of them.

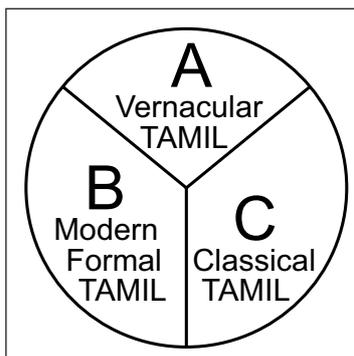


**Figure 1:** Beginning of the 16th cent. *Arte* attributed to Henrique Henriques (HH) (cod. 3141, Biblioteca Nacional de Portugal, Lisbon).

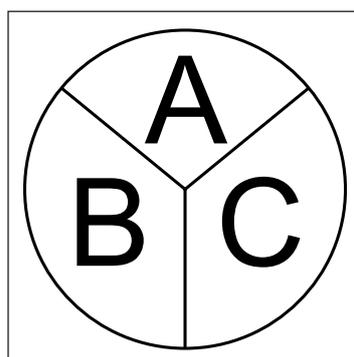


**Figure 2:** Five entries (342\_L\_h to 342\_L\_l) from the 1679 VTCSP (*Vocabulário Tamulico Com a Significação Portuguesa*).

<sup>2</sup> See Auroux (1994). The currently accepted English translation for French *grammatisation* is ‘grammatization’. See Pellin (2019). Briefly stated, *grammatization*, which is not the same as *grammaticalization*, is what happens to a language when it is equipped with grammars and dictionaries.



**Figure 3:** Three languages called ‘Tamil’ (the Tamil triglossia).



**Figure 4:** Tamil Triglossia (simplified form of Figure 3).

- (1) த மி ழ [tamiḷ]. 1, [‘or’] த மு ழ [tamul], a lingua, Ta- // = mul ‘the language, Tamil’.
- (2) த மி ள ன [tamiḷan]. H. Tamul. ‘Tamil man’.

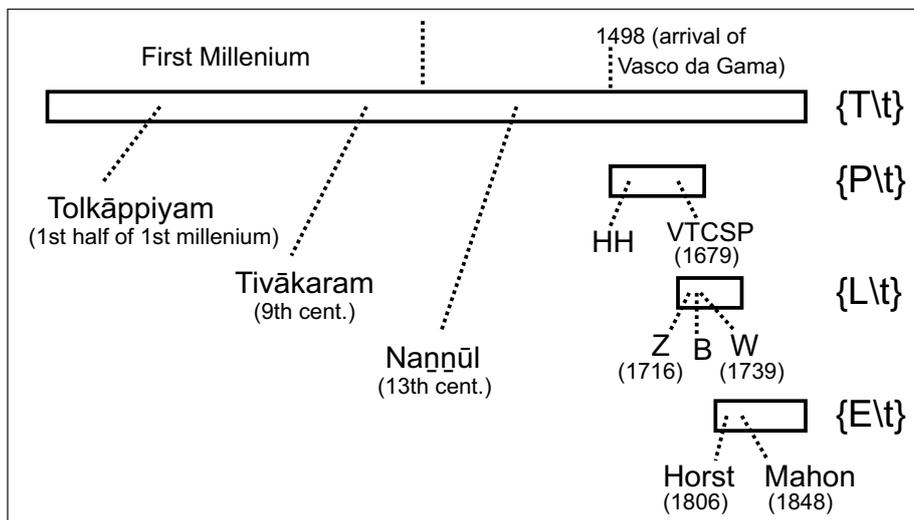
As can be seen in the above citations (1) and (2), which contain as additions standardized transcriptions of the Tamil characters and English translations of the Portuguese glosses found in the VTCSP entries 342\_L\_j & 342\_L\_k, two variants of the name of the Tamil language are mentioned by Antam de Proença [1625–1666] (henceforth AP), who had compiled the data which would be printed after his death in the form of the 1679 VTCSP book. Those two variants, which are *tamiḷ* and *tamul*, are a clear illustration of what is normally referred to as the Tamil Diglossia (B vs. A), expanded by me into a Triglossia, inside **Figure 3** (C vs. B vs. A), which I shall also symbolize occasionally by means of  $\text{Ⓢ}$ .<sup>3</sup> The first form mentioned in (1), namely « த மி ழ », transliterated as [tamiḷ], is the standard form, representative of Variety B (formal Tamil), and is also found in Variety C (Classical Tamil). The second form mentioned, namely « த மு ழ », transliterated as [tamul], is one of the possible variants found in some dialects of spontaneously spoken vernacular Tamil (Variety A). Since it is the form which is the closest to the Portuguese form *tamul*, to which the French form *tamoul* is comparable, we can probably conclude that this dialectal variant<sup>4</sup> is the one with which the

<sup>3</sup> I have first made use of the  $\text{Ⓢ}$  notation seen in **Figure 3** and in **Figure 4** in Chevillard (2018).

<sup>4</sup> I would like to express here my thanks to Bharathan R D (@bharathanrd) and to Gopalakrishnan R (@cobbaltt), who are both very knowledgeable on existing living Tamil dialects, and on many other Indian languages, and with whom I frequently interacted, on the Tamil side of *Linguistic Twitter*, during the COVID 2020 pandemic year, when travels were impossible. I (as @JLC1956) have explored with them many questions starting with the question why my mother-tongue, French, writes as ‘tamoul’ what English writes as ‘Tamil’. It is thanks to their encouragement and to that of others, including my wife, Eva Wilden, that I have finally managed in October 2020 to finish entering in XML format, after approximately 4000 hours of typing work, over the

Portuguese travelers first came into contact, when Vasco da Gama reached India. As for the form seen in (2), namely « த ழி ள ன », transliterated as [tamiḷan], it looks like a kind of compromise between the A and B varieties, because the use of the *i* vowel in the second syllable seems to indicate an attempt at using formal Tamil (i.e. the B-variety), whereas the use of the consonant *ḷ* instead of the standard *l* seems to indicate that the word was uttered by a speaker of a dialect in which *ḷ* and *l* had merged, as is the case for instance in Madurai nowadays, although a school teacher in Madurai would now, if seeing (2) in the notebook of a student, immediately correct it and criticize it as a serious mistake.

As can be seen by a rapid look at **Figure 5**, the temporal span covered by all the sources mentioned in this article is extremely vast. It starts with the *Tolkāppiyam*, an ancient grammatical treatise, which belongs to the most ancient period for Classical Tamil literature (the Variety C of **Figure 3**). It goes up to the modern period, where historians of linguistics are trying to evaluate what took place during this long time span, and where the linguistic situation has also changed. We can indeed see that if there exist nowadays many human beings who have a simultaneous mastery of English and of Tamil (in several of its varieties), the number of human beings who have a simultaneous mastery of Portuguese and Tamil is probably very small. It is even smaller if we restrict ourselves to those who are simultaneously capable of deciphering the Portuguese sentences contained in a 16<sup>th</sup> or 17<sup>th</sup> century MS or book, along with the Tamil passages, in ambiguous Tamil script or in approximate Roman transcription, which are found in the same book or MS. The same remarks can be made concerning the grammars and dictionaries composed in Latin, such as the three grammars which are represented by the letters Z, B and W inside **Figure 5**. Those letters stand for the works of B. Ziegenbalg [1682–1719], C. J. Beschi [1680–1747]<sup>5</sup> and C.T. Walther [1699–1741]. If two of these grammars had not been translated into English, very few people nowadays would have even an approximate idea of their content. It has been my intention, in writing this article, to rediscover myself and to help others to



**Figure 5:** Brief chronology of the early stages in the description of Tamil HH = Henrique Henriques; Z = Ziegenbalg; B = Beschi; W = Walther; P = Portuguese; L = Latin; E = English; t = Tamil as object-language; {T\t} stands for a description of Tamil (=t) written in Tamil (=T).

course of seven years, the text of the 1679 VTCSP (*Vocabulario Tamulico Com a Significaçam Portuguesa*), a text which has a seminal importance for the study of Tamil dialectology, being, along with its manuscripts counterparts, one of the earliest systematic evidence for the exploration of the Tamil diglossia, in which the A-varieties and the ideal B-variety of Tamil constantly interact, in the every-day practice of living Tamil speakers, many of whom also have a strong devotion for the C-variety of Tamil.

<sup>5</sup> For a description of Beschi’s successful career in South-India, see Ebeling & Trento (2018).

rediscover many forgotten topics, by exploring what is in fact ancient primary linguistic field work. That field work —or initial *terrain*— is especially important, from the point of view of general linguistics, because it represents a first contact between languages which were not part of the same linguistic area, and were thus in a situation of maximum alterity.<sup>6</sup>

In order to illustrate in a compact manner the linguistic diversity, a number of symbols, such as {T\t}, {P\t}, {L\t} & {E\t},<sup>7</sup> have been placed on a vertical axis found on the right side of **Figure 5**, and complement the temporal information of the horizontal axis. Inside those four symbols, the capital letters T, P, L & E stand respectively for the Tamil, Portuguese, Latin & English languages, suitably extended in order to be usable as metalanguages for describing Tamil, noted here by the lower case letter t, in order to signify that it is the object-language in a description, which can be either a grammar, as in the case of the *Arte* attributed to HH, or a dictionary, as in the case of AP's 1679 VTCSP.

## 2. A preliminary question: the infinitive

The original inner incentive for the writing of this specific article has been my attempt at finding an answer to a question which I would now, with hindsight, reformulate in the following manner:

- (3a) How was a consensus historically reached, leading to the quasi-unanimous modern use of the label ‘infinitive’<sup>8</sup> inside grammars and dictionaries falling under the {E\t} type —where (capital) E stands for English used as a metalanguage and (lower-case) t stands for Tamil being the object language inside a linguistic description— when referring to those forms belonging to the Tamil verbal paradigm which would be referred to by a traditional Tamil grammar of the {T\t} type — where (capital) T stands for Tamil used as a metalanguage— as instances of [*ceyya-v-ennum vāypāṭṭu*]<sub>specifier</sub> [*viṇai-y-eccam*]<sub>head</sub>, an iconic Tamil technical expression in which a general head, namely *viṇai-y-eccam* (henceforth VE),<sup>9</sup> which is the common designation of several grammatical categories —converb (VE1), infinitive (VE2), conditional (VE3), etc. in the parlance of some English speaking modern linguists— is made into a more precise designation by means of a specifier, namely *ceyya-v-ennum vāypāṭṭu*, in which the underlined term, namely *ceyya* ‘to do’, is an individual item chosen as iconic symbolic representative (i.e. a *vāypāṭṭu*)<sup>10</sup> of the formal class to which it belongs (i.e. the class of ‘infinitives’)?<sup>11</sup>

<sup>6</sup> This work is also a continuation of Chevillard (2015) & Chevillard (2017), drawing on the progress made in my preparation of an electronic edition for the VTCSP.

<sup>7</sup> The {A\b} symbol, which stands for a description of a natural language b by means of a metalanguage A, obtained by extending a natural language a, was introduced in Aussant & Chevillard (2020), as a simplification of an earlier notation seen in Auroux (1994).

<sup>8</sup> Covering all the three components of the Tamil (ᱵᱤᱨᱫᱟᱹᱜᱟᱲ) triglossia, one can compare for instance Schiffmann (1979) describing Modern Spoken Tamil (A-variety), Lehmann (1989) describing Modern Formal Tamil (B-variety) and Wilden (2018) describing Classical Tamil (C-variety).

<sup>9</sup> In traditional Tamil grammar, a *viṇai eccam* (VE) is a sub-type falling under a general type called *eccam* ‘incompletion’. In the parlance of that grammatical school, the incompletion —which is syntactic— of the *viṇai eccam* is ‘made complete’ (i.e. removed) as soon as the *viṇai eccam* is followed by a *viṇai* ‘verb’. The most frequent type of *viṇai eccam* —for which the model (or *vāypāṭṭu*) is iconically called *ceytu* (approx. ‘having done’)— is referred to by some English-speaking typologists as a ‘converb’ (VE1), but has received many other designations in the course of history. See (3b).

<sup>10</sup> For more information on several technical devices seen in traditional Tamil grammars, see Chevillard (2009).

<sup>11</sup> For the sake of completion, we can add that: the head noun phrase *viṇai-y-eccam* can be approximately rendered by ‘incomplete verbal form’; the form *vāypāṭṭu* is the oblique of *vāypāṭṭu* ‘model’; the iconic label *ceyya* for that model is taken from the paradigm of a verb to which the 20<sup>th</sup> cent. Madras Tamil Lexicon (MTL) refers as *ceyṭal* ‘to do’, although we shall in this article refer to it as the verb *ceykiratu* ‘fazer’, following the 1679 VTCSP.

That question (3b) had first in its background, my personal opinion that the traditional Tamil terminology, which has been developed by Tamil grammarians over the course of many centuries, is in fact preferable and more natural. It was also reinforced by a second observation, formulated in what follows as a question, namely:

- (3b) Why is there much less unanimity among the English-speaking descriptors of Tamil when coining a label for referring to the most frequent type of VE (*viṇai-y-eccam*), henceforth VE1, which Tamil grammarians refer to as the *ceytu-v-ennum vāypāttu viṇai-y-eccam* ‘*viṇai-y-eccam* of the type/model called *ceytu*’, and for which we can find among others the following attempted English renderings: ‘Gerund’ (Mahon 1848: 54, parag. 71)), ‘adverbial participle’ (Pope, as quoted by Arden 1976: 199, fn. 1), ‘verbal participle’ (Arden, 1976: 199), ‘past participle’ or ‘verbal participle (past)’ (Schiffmann 1979: 20 & 68), ‘conjunctive’ (Steever: 2005: 78–85), to which must be added ‘converb’.

If we were to stay within the {T\t} framework, where the types discussed in (3a) and in (3b) are both called *viṇai-y-eccam* (henceforth VE), we could refer to them compactly as VE2 (whose model is *ceya*) and as VE1 (whose model is *ceytu*), adopting the order in which Tamil grammarians enumerate them, which corresponds in fact to their frequency, because VE1 forms are more frequent than VE2 forms, as revealed by the observation of Tamil textual corpora.

However, since we are in this article exploring the roots of {E\t} type grammars, and since all English-speaking grammarians and linguists refer to the VE2 form as ‘infinitive’, we shall also make frequent use of that expression, although we shall see, for instance in section 6, that the current unanimity has not always existed.

My attempt at answering question (3a), by moving back in time, from the 21<sup>st</sup> to the 18<sup>th</sup> century, had first taken me, when making the initial conference presentation, to an exploration of the two most visible {L\t} types of grammars —where L stands for Latin— namely the 1716 *Grammatica Damulica*, composed by B. Ziegenbalg and printed in Halle, and the 1728/1738 *Grammatica Latino-Tamulica*, written in 1728 by C.J. Beschi and printed in Tranquebar in 1738. To those two should be added a third {L\t} type grammar, written by C. T. Walther and printed in 1739 in Tranquebar, as an (unasked for)<sup>12</sup> supplement to Beschi’s grammar. Placed in the wider chronology, those three sources are represented by the three letters Z, B and W inside **Figure 5**, as has already been said in the previous section. In the same **Figure 5**, one can also see, just below, inside the {E\t} line, two more names, with a date, Horst (1806), and Mahon (1848). These are the names of the authors of the two English translations<sup>13</sup> of Beschi (1738) and it is certainly thanks to the existence of these two translations that Beschi’s grammar has undoubtedly remained directly influential for the longest duration of time. Additionally, it has been reprinted (both in the original and in translation) several times, whereas Ziegenbalg’s grammar has had to wait until 2010 for finally becoming accessible to the English-speaking world, thanks to the efforts of Daniel Jeyaraj. As for Walther’s grammar, which has never been translated, it has probably had very few readers, especially in the modern period. However, because Ziegenbalg’s grammar has been the (veiled) target of Beschi’s criticism, who for instance did not agree with his use of the label *infinitivus substantivus*,<sup>14</sup> his positions

<sup>12</sup>The (catholic) Beschi is known to have been unhappy that the protestants in Tranquebar had been bold enough to try to publish Walther’s 1739 *Observationes* as a supplement to his 1738 *Grammatica*. See for instance Arno Lehmann (1955: 209).

<sup>13</sup>See bibliography: Beschi (1831) and Beschi (1848).

<sup>14</sup>See section 6 in this article, which contains the Latin text and the English translation of Beschi (1738: 95. parag.107).

were somehow present in some of the implicit debates. As for Walther, who is indebted, in different ways, to both Ziegenbalg and Beschi, but was also a kind of general linguist, interested in many languages, including Hebrew, his distinct formulations also throw some new light on the debate between his two immediate predecessors, but we must reserve their detailed examination for a future study.<sup>15</sup>

However, as I have tried to suggest in a preliminary manner in the introductory section, the nature of the linguistic configuration which the European missionaries encountered when they reached South India was much more complex than is usually reported in some simplified accounts of the discovery of Tamil. For that reason, it has progressively appeared to me that an answer to the limited question asked in (3a), above, could really be informative only if it started from the very beginning of the external description of Tamil, or at least from what is reachable of that very beginning, namely the *Arte* composed by HH, which falls under the formula {P\t}, where P stands for Portuguese.

The simple truth is that one cannot really understand what Ziegenbalg, Beschi and Walther were trying to do if one does not study the work of their predecessors, who wrote in Portuguese. Ziegenbalg, Beschi and Walther had of course their own sources of information in the Tamil-speaking world but they also had at their disposal a relatively large body of already compiled knowledge, which took the form of grammars and dictionaries, written in Portuguese (and sometimes in French), some of which are still available to us, as MSS or as printed books, although much remains to be done for the information which they contain to be really accessible to a modern reader. When they make statements about Tamil, we should not believe that everything they say is based on independent direct experience. What they tell us is partly based on the sources which they had studied and which they reformulate, trying to improve upon them, but, by so doing, they tend in fact to replace the initial discovery of the A-variety of Tamil, by an exposition targeted towards the B-variety of Tamil. Such an improvement may not necessarily be a progress, from the point of view of descriptive linguistics.

Additionally, another parameter which is also important in the evaluation of the linguistic survey made by Ziegenbalg, Beschi, Walther or AP, is what we could call their virtual 'Tamil Library environment'. What this expression indicates is that we should try to evaluate their own estimation of the complexity which they faced in trying to have a global view of Tamil literary culture, which was, and still is to some extent, totally dominated by the C-variety of Tamil. This environment is partly accessible to us thanks to the documents which they collected. Among these:

- some are known to us by catalogues which they left although the artefacts catalogued may have disappeared (as is the case with Ziegenbalg),<sup>16</sup>
- some are still available thanks to actual collections which have been preserved (such as the BnF)

As a general observation, I shall remark that all those collections, virtual or still in existence, are dominated by the C-variety of Tamil, which is the only variety which was really admired and treasured by the inhabitants of Tamil Nadu, who would never have considered it a useful task to attempt to describe the language (Variety A) which they used in their everyday life. This is the reason why HH composed his *Flos Sanctorum* in the B-variety of Tamil. This is also the reason why Beschi went much further and composed a long Christian poem, the *Tēmpāvāṇi*,<sup>17</sup> in the C-variety of Tamil. And, from the point of view of descriptive

<sup>15</sup> It is currently my plan to translate Walther's grammar, from Latin into English, once my work on the VTCSP is completed.

<sup>16</sup> See *Bibliotheca Malabarica* (Sweetman & Ilakkuvan 2012).

<sup>17</sup> See see Ebeling & Trento (2018). For a complete English translation of the 3615 stanzas of the *Tēmpāvāṇi*, see Dominic Raj (2019).

linguistics in general and Tamil dialectology in particular, this is why the *Arte* prepared by HH and the VTCSP prepared by AP are such invaluable sources of information, without equivalent, as far as the first component of the  triglossia is concerned.

### 3. A brief overview of HH's 16<sup>th</sup> century *Arte*

Regarding the *Arte* attributed to HH, we are lucky that a number of scholars have worked on it. It was rediscovered and identified in Lisbon as a unique MS, the codex 3141 (see **Figures 1** and **6**), in the BNP (Biblioteca Nacional de Portugal) by X. Thani Nayagam. It has been critically edited by H.J. Vermeer in 1982. An English translation by J. Hein & V.S. Rajam is available since 2013. Nevertheless, it remains a very difficult task to navigate that text and to locate elements of information in it.<sup>18</sup>

Our starting point will be a brief presentation of the section dedicated to verbs, because it is inside that section that some elements of an answer to the preliminary question (3a) concerning the beginnings of the use of the English technical term 'infinitive' for Tamil can be found, as can be verified inside **Figure 7**, where a line reads: "15 o *ēfinitiuo*<sup>19</sup> [emphasis mine] ē thca ou rca". It is also there that the seeds for answering our enlarged question can be located. That enlarged question addresses the challenge of understanding the implicit organizational practices which govern a non-lemmatized dictionary such as the VTCSP, which is our central target in this article, and which stands between a grammar and a dictionary.

At the beginning of the section dedicated to verbs the upper part of fol. 37r in that MS announces that nine distinct verbal 'conjugations' (i.e. *comjugações*) are distinguished (see **Figure 6**).<sup>20</sup> This announcement is followed by a prologue (fol. 37r to fol. 39v), where the general characteristics of each of the nine conjugations are briefly explained. See for example **Figure 7** which contains a transcription, extracted from Vermeer (1982), of the brief characterization, given in that prologue, of the various stems seen in the two classes distinguished by the *Arte* in the 3<sup>rd</sup> conjugation.

After the prologue, the verbal paradigm as it appears in each of the nine conjugations is detailed over circa 100 folios (from fol. 39v upto fol. 139r). This long section, which corresponds to pp. 33–113 in Vermeer (1982) and to pp. 81–220 inside Hein & Rajam (2013), is then followed by a shorter one, from fol. 120r upto fol. 128r, where the nine conjugations are revisited with respect to the specific morphology of what would now be referred to as *neuter forms*.<sup>21</sup> Then comes another subsection, from fol. 129 r upto fol.

<sup>18</sup>I do not have the same assessment as some of my predecessors concerning the nature of the grammar composed by HH and preserved in Lisbon, in the BNP, which became more easily accessible after H. J. Vermeer published his critical edition in 1982. Vermeer himself has written, concerning 'the explanation of the verb forms' used in this grammar, that 'from a practical point of view, this classification is easier to learn than the modern one' (Vermeer 1982: p. xviii, lines 23–24 & p. xix, lines 13–15). This might be a slightly optimistic position. HH's grammar is extremely difficult to understand, and would be even more difficult to use. Even now that an English translation is available, thanks to Hein & Rajam (2013), it does not seem conceivable that someone could use it for learning Tamil. HH's grammar is first of all a historical document, which reminds us how difficult it is to make a description of a completely unknown language. This is probably the only answer one can provide to one of Otto Zwartjes' questions, in his 2011 book where he wonders, on p. 44, why modern linguists seem to ignore HH's *Arte*.

<sup>19</sup>This is one among several possible spellings seen in the MS of the *Arte*. Another spelling is *imfinitiuo*, as seen in the column ζ of Table 1. This designation is used in the 16th cent. *Arte* for the form which Ziegenbalg (1716) proposed to call *Infinitivus absolutus* (See Figure 11).

<sup>20</sup>The transcription given in Vermeer (1982: 30) for the MS passage visible in Figure 6 is: *Segêse as comjugações dos verbos as quais são 9. As 3 primeiras se acabaõ em quiren, as outras tres que se seguem se acabaõ e Ren, as tres derradeiras se acabaõ em guiren.* Hein & Rajam (2013: 76) translate: 'The conjugations of the verbs follow. There are nine. The first three end in *quiren*; the next three end in *RRen*; the last three end in *guiren*.'

<sup>21</sup>HH's MS contains as a title for this section *Verbos jmpersoais ou das cousas jrracionaês.* Hein & Rajam (2013: 202) translate: 'Impersonal verbs, or concerning irrational things.' This section first discusses the 3<sup>rd</sup> person Neuter forms of all the verbs which have been previously discussed in the long section going from fol. 39v upto fol. 139r. This is continued by a discussion of a small group of verbs which are attested only in the 3<sup>rd</sup> person Neuter, like for instance *verhiquidu* 'estrala' ['it bursts']. Interestingly, when using the VTCSP scheme,

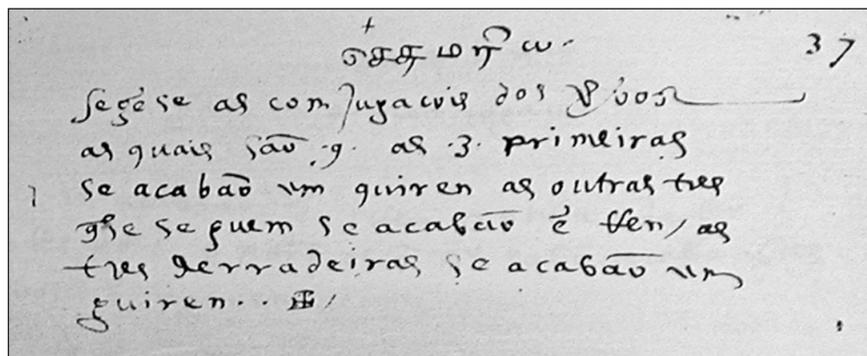


Figure 6: Extract from top of folio 37 r in Codex 3141 (BNP, Lisbon).

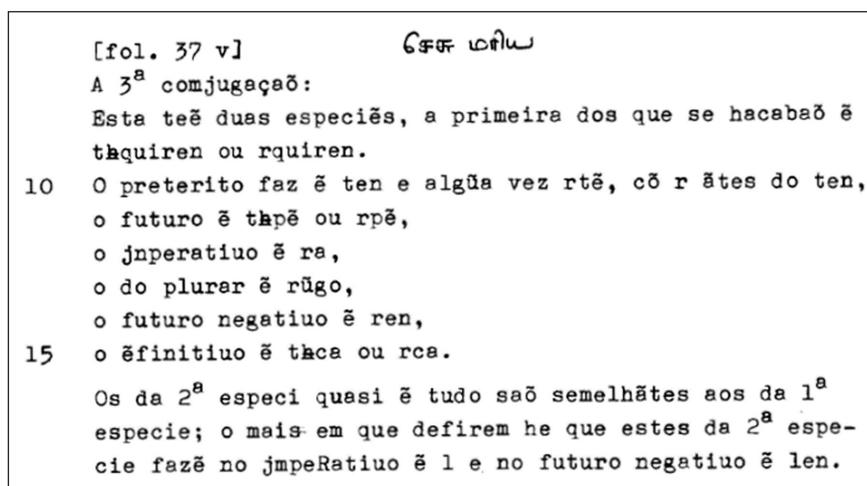


Figure 7: Vermeer 1982 transcription of folio 37 v in Codex 3141 (BNP, Lisbon): Preliminary presentation of the 3<sup>rd</sup> conjugation.

139 r, for which the title is *Dos verbos passiuos*.<sup>22</sup> Then comes a coda, from fol. 140 r upto fol. 142 v, which is not part of the nine conjugation system and for which the title is ‘sum, es, fui’. The goal of that coda is to give a separate presentation of a special verb, namely *aguires* ‘eu sou’ [‘I am’], which plays mostly the role of a copula.<sup>23</sup> All this is of course a huge amount of information, and I shall draw in what follows only from the initial long subsection. While doing so, I shall take advantage of the fact that Hein & Rajam have introduced some special notations, making use of what they call *sub-tenses*, which have to be used as additional specifications, along with the *tenses* (i.e. *tempo*s) mentioned in the *Arte* attributed to HH. The use of those subtenses, indicated by capital letters, from A to R as maximum value, greatly simplify the navigation of this ocean of data, while at the same time clearly showing why it is so difficult to really master it.

In a chart which is found below (see **Table 1**), I have provided in compact form implicit pointers towards all the elements which are discussed inside that larger sub-section of the

where the citation form is the 3<sup>rd</sup> person neuter singular, and not the 1<sup>st</sup> person singular, there is no need to have a special category for such items, as can be seen in the corresponding VTCS entry *vetikkiratu* (432\_R a).

<sup>22</sup> Hein & Rajam (2013: 211) translate: ‘Concerning the passive verbs.’ The content of this section in the *Arte* is related with the topic which I shall discuss later inside section 10 with respect to the VTCS. I shall briefly revisit the *Arte* there, and try to point out how this is clearly a feature concerning which a lot of progress took place, in terms of descriptive adequacy, in the hundred years which separate HH’s *Arte* from the VTCS. This same topic is also discussed, with a larger textual basis, inside Muru (2021), inside this issue of the JPL.

<sup>23</sup> See also the references given in the last row of Table 2.

**Table 1:** Distribution of 759 items, and additional information, inside the largest section for the verbal paradigm in Hein & Rajam (2013: 81–202).

Con- jug.	page span	$\alpha$ Presente (Present) A-K	$\beta$ Preterito (Preter- ite) A-R	$\gamma$ Futuro (Future) A-G	$\delta$ Impera- tuo (Impe- rative)	$\epsilon$ Futuro negatio (Nega- tive) A-I	$\zeta$ infinitiuo (Infinitive) A-R	Rules & sum- maries	Examples (&Irregu- larities)
C1	81–114	81–88	88–96	96–100	100	100–104	104–111	111–112	113–114
C2	114–125	114–116	116–118	118–120	120	120–121	122–124	124	125–126 (124–125)
C3a	126–137	126–127	127–129	129–130	130	130–131	131–133	134–135	135–136
C3b		133			133	133–134			136–137
C4	137–148	137–139	139–142	142–143	143	143–144	144–146	146	146–148
C5a	148–167	149–150	150–152	153	154	154–155	155–156	164–165	165
C5b		157–158	158–160	160–161	161	161–162	162–164	166	166–167
C6	167–176	167–169	169–171	171	172	172–173	173–175	176	176
C7	177–185	177–178	178–180	180–181	181	181–182	182–184	184–185	185
C8	185–194	185–187	187–189	189–190	190	190–191	191–193	193	194
C9	194–202	194–195	195–197	197–198	198	198–199	199–201	201–202	202

section dedicated to the verbal paradigm in HH's *Arte*. The pointers are implicit because they have to be constructed on the basis of two or three components:

1. The row information, which tells us which conjugation (among the nine, with their eventual sub-classes) is concerned
2. The column information, which tells us which tense is intended
3. The sub-tense information, which consists in choosing a capital letter, from the ranges provided at the top of the column

Combining the *conjugation* and the *tense* information from the rows and the columns, it can be seen that the chart provided as **Table 1** gives access to 66 items, falling under 6 types, in the description of the verbal paradigm. If we add to that the third element which Hein & Rajam call a sub-tense, we have access to 693 additional items, falling under 63 types, which gives a total of 759 items, falling under 69 types. For each of these 759 items, the chart tells us inside which page range of Hein & Rajam (2013) the information (in English) provided by those two scholars can be located. I must add to this that some of those 759 items are paradigms, others are list of forms and still others are individual items. The chart is as follows:

A possible way of making use of this huge amount of data, is to select a given type, and to go through the successive rows, comparing what is said about that type, in the successive rows. This will be done here first for the type which Hein & Rajam (2013) refer to as 'present tense, subtense J', making use of that designation for the first time on p. 87, while translating folio 44 r from the *Arte*. I shall refer to this type in a more compact manner as the type  $\alpha$ -J. This type does not receive any Portuguese technical designation in the *Arte*, where we simply see (info-1) the form named, to which is added (info-2) an explanation of its morphology —How to derive it from the present tense?— followed by (info-3) an explanation of its meaning —How to translate it?— and then by (info-4) some examples of its use.<sup>24</sup> Such a detailed treatment is however not available for all the conjugations, and it is mostly the first conjugation which serves as the occasion for providing the most detailed

<sup>24</sup>This will be revisited inside Section 7, where I shall also establish a link with the important type of construction which Lindholm (1972) has called *cleft sentences*.

set of information features.<sup>25</sup> As a first example, I shall provide here in chart form the occurrences of the type  $\alpha$ -J in the *Arte*. I could also make use of Ziegenbalg's designation for that form, namely *Infinitivus Substantivus*, but that would be going too fast in the unfolding of the chronology.<sup>26</sup> That type  $\alpha$ -J has been chosen by me as the initial item to be examined because of the immense importance which it has in the organization of the 1679 VTCSP, where it acts as the head for what I shall call a 'family of entries', as we shall see in the following section. However, at this early stage, because the type  $\alpha$ -J is, in the organization of the 16th cent. *Arte*, secondary to the  $\alpha$  type (i.e. the present), from which it is derived in the description, we have no choice (except in the first row C1) but to provide its meaning on the basis of the meaning given for the primary entry  $\alpha$ . The chart is as follows:

Finally, this chart can play the role of a transition between, on the one hand, the classification system of the *Arte* attributed to HH, in which each conjugation is presented on the basis of several main stem forms, starting with the 1st person present (see for instance **Figure 7** for the case of the 3<sup>rd</sup> conjugation) and, on the other hand, the classification system of the VTCSP, which will be our object of investigation in the next section.

#### 4. A brief overview of AP's 1679 VTCSP

After this long contextualizing introduction, we now turn to what might be the more original part of the present study, containing elements of information which have not been available for a long time. We shall now examine more in detail the 1679 VTCSP, which is the posthumous work of AP, printed in book form in Ambalacatta (Kerala) by his colleagues. The VTCSP was rediscovered in the Vatican Library by several scholars, although one can also argue that the single available Vatican Library copy of the VTCSP was never really lost (see Vinson 1910). It has been published as a Facsimilé by X. Thani Nayagam in 1966, and has been studied several times by G. James (See James 2000, 2007 & 2009). It is however not yet as accessible as HH's *Arte*. I have been engaged since 2013 in the preparation of an electronic edition<sup>27</sup> of that text, which contains 16,215 entries, as per my current estimate,<sup>28</sup> on 508 unnumbered pages. The odd pages in the VTCSP contain folio numbers in the Upper right corner but unfortunately, because of some mistake in the printing the great majority of these numbers are faulty and there are even duplicate numbers. As a clear example, the last folio in the VTCSP, whose recto is (logical) page 507 and whose verso is page 508, bears the folio number 247, when it should have been numbered as 254.<sup>29</sup> The consequence of this state of affairs is that the VTCSP folio numbers are unusable for all practical purposes. For that reason, I have decided to make use of the logical page numbers in my references to the pages: entries are referred to by means of a coordinate system indicating:

- the logical page number (from 1 to 508)
- the column: L for 'left' or R for 'right'
- the rank: indicated by a lower case letter (from a to y)<sup>30</sup>

<sup>25</sup> The features 'information-2' and 'information-4' are provided only for the first conjugation. The feature 'information-3' is also missing.

<sup>26</sup> This will be discussed in Section 6, where I shall also reproduce Beschi's disapproving comments.

<sup>27</sup> A number of statistical elements provided in the continuation of this article are based on the results of that editing task.

<sup>28</sup> For a general presentation, see Chevillard (2017). A number of sources state that the VTCSP contains 16,546 entries. See for instance James (2000: 96) & Zwartjes (2011: 272), but this is simply a consequence of the unverified repetition of an initial erroneous estimate first found in Thani Nayagam (1966: 9).

<sup>29</sup> See for instance the bibliographical entry 255 in Dhamotharan (1978: 69), which reads: 255 PROENÇA, Antam de. // Vocabulário Tamulico com a significação Portuguesa. [A Tamil-Portuguese dictionary.] Na imprensa Tamulica da Provincia do Malabar, por Ignacio Aichamoni impressor della, Ambalacatta, 1679. (lo), 247 fo. // [First printed dictionary in Tamil.]

<sup>30</sup> The column having the biggest number of entries is column 245\_L, which has 25 entries, numbered from 245\_L\_a to 245\_L\_y.

This coordinate system is the one which is used, for instance, inside the seventh column of **Table 2**, inside the previous section.

**Table 2:** occurrences of the items belonging to the  $\alpha$ -J type in HH's *Arte* (along with the Present stem forms to which they are attached): the  $\alpha$ -J type will be baptized *Infinitivus Substantivus* by Ziegenbalg 150 years later.

Conjug.	<i>Arte</i> <sup>31</sup> (with English from H.&R.)	fol.	Ve.	H.&R.	VTCSPP (entries) <sup>32</sup>	VTCSPP coord.	VTCSPP (meaning)
C1	vichuadiquiradu 'o crer' ['the believing'] (→ nan vichuadiquiren 'eu creo' ['I believe'])	44r	37	87	viccuvatikkiratu vicvacikkiratu	434_R_m 447_R_m	'Pro vicvacikkiratu crer' 'Crer'
C2	pilaquiradu (→ nan pilaquiren 'eu fendo' ['I am splitting'])	74r	59	116	piḷakkiratu	222_R_f	'Fender, rachar'
C3a	pathquiradu (→ pathquiren 'eu olho' ['I look at'])	83r	65	127	pārkkiratu	213_R_a	'Olhar'
C3b	[form non provided] (→ vithquiren 'vêdo' ['I sell'])	(86r)	(68)	(136)	viṛkkiratu	444_L_b	'Vender'
C4	etiradu (→ etiRen, 'eu carego' ['I load'])	89v	71	139	ēttukiratu ēṛṛukiratu	77_L_d 75_L_p	'Vid. erri.' 'Aleuantar, actiué'
C5a	coliradu (→ coliren 'cōpro' ['I buy'])	95v	76	150	koḷḷukiratu	299_L_o	'Receber. itē comprar'
C5b	puduradu (→ puduren 'entro' ['I enter'])	98r	79	158	pūtukiratu pukutukiratu	250_R_f 246_L_f	'Entrar' 'Entrar'
C6	paýRadu (→ paýRen 'eu salto' ['I jump, leap'])	103v	84	169	pāykiratu	210_L_f	'Saltar, pullar, ...'
C7	jrhuگیرadu (→ jrhuگیرen ['meter'] <sup>33</sup> ['to put'] <sup>34</sup> )	108r	89	178	iṭukiratu	94_R_j	'Por, metter: ...' <sup>35</sup>
C8	aluguiradu (→ aluguiren 'eu choro' ['I cry'])	112v	93	187	aḷukiratu	13_L_f	'Chorar'.
C9	canguiradu (→ canguiren 'veio' ['I behold'])	116v	96	195	kāṅkkiratu	281_L_o	'Ver'.
Copula	aguiradu (→ aguiren 'eu sou' ['I am'])	140r	114	220	ākiratu	46_R_i	'Ser, fazerse'

<sup>31</sup> The transcription of the *Arte* entries is based on Vermeer (1982) and frequently differs from the transcription given in Hein & Rajam, because of the difference in their readings or in their transcription scheme. For instance, concerning the item in row C4, Vermeer writes *etiRen* where Hein&Rajam write *etiRRen*. The MS which they transcribe has  $\text{e } \text{r} \text{ i } \text{r} \text{ e} \text{ n}$ . I have discussed that topic in Chevillard (2017: 113).

<sup>32</sup> The text appearing in this entry is not the VTCSPP text but a normalized transliterated form, obtained after removing the ambiguity in the original Tamil text: For instance, *viccuvatikkiratu* is a transliteration of  $\text{விச்சுவதிக்கிறது}$  which is the modernized form of the ambiguous original  $\text{விச்சுவதிக்கிறது}$ , in which the consonants without vowels are not indicated by the presence of a dot (*pulli*) over them.

<sup>33</sup> HH does not give a meaning for *jrhuگیرen*, but explains that *Este verbo jrhuگیرen tē muitas significações segundo a palavra cō que se ajunta, de que quisais adiante se falara* (Vermeer: 88). Hein & Rajam translate: 'The verb **irhuگیرen**, which perhaps will be discussed later on, has many meanings depending on the verb with which it is joined'. However, for the 'preterito' *jthen*, for the 'futuro' *jrhuieren* and for the 'Jmfinitiuo' *jrha*, we are provided with Portuguese translations which are, respectively, 'meti', 'meterei' and 'meter', rendered into English by Hein & Rajam as 'I put', 'I shall put' and 'to put'.

<sup>34</sup> See Hein & Rajam (2013: 178, fn. 341).

<sup>35</sup> The continuation of the entry says: [...] *uestesse [emphasis mine] este uerbo de uarias significações, cōforme o nome cō que se ajunta. ānai-y-iṭukiratu. jurar, kaṭṭalai-y-iṭukiratu. mandar, &c.* ['this verb **wears on itself** various significations, depending on the name with which it is combined. *ānai-y-iṭukiratu*. to swear. *kaṭṭalai-y-iṭukiratu*. to order, &c.'].

We are now going to compare, from a point of view as global as possible, the coverage of the Tamil verbal system which is performed in HH's *Arte* on the one hand and in AP's VTCSP on the other hand, although these two works do not have the same nature, because, from the point of view of the Tamil (ஊ) triglossia, we can indicate, in a nutshell, that:

- HH's *Arte* is the incomplete draft of a preliminary grammar, possibly summarizing the linguistic observations made during half a century, but mostly centered on the A-variety of Tamil, although containing a few clear indications of the fact that its author knew that the B-variety existed and was important.<sup>36</sup>
- AP's VTCSP is an incompletely lemmatized Tamil dictionary, with its entries ordered in the Portuguese alphabetical order. It is a mature work, based on 150 years of linguistic observations, which has in part moved away from being based exclusively on the A-variety, as is for instance visible on the rows C4, C5a, C5b & C6 of **Table 2**, where we see that the A-variety forms *etiratu*, *coliRadu*, *pu-duradu* and *paýRadu*, all found in column 2, have been replaced, respectively, by the B-variety forms *ēttukiratu/ēṟṟukiratu*,<sup>37</sup> *koḷlukiratu*, *pūtukiratu* and *pāykiratu*. However, the number of A-variety forms remains very important, and there are many cross-references between the two varieties.<sup>38</sup> Additionally, as we shall see, there are many entries inside the VTCSP which are exclusively C-variety forms. Several of those entries are explicitly said to belong to the language used by Tamil poets.

## 5. Taking the measure of the incompletely lemmatized VTCSP

We now return to the challenge of (virtually) lemmatizing the non-lemmatized VTCSP, a preoccupation which was my initial reason for moving away from the question asked in (3a) to a more general group of questions. Before however entering the lemmatizing challenge, I must provide in this section more information concerning the organization of the VTCSP, taking as a starting point, for the sake of continuity in the exposition, the treatment seen in the VTCSP of some of the verbs which were used as models in HH's *Arte*. I shall take as first object of my examination the VTCSP item which is referred to as 213\_R\_a in the second half of row C3a inside **Table 2**.

The corresponding entry, in image form, is made available here as **Figure 8**. That image also contains another VTCSP entry, namely 213\_R\_b.

We shall also refer in the discussion which follows, to still another VTCSP entry, namely 214\_L\_a, visible here in image form, inside **Figure 9**. Inside my current draft XML electronic edition of the VTCSP, those three entries are represented by three entry elements inside a file, which appear, in a simplified form, as follows.

<sup>36</sup> Later in his life, HH himself would compose a text in the B-variety of Tamil, namely the *Flos Sanctorum*. His awareness of the existence and of the importance of the C-variety Tamil (i.e. poetical Tamil) is clearly visible through the presence of many remarks such as 'de uerso' ['poetical'] in entry 478\_L\_i (செனனி 'Cabeça' ['head']), 'pal: de uerso' ['poetical word'] in entry 15\_L\_m (அமடலி 'Mundo' ['world']), etc. which indicate that those items are exclusively found in Poetry. It is also clear from the presence of entries describing some features of poetical activity and of entries containing references to several literary compositions.

<sup>37</sup> In the case of *ēttukiratu*, we can say that we have an intermediate form, standing between the A-variety form *etiradu* and the B-variety alternate variant form *ēṟṟukiratu*.

<sup>38</sup> Many of those cross-references between variants make use of the Latin word *quod*, which occurs in this role at least 490 times inside my XML text of the VTCSP.

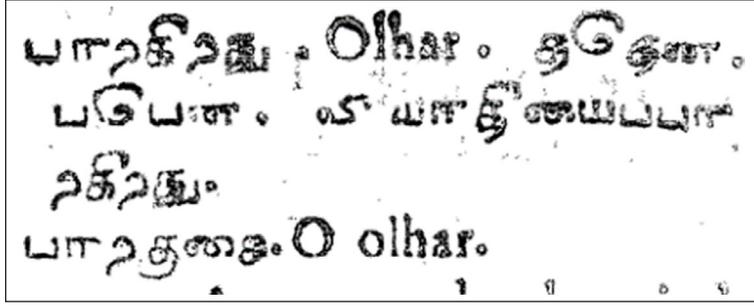


Figure 8: Entries 213\_R\_a & 213\_R\_b in the 1679 VTCSP.



Figure 9: Entry 214\_L\_q in the 1679 VTCSP.

- (4a) <entry coordinate = “213\_R\_a”>  
 <line\_length> 3</line\_length>  
 <entry\_head>ப # ற கி ற து</entry\_head>  
 <disambiguation> pārkiratu </disambiguation>  
 <CATEGORY> V</CATEGORY>  
 <entry\_body>Olhar.  
 <p>த @ த ன</p>.  
 <p>ப @ ப ன</p>.  
 <expression>வி ய # தி @@ ய ப ப # ற கி ற து.</expression>  
 </entry\_body>  
 </entry>
- (4b) <entry coordinate = “213\_R\_b”>  
 <line\_length> 1</line\_length>  
 <entry\_head>ப # ற கு @@ க</entry\_head>  
 <disambiguation> pārkukai </disambiguation>  
 <Under> 213\_R\_a</Under>  
 <CATEGORY> INF1</CATEGORY>  
 <entry\_body>O olhar. </entry\_body>  
 </entry>
- (4c) <entry coordinate = “214\_L\_q”>  
 <line\_length> 1</line\_length>  
 <entry\_head>ப # த த</entry\_head>  
 <disambiguation> pātta </disambiguation>  
 <Under> 213\_R\_a</Under>  
 <CATEGORY> PE</CATEGORY>  
 <entry\_body>C. que uio, ou uifta. </entry\_body>  
 </entry>

In this rather verbose collection of tags,<sup>39</sup> only some items are a transcription of what is in the printed book, and everything else belongs to the interpretive layer. These three entries, which occupy 5 printed lines in the VTCSP represent approximately 0.0177%

<sup>39</sup>This is in fact a simplified version of the actual encoding.

of the totality of the text which is encoded in XML, because the total number of printed lines occupied by the 16,215 entries is 28,227, with an average of 1.74 lines per entry.<sup>40</sup> Regarding the added interpretative layer, we can say the following:

- The coordinate attribute, which appears inside each entry head element, has been explained in the previous section.
- The disambiguation element is here to provide my interpretation of the ambiguous printed Tamil text.<sup>41</sup>
- The Under element, which indicates that one entry is secondary with respect to another entry, named by means of its coordinate,<sup>42</sup> allows us to make a virtual family out of several entries, if they are secondary with respect to the same primary entry, considered as the head of the family. This is the case here, and therefore {*pārkiṛatu*, *pārkkukai*, *pātta*} are made into a family of entries,<sup>43</sup> in which the head is *pārkiṛatu*.
- The CATEGORY element, which in those three entries contains respectively the values ‘V’, ‘INF1’ and ‘PE’, is used here for indicating which element in the paradigm an entry represents. We shall come back to this when explaining in which manner the VTCSP continues HH’s *Arte*, and in which manner it goes beyond it.
- The p element, which stands for ‘paradigm’, and which appears inside the entry body element, was intended for giving morphological information. However, it must have been difficult to use because the information given here, about the past and future forms of first person, is truncated. We simply read ‘*ttēn*’ and ‘*ppēn*’ —after disambiguation of the ambiguous sequences « *த @ த ன* » and « *ப @ ப ன* »— and have to guess whether it means that the past tense 1<sup>st</sup> person singular is *pāttēn* and that the future tense 1<sup>st</sup> person singular is *pāppēn*, which are the A-variety forms, or whether it means that those two forms are *pārttēn* and *pārppēn*, which are the B-variety forms.
- The expression element singles out an idiomatic expression which must have been considered as important but which is difficult to interpret now because it is not translated into Portuguese and no similar expression is found in modern dictionaries of Tamil. Since « *வி ய # தி @@ ய ப ப # ற கி ற து* » disambiguates as *viyātiyaip pārkiṛatu*, where *viyātiyai* is the accusative of *viyāti* ‘doença’ (436\_L\_j), i.e. ‘disease’, the literal translation is ‘to see the disease’, which might be a medical expression referring to the performing of a medical diagnosis. Some external confirmation would be necessary.

<sup>40</sup> In order to clarify the distribution, I shall add that, when interrogating my current XML database, I find that: among the 16,215 entries, the longest entry (which is entry 156\_L\_p) occupies 16 lines; only 12 entries have more than 10 lines; only 213 entries have more than 5 lines.

<sup>41</sup> Concerning the ambiguous Tamil writing system used in the VTCSP and other ancient texts, see for instance Chevillard[2015]. The presence of signs like ‘@’, ‘#’ and ‘@@’ inside the tamil passages is motivated by my desire to preserve VERBATIM the ambiguous character of the Tamil writing system of that period, which for instance does not distinguish between short e and long ē, both noted by the same signs. As an example, the sequence « @ க » can be read either *ke* (modern கெ) or as *kē* (modern கே). Additionally, an even more difficult situation arises because the same sign, represented here by ‘#’, can be read either as *r*, or as *ra*, or as the marker of a long -ā. For example, « க # » can be read either as *kar* (modern கர்), or as *kara* (modern கர்), or as *kā* (modern கா). Finally, because the long *ō* and the short *o* are not distinguished in writing and their presence in combination with a consonant is noted by two diacritics, namely ‘@’ before the consonant sign and ‘#’ after the consonant sign, the sequence « @ க # » can be read in six different manners: *ko* (modern கொ), *kō* (modern கோ), *ker* (modern கெர்), *kēr* (modern கேர்), *kerā* (modern கெர்), *kērā* (modern கேர்).

<sup>42</sup> Reformulated here verbosely, for the sake of clarity, the two instances of the tag <Under>213\_R\_a </Under> inside (4b) and (4c) ARE the *pointers* which *subordinate* those two entry elements to the other entry element visible in (4a). That entry element becomes, thanks to the presence of those two pointers, the *head of the family* thus created.

<sup>43</sup> More information will be provided concerning the order of magnitude of the task inside sections 9 and 10.

In the remaining part of this section, I shall make a few remarks, intended as supplements both to what has been said concerning the CATEGORY element and to what has been said concerning the p element. These remarks will try to establish a link between the description of the Tamil verbal conjugation as seen in HH's *Arte*, and the verbal conjugation as seen in the VTCSP. We shall revisit some of the apparent contradictions, connected with the tension which must always have been present in the description of diglossia, not to talk about the added layer of complexity which comes from the triglossia.

The first remark concerns my choice of *pārkiṛatu* as the head of what I have called the {*pārkiṛatu*, *pārkiṛukai*, *pāṭta*} family of entries. What is the criterion for deciding that one item is the natural choice for such a role? As we have seen in the previous section, and notably inside **Table 2**, in the system used by HH in his *Arte*, the head of the 'family' in the description of a verbal paradigm was the 1st person present tense form, which I referred to as the primary  $\alpha$  entry. In the case of the lexeme which we are currently examining, the head was, in HH's transcription system, the form *paṭhquiren* 'eu olho' ['I look at'], which in a disambiguation XML element would appear as *pārkiṛēn*. In that system, *pārkiṛatu* is 'sub-tense'  $\alpha$ -J, if we make use of H&R's terminology, slightly adapted. As for the form *pāṭta*, it is also a sub-tense, although it is part of another group because it is 'preterite tense, subtense J',<sup>44</sup> or, more compactly, the form  $\beta$ -J, enumerated as part of a succession of sub-tenses, morphologically derived from the tense  $\beta$ , i.e. the 'preterite' (HH's *preterito*), which is *paten* 'eu olhei' ['I looked at'], which would be transcribed inside the disambiguation element adopted here as *pāṭṭēn*, or possibly as *pāṭṭēn*, as discussed earlier in this section. To sum up, a system in which verbal forms are enumerated as {*paṭhquiren*, ..., *paṭhquiratu*, ..., *paten*, ..., *pata*, ..., *paṭhpen*, ...} has been replaced by a system where the enumeration is {*pārkiṛatu*, ..., *pāṭṭēn* (or *pāṭṭēn*), *pāṭta*, ..., *pārppēn* (or *pārppēn*)}. To this, which might appear to the reader as speculative, can be added the massive evidence which results from statistical evidence made on a data base consisting of 16,215 entries, inside which every entry element whose entry head is a verbal form ending in *-kiṛatu* or *-kiratu* or *-ratu* has been tagged by means of a CATEGORY element containing the string 'V'<sup>45</sup> (for 'verb') and in which other entries semantically related to them, in a manner to be explained, have a CATEGORY element with a string value which is 'PE' (for *peyar eccam*),<sup>46</sup> or 'INF1', or still many other possibilities. The statistic concerning those three types, plus another one to be discussed shortly, is as follows.

At the end of this already long section, and before examining other aspects in the coming sections, I shall briefly comment on the items contained in the last two lines of that chart. Concerning the label INF1, which I have provided for the form *pārkiṛukai* 'O olhar' ['the looking at'],<sup>47</sup> I should first observe that this form does not seem to be mentioned inside HH's *Arte*. It might be more typical of the B-variety inside the Ⓢ triglossia. Nevertheless, HH himself has made use of it in a book which he composed, *Flos Sanctorum எஸ்ஸி அடியார் வரலாறு*, which is a collection of the lives of Saints. It is also found in the C-variety. As can be seen from the statistics in **Table 3**, items having this CATEGORY label form appear 1,121 times inside the VTCSP, as per my most recent count. This is certainly an argument for saying that the VTCSP is a fully diglossic dictionary.

<sup>44</sup> Compare, in Hein & Rajam (2013), the page 93 (where *vichuvadita* is discussed) and the page 128 (where *pata* is mentioned).

<sup>45</sup> In practice, this means that such entries contain the sequence <CATEGORY>V</CATEGORY>.

<sup>46</sup> This traditional Tamil grammatical term, which usually translates into English as 'relative participle', will be discussed in Section 7, when we revisit example (4c).

<sup>47</sup> One of the biggest difficulties, when translating Portuguese glosses of Tamil words into English, is that the target language, i.e. English, is not always well equipped for faithfully conveying the minimal distinctions which Portuguese can perform by using the definite article *o*, which is seen in the difference between *pārkiṛatu* 'olhar' ['to look at'] and *pārkiṛukai* 'O olhar' ['the looking at']. Almost all the 1,121 items categorized as INF1 in the VTCSP are glossed in this manner, with a few rare exceptions.

**Table 3:** Statistics concerning four of the possible values of the CATEGORY element in the VTCSP.

CATEGORY	Examples	Total number of such items in the VTCSP	
V	<i>pārkiṛatu</i> (213_R_a), <i>iṭukiṛatu</i> (94_R_j)	3,168 items	A or B
PE	<i>pāṭta</i> (213_R_b), <i>iṭṭa</i> (94_L_n)	719 items	A, B or C
INF1	<i>pārṅkukai</i> (214_L_q), <i>iṭukai</i> (94_R_l)	1,121 items	A, B or C
INF2	<i>iṭutal</i> (94_R_k)	168 items	C only

As for the last row in **Table 3**, it contains the statistics for a much less frequent form, which is labeled as ‘INF2’ and is found 168 times inside the VTCSP. That form is characteristic of ancient poetical texts, belonging to the C-variety, i.e. ‘Classical Tamil’. The example chosen is taken from the family  $\{iṭukiṛatu, iṭṭa, iṭukai, iṭutal\}$ , which we have already encountered on the row C7 of **Table 2**, representative of a time when the head of the family was *jrhuḡiRen*. The fact that such a form as *iṭutal* has its own autonomous entry inside the VTCSP shows that Proença had not completely separated himself from the explorations made by one of his predecessors, Ignacio Bruno, to whom he refers in his introduction.<sup>48</sup>

## 6. The one-sided debate between Ziegenbalg and Beschi concerning the two types of *infinitivus*

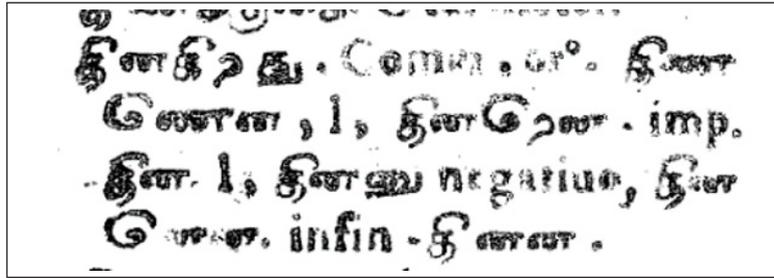
In this section, we shall move forward in time to a brief examination of the one-sided debate<sup>49</sup> between Ziegenbalg and Beschi, which has already been alluded to inside the section 2 of this article, as an anticipated comment on the question numbered (3a) in that section, which refers to the modern unanimity among English-speaking descriptors of Tamil, in their use of the English technical word ‘infinitive’, which was in the beginning simply a translation of an earlier use of the Latin technical word ‘infinitivus’.

Having now examined the prehistory of the  $\{L\backslash t\}$  type grammars of Tamil, thanks to our brief explorations of HH’s *Arte* and AP’s VTCSP, which both fall under the  $\{P\backslash t\}$  type, we know that the use of *infinitivus* (with respect to Tamil) was itself preceded by the use of *infinitivo*, a term which occurs rather frequently in HH’s *Arte*, because it is considered as one of the main tenses (see the 8<sup>th</sup> column inside **Table 1**), from which 18 sub-tenses are morphologically derived in the scheme presented by HH’s *Arte*.

If we now consider the VTCSP, the use of the *infinitivo* label is more difficult to detect, but having now at my disposal the integrality of its text, I have located one occurrence of the abbreviation ‘inf.’ (as a label for the form *uṇṇa* inside entry 391\_R\_l, *uṅkiṛatu*) and two occurrences of the abbreviation ‘infin.’ (as a label for the form *tiṇṇa* inside entries 361\_L\_d *tiṅkiṛatu* and as a label for the form *vayya* inside entry 405\_L\_o *vaykiṛatu*). See for instance the entry 361\_L\_d, as reproduced in **Figure 10**, where we can see the use of four of the main *tenses* mentioned in **Table 1**, namely ‘Preterito’ (pr<sup>o</sup>) *tiṅrēṇ*, ‘imperatiuo’ (imp.) *tiṅ* or *tiṅṇu*, ‘negatiuo’ *tiṅṇēṇ* and ‘infinitiuo’ (infin.) *tiṅṇēṇ*. On the basis of those three examples, we can conclude that the usage of the technical term *infinitivo* is identical between HH’s *Arte* and the VTCSP, and has not changed during the more than 100 years which have elapsed between the two works.

<sup>48</sup> An English translation of AP’s reference to Ignacio Bruno is found on page 11 in the Facsimilé published in 1966 by X. Thani Nayagam.

<sup>49</sup> Ziegenbalg (1682–1719) had already been dead for a number of years, when Beschi (1680–1747) was writing his grammar in 1728. It would be printed ten years later, in 1738, in Tranquebar, which is the place where Ziegenbalg had been posted during his stay in India and where he is buried.



**Figure 10:** VTCSP Entry 361\_L\_d (tinkiratu ‘comer’).

However, contrary to what we have seen for other elements of the morphology, such as PE, or INF1, the number of entries which could be labeled as VE2, which is my {T\t} label for the standard *infinitiuo* is very small and seems to be restricted to a number of adverbs (currently 18 items) which are in fact grammaticalized uses of *infinitives*.<sup>50</sup> What we find on the other hand is the use of *infinitives* as components in entries which I have labeled as VE2-AUX2 constructions.

Moving now to the central topic in this section, we can say that the formal type which I have referred to until now, following Hein & Rajam, as the sub-tense  $\alpha$ -J, starting at **Table 2**, did not have a convenient designation in HH’s *Arte* but rose to such a prominence, for reasons which could be further debated, that we find 3,168 entries in the VTCSP falling under that type, a proportion which is almost 20% of the total number of entries.<sup>51</sup> It seems that Ziegenbalg proposed to innovate<sup>52</sup> by referring to forms such as VICUVĀCIKKIRATU by means of the *Infinitivus Substantivus* label, as can be seen in **Figure 11**, where, at the same time, he proposes to use another technical designation, namely *Infinitivus Absolutus*, for the sake of referring to forms such as VICUVĀCIKKA, which is what HH’s *Arte* refers to by means of the *infinitiuo* label (see the  $\zeta$  column inside **Table 1**).

Beschi’s disagreement with this suggestion can be clearly seen in his 1738 grammar, on page 95, inside parag. 107, which is found inside §. II, *De Nomine*, which is itself

<sup>50</sup> As an example, I could have been tempted to add to the {pārkiratu, pārkukai, pātta} family of entries examined earlier the item *pārka*, which is entry 213\_L1, if it were not clear from its Portuguese gloss, namely *Mais, comparatiuo, com acusatiuo*, and also from still existing usage, that it is a grammaticalized form of the lexeme *pārkiratu*, which plays in comparative constructions in Tamil the role that *than* plays in comparative constructions in English.

<sup>51</sup> As has already been explained, this means in practice that they have been assigned the tag <CATEGORY>V</CATEGORY> because of being verbal forms ending in *-kkiratu* or *-kiratu* or *-ratu*.

<sup>52</sup> I was wondering whether Ziegenbalg had been following some intermediate Portuguese missionary, when Cristina Muru gave me very clear elements of answer based on her deep knowledge of the works of Aguilar [b. 1588] and of Da Costa [c.1610–1673] —see for instance Muru (2020)— and also on the remarks made by Jeyaraj, with which she agrees, that Ziegenbalg’s grammar is based on Da Costa’s *Arte* (Jeyaraj 2010: 20: ‘Ziegenbalg made use of Da Costa’s Tamil grammar entitled *Arte Tamulica*’). Jeyaraj identified a copy of Costa’s grammar at the British Library (ms OC Sloane 3003) which is almost identical with the copy of Costa’s *Art* found inside Ms 50 (previously 34) which she has transcribed and translated. Ziegenbalg must have left in London, when he stopped there during his second journey to India, a copy which he had of Da Costa’s *Arte Tamulica*. As is clear from the indications given by Cristina Muru, Aguilar (as seen in MS Cod.Orient. 283, fol.29 v, lines 31–33) does not innovate on the terminological side by creating two distinct sub-designations inside what he calls the *Modo infinitivo*, where he enumerates (as transcribed by Cristina Muru):

- {31} Modo infinitivo.
- {32} Præsente vichuvadica. Crer, crendo, pera crer.
- {33} Vichuvadicradu. Isto que he crer.

On the other hand, inside Da Costa’s grammar (MS 50, previously MS 34), there is a clear distinction made (as seen in the transcriptions communicated to me by Cristina Muru) between, on the one hand, a *Modo infinito absoluto* (fol. M-34–28) and an *Infinito Substantivo* (fol. M-34–29). It is to be hoped that the texts of those two grammars will be published by Cristina Muru in the near future and will become available to the larger circle of all those who are interested in the global history of terminology. Another possibility which could be explored is the influence on the refinement of terminological distinctions of the study at that period in Europe of languages such as Hebrew.

	<i>Infinitivus absolutus.</i>	
உசு உாசீகக		Credere
உசு உாசீகக தககக தாக		idem.
	<i>Infinitivus Substantivus.</i>	
உசு உாசீககி மது		Credere
உசு உாசீததது		Credisse
உசு உாசீபபது		Crediturum esse
உசு உாசீயாதது		Non credere.

Figure 11: Ziegenbalg (1716: 72) making the distinction between two types of infinitive.

contained in *Caput IV, De Syntaxi*. Beschi writes the following (accompanied here by the 1848 translation by Mahon):

- (5a) Ex quolibet verbo fiunt prætereà plura verbalia, quæ ad modum nominum declinantur, et ... (Beschi 1738: 95, parag. 107)  
From every verb are formed moreover many verbals, which are declined in the manner of nouns, and ... (Mahon 1848: 79)
- (5b) ... 1) ex participiis præsentis et præteriti fiunt verbalia செயகிறது, படிக்கிறது, et செயதது, படித்தது &c. ... (Beschi 1738: 95, parag. 107)  
... 1. From the participles of the present and præterite come the verbals *ceykiratu*, *paṭikkiratu*, and *ceytatu*, *paṭittatu*, &c. ... (Mahon 1848: 79)
- (5c) ... Hinc apparet, quàm improprie hæ voces v.g. செயகிறது, சொல்லுகிறது &c. dicantur voces infinitivi: quod certè fallum est, ... (Beschi 1738: 95, parag. 107)  
... Hence it appears, how improperly these words, e.g. *ceykiratu*, *collukiratu*, &c. are called words of the Infinitive: which certainly is not true, ... (Mahon 1848: 79)
- (5d) ... Cùm autem has dicunt esse, ut sunt, voces participii præsentis, cum addito pronominis அவன, அவள், ex quo fiunt nomina appellativa hominum: sic et செயகிறது est vox ejusdem participii cum addito pronominis அது, ex quo fit nomen verbale, quod per suos casus declinatur et non correspondet infinitivo Latinorum ... (Beschi 1738: 95, parag. 107)  
... But since they say, that these are, as they are, words of the participle present, with the addition of the pronoun *avan*, *aval*, from which are formed appellative nouns of men: so also, *ceykiratu* is a word of the same participle, with the addition of the pronoun *atu*, from which comes the noun verbal, which is declined throughout its cases, and does not correspond with the Latin Infinitive, ... (Mahon 1848: 79)
- (5e) ... nisi quando hoc sumitur tanquam nomen, ut in hac propositione, *studere bonum est*, quam vertam, படிக்கிறது நல்லது. At in hoc sensu infinitivo Latinorum correspondent cætera quoque verbalia, de quibus infra, படித்தல, படிப்பு &c. unde illa propositio potest reddi, படித்தல நல்லது படிப்பு நல்லது, படிப்பது நல்லது &c. Numquid propterea hæc omnia voces infinitivi dicenda sunt? (Beschi 1738: 95, parag. 107)  
... except when this is taken as a noun, as in this proposition, *to study is good*; which I may translate, *paṭikkiratu nallatu*. But in this sense other verbals also, of which hereafter, correspond with the Infinitive of the Latins; *paṭittal*, *paṭippu*, &c., whence that proposition may be rendered, *paṭittal nallatu*, *paṭippu nallatu*, *paṭippatu nallatu*, &c. But pray, are all these therefore to be called words of the Infinitive? (Mahon 1848: 79)

Although it is difficult to comment on this, because there is a large degree of arbitrariness in terminological choices, it seems that Beschi was successful, in front of the audience which he addressed. We shall not however directly comment now on what he said. We shall rather, in the following section, revisit some of our former steps, both in connection with the 16<sup>th</sup> cent. *Arte* and the 1679 VTCSP, in order to clarify Beschi's references to participles.

## 7. Understanding relative constructions in Tamil, as seen through the VTCSP and through HH's *Arte*

In this section, we shall first revisit the item (4c) which was provided in section 5 but on which we have not yet commented, except when relating its entry head, namely *pātta*, with the example given for sub-tense  $\beta$ -J of conjugation 3a in the *Arte* and except when indicating that the CATEGORY label chosen for it, namely PE, was an abbreviation of a traditional Tamil technical expression which is *peyar eccam*, known in English as 'Relative Participle'. I shall now add to this that inside traditional {T\t} grammars, the PE label is the symmetrical of the VE label evoked in (3a) and (3b), because if the VE is an incomplete form (*Eccam*) which needs a *Vinai* (verb) in order to become syntactically complete (*murru*), similarly the PE is an incomplete form (*Eccam*), but it needs a *Peyar* (noun) in order to form a complete syntagm. To which I shall add that if we choose a noun N in order to make the PE *pātta* 'complete', that N will come after *pātta*, resulting in the Noun-Phrase '*pātta* N', which we still have to translate, but we shall make use of the VTCSP entry 214\_L\_a, alias (4c), for that purpose, reproduced below in a simplified, tag-less form, (6a), which is as follows:

(6a) *pātta*. 'C. que uio, ou uifta'. [entry 214\_L\_q, under entry 213\_R\_a, *pārkiṛatu*, 'olhar'].

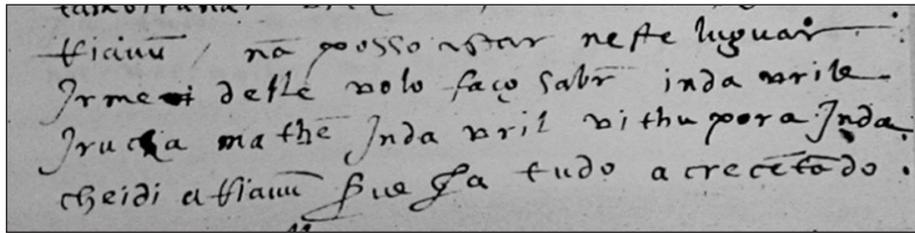
The two features on which I must comment now are the presence of the place-holder 'C.' (for Portuguese *cousa*, i.e. 'thing, entity') and the presence of a double translation. I shall react to the double translation by duplicating the entry as (6a1) and (6a2). Additionally, I shall make the place-holder role of 'C.' explicit by inserting a phantom (N) in the Tamil, Portuguese and English sequences which translate each other. The result is:

(6a1) *pātta* (N). 'C. (=N) que uio' ['entity (N) which saw']

(6a2) *pātta* (N). 'C. (=N) uifta' ['entity (N) [which was] seen']

What this linguistic situation may have suggested to the compiler of the VTCSP is the fact that a form like *pātta* was the close equivalent of both the Latin Past Passive Participle *visum* and of the Latin Present Active Participle *videns* (if we overlook the tense difference) because in (6a2) the head noun N stands for the object of the action whereas in (6a1) the head-noun N stands for the agent of the action. The ambiguity caused by this double equivalence must have appeared to the successive compilers of dictionaries as very striking. This must be the reason why, as we saw in **Table 3**, the VTCSP contains 719 entries tagged as PE, to which we can add 67 entries tagged as PE\_pr (Present relative participle) and 77 entries tagged as PE\_neg (Negative relative participle). I could of course have CATEGORY-tagged the PE as PE\_past (Past relative participle), but since it is so much more frequent than the other two, that would have been a misleading symmetry.

How, in a given situation, we must translate the noun-phrase *pātta* N depends of course on a number of parameters, such as the nature of N (is it agentive? is it inanimate? etc.) and such as the context. I must however hasten to add that giving the choice between these two translations is, as a matter of fact, insufficient: there are other possibilities, in



**Figure 12:** Extract from folio 43r in Codex 3141 (BNP, Lisbon) [containing text of example (6b), preceded & followed by other words].

addition to these two, as was in fact clear already to HH, when he composed the *Arte*. We can see that in the following example, extracted from folio 43r (see **Figure 12**), where HH explains the behaviour of the form which I have tagged as PE<sub>pr</sub> and to which H&R refer to as subtense  $\alpha$ -D, by means of two examples from which I shall now reproduce the second:

- (6b1) *inda vrile jručamathē, jnda vril vithupora jnda cheidi aRiauū* ‘Nã posso estar neste lugar, jrmeei delle, volo faço saber’ (Vermeer 1982: 36) (see original in **Figure 12**) [‘I can’t be in this place. Leaving it, I inform thee’] [Hein & Rajam 2013: 84–85]

This example can be written in modernized transcription as follows:

- (6b2) *inta ūrilē irukka māttēn. inta ūril viṭṭup pōra inta ceyti ariyavum.*  
[‘I shall not stay in this village. Please be informed of the fact that I shall go, leaving from this village.’] (My translation)

Inside this example, the present relative participle (alias PE<sub>pr</sub>) form *pōra* is followed by a head, which is the noun-phrase *inta ceyti* (‘this fact’, or ‘this action’, or ‘this information’). As for the relationship between the relative participle and its head, the easiest way to explain it seems to use the language of traditional {T\t} grammars, such as the *Tolkāppiyam* (see **Figure 5**). Those grammars explain that the noun which is the head-noun for a PE (*peyar-eccam*) ‘relative participle’ can have as a referent:

- the agent of the action expressed by the PE, as is the case in (6a1)
- the object of the action expressed by the PE, as is the case in (6a2)
- the time of the action, as would be the case in [...] *pātta nēram* (‘the time when [...] saw’)
- the location of the action, as would be the case in [...] *pātta ūr* (‘the village where [...] saw’)
- the instrument for the action, as would be the case in [...] *pātta kaṇṇāṭi* (‘the pair of glasses with which [...] saw’)
- the action itself, as would be the case in [...] *pātta ceyti* (‘the fact that [...] saw’) and as is the case in example (6b2)

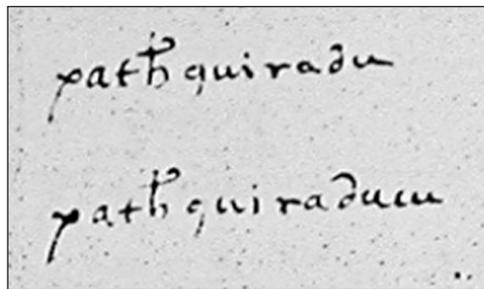
More details can of course be obtained by studying the commentary to these grammars,<sup>53</sup> which were written for describing the C-variety of Tamil, but which are nevertheless applicable in many respects (but not all) to the B-variety and the A-varieties of Tamil, even now.

<sup>53</sup> See Chevillard (1996: 362–365), which contains a French translation of Cēṅāvaraiyar’s commentary on the *Tolkāppiyam* sūtra TC234c.

These points being established, we are now in a better situation for understanding some of the elements evoked by Beschi, which were reproduced in Section 6, from (5a) to (5e). When he refers in (5d) to ‘words of the participle present, with the addition of the pronoun *avan*, *aval*, ...’, he is in fact proposing an analysis of words such as *ceykiravan*, *ceykiraval*, ... as a combination of the PE\_pr form *ceykira* with Human pronouns such as *avan* ‘he’, *aval* ‘she’, ..., considered as the head-nouns for *ceykira*. Since he does not translate, it is not completely clear which role he has in mind for the head pronoun. Are those heads meant to refer to the agent of the action, in which case we must translate them by ‘He who does’, ‘She who does’, ... ? Are those meant to refer to the object of the action? This is much more rare, if the head noun refers to a human being but not strictly impossible. When however Beschi continues his analogy and declares in (5d) that ‘*ceykiratu* is a word of the same participle, with the addition of the pronoun *atu*,’ we are now faced with a greater number of possibility. The *atu* ‘that-NEUTER’ which is the head pronoun incorporated inside the fusion form *ceykiratu* (= ‘*ceykira* + *atu*’) can have any of the roles enumerated above.<sup>54</sup> It can refer to the agent and mean ‘that which does’. It can refer to the object and mean that ‘that which is done’ ... It can refer to the action itself and mean ‘that which is the fact of doing’.<sup>55</sup> This is the reason why Portuguese lexicographers thought it was a good equivalent for the Portuguese infinitive and chose it as the citation form in their dictionaries, a tradition which continued for several centuries.<sup>56</sup>

### 8. How should one write *pārkiṛatu*? (between phonetics & orthography)

In this section, we shall return to the topic of the Tamil ⚙️ triglossia, through an observation concerning changing orthographies. The occasion for this remark, is the observation that there is a difference of spelling between Ziegenbalg and Beschi, concerning the word *pārkiṛatu*, which has already been one of our targets of observation, inside section 5, where it can simultaneously be seen in **Figure 8**, which contains an image extracted from VTCSF page 213, and in (4a), which is a transcription of the content of that image. Additionally, the word also appeared on row C3a of **Table 2**, where it was spelled *paṭhquiradu*, a spelling which tries to be faithful to what is observed in the Lisbon BNP MS of HH’s *Arte*, here visible on this page as **Figure 13**, where it is accompanied by another form, *paṭhquiraducu*, which would be written in normalized transcription as *pārkiṛatukku*, and is almost identical, except for the final additional suffix.



**Figure 13:** Paṭhquiradu on folio 83r of Codex 3141 (BNP Lisbon).

<sup>54</sup> This is in fact what makes possible the existence of what Lindholm (1972) has called the *Tamil Cleft Sentences*.

<sup>55</sup> Compare this with the gloss *Vichuvadīradu. Isto que he crer*, found in Aguilar’s grammar (MS Cod.Orient. 283, fol.29 v, lines 33), which is quoted in Note 52.

<sup>56</sup> See also the argument mentioned at the end of Note 21, for choosing the 3<sup>rd</sup> person neuter singular, rather than the 1<sup>st</sup> person singular, as a citation form. The MTL decided, on the contrary, to make use of the C-Tamil form ending in *-tal* as a main citation form. See an example of that form, labeled as INF2 inside Table 3. Several modern dictionaries make use of the verbal root (which is also the imperative), as a main citation form. This is however problematic in the case of paired verbs, because two entries in the MTL become one entry in those dictionaries.

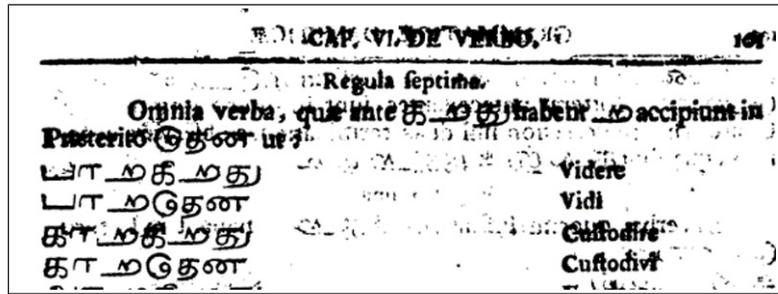


Figure 14: Ziegenbalg 1716: 101, Top.

Coming now to Ziegenbalg, we have an attestation of the same word on page 101 of his grammar, where we can see, if we examine Figure 14, a rule, formulated by Ziegenbalg, which states.

- (7a) Regula septima.  
 Omnia verb, quæ ante கிறது habent ற accipiunt in Præterito தென ut:  
 பாறகிறது Videre  
 பாறதென Vidi (Ziegenbalg 1706: 101)
- (7b) Seventh rule:  
 All the verbs which before *kiratu* have *r* receive *tēn* in the preterite, thus:  
*pārkiratu* to look at  
*pārtēn* I looked at (My translation)

Interestingly, a 2010 book, *Tamil Language for Europeans: Ziegenbalg's Grammatica Damulica (1716)*, by Daniel Jeyaraj which presents itself as a translation of Ziegenbalg's grammar, contains the following translation, which can be usefully compared with my translation.

- (7c) Seventh Rule [Page 101]  
 All verbs which have -தல்[-tal] before ர[r] take -தேன் [-tēn] in the past, like this  
 பார்த்தல்[pārttal]<sup>171</sup> to see, look at  
 பார்த்தேன்[pārttēn] I saw  
 [Footnote 171: 'பாறகிறது'['pārkiratu'] ] (Jeyaraj 2010: 133)  
 [[NB: the transliterated items between square brackets have been added by me (jlc)]]

The reader of this article may try to imagine the nature of a collective linguistic pressure for correctness which is so strong that it induces the 21<sup>st</sup> century translator of an 18<sup>th</sup> century important historical text to modify that text, while translating it, in such a manner that the translation becomes misleading for anyone who does not have access to the original. We can indeed verify that inside (7c):

- the rule has been changed, but we are not informed of the change
- the original target of the rule has been replaced by another target, and put in a footnote.
- the result of the rule has been changed, but we are not informed of the substitution.

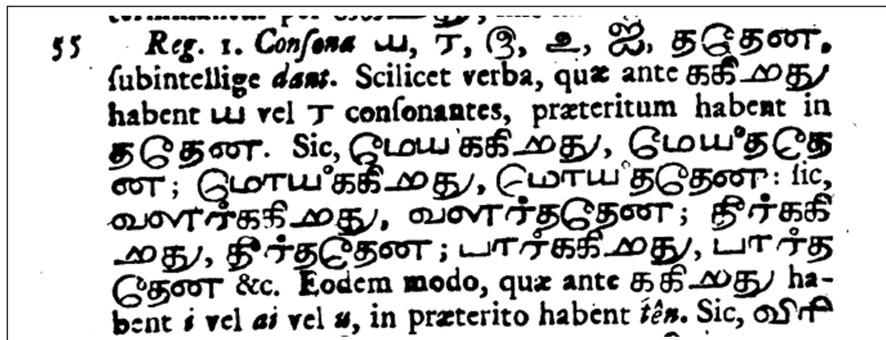
I can only conclude that the translator must have thought that the prestigious original was written in such a sub-standard form of language, that one should not run the risk of having the prestigious original setting a possibly contagious bad example. Without making additional comments on (7c), I shall now turn to the text of Beschi's 1738 *Grammatica Latino-Tamulica*, as it was printed in Tranquebar. We find inside this text, on page 50, in

paragraph 55, a rule which is reproduced here in **Figure 15**. That Rule explains how to form the preterite of the verbs whose stem ends in *y*, *r*, *i*, *u* or *ai*. Inside the image, on the 7<sup>th</sup> line, we can see one of the targets of the rule, which is பார்க்கிறது [to be normalized as *pārkkiratu*]. The result of the application of the *First preterite rule* appears on the 7<sup>th</sup> and 8<sup>th</sup> lines and is பார்த்தேன் [to be normalized as *pārttēn*]. As we can see, the target is identical with the value which Daniel Jeyaraj has substituted to Ziegenbalg target, which was பாற்தென் [to be normalized as *pārtēn*]. I should add that there are several levels in ungrammaticality from the point of view of Traditional Tamil grammarians. For them:

- both *pārkkiratu* ‘to look at’ and *pārttēn* ‘I looked at’ are acceptable phonetically and are the correct forms in the B-variety
- *pārkiratu* is a phonetically acceptable combination of sounds, which is distinct phonetically from *pārkkiratu*, but which is NOT the correct form, inside the B-variety, if the intention is to convey the meaning ‘to look at’, and can only appear as a mispronunciation.
- *pārtēn* is a succession of written symbols, which is the transcription of a forbidden combination of sounds, because (alveolar) *r* is NEVER followed by (dental) *t*, according to the rules of the C-variety, as has been stated in the *Tolkāppiyam* (see **Figure 5**), and because the phonetic rules of the C-variety are supposed to govern the B-variety. The educated guess of an educated Tamilian is that whoever wrote *pārtēn* must have made a mistake and wanted to write *pārttēn*.

Of course, a descriptive linguist may have a different point of view and remark that since the A-variety component of the Tamil (☉) triglossia is in fact a collection of dialects and that each of them is likely to have its own phonological system, it should be possible to make sense of at least some of the aberrant combinations of letters found in ancient written documents but that is of course a tall order, which will not be attempted here, although that should be at least one of the possible outcomes of my forthcoming edition of the 1679 VTCSP.

A final question must now be dealt with in this section concerning *pārkkiratu*: is it possible that this usage is limited to Christian documents and is due to the (bad) influence of Christian Missionaries? In order to answer this question, I have been lucky enough to receive the help of a doctoral student of Hamburg University, Neela Bhaskar, who is currently preparing a Ph.D. on a śaiva collections of stories known as the *Tiruvīlaiyāṭarpurāṇam*, which narrates the Sixty-four ‘Sacred Sports’ (*tiruvīlaiyāṭal*) of Śiva in Madurai. She has easily located for me several occurrences of the spelling *pārkkiratu* inside a group of śaiva MSS on which she works.<sup>57</sup> I shall now display one of those examples in **Figure 16**. Although the MS is not very easy to read because it is incised but not inked, one can clearly see that the first of



**Figure 15:** Beschi 1738: 50, (Cap. III, §. II, parag. 55, Præteritum, Regula 1).

<sup>57</sup> I express here my gratitude to her.



**Figure 16:** *pārkiṛatu* in extract from Folio 55 r in MS *Indien 291* (BnF, Paris).

the two lines contains the written form பாரகிரது [*pārkiṛatu*] preceded by the sequence த்தைப [ttaip] and followed by the glyph ச [cu].<sup>58</sup>

### 9. Going beyond families: the case of the causatives

In this section and in the following one, I shall return to what could be called the *global numerical perspective*, where we attempt to measure the progress made by Portuguese-speakers between the 16<sup>th</sup> and the 17<sup>th</sup> century in the twin tasks of learning and describing Tamil. That progress is seen in the comparison between HH's *Arte* and Proença's VTCSP. A preliminary evaluation of the content of the VTCSP, *as a whole*, has already been provided here in **Table 3**, inside Section 5. That table contains basic statistics on the number of occurrences of four CATEGORY labels, starting with the omnipresent V label, which is used 3,168 times in my tagging of the VTCSP in order to characterize those forms which BZ referred to as *Infinitivus Substantivus* (see section 6), and continuing with three other labels, PE (719 occ.), INF1 (1,121 occ.) and INF2 (168 occ.), which were also commented upon in the previous sections. When we compare this with the fact that Vermeer's 1982 critical edition of HH's *Arte* includes as ultimate section, on pp. 155–166, an 'Index of the Tamil words occurring in the grammar' which contains 309 entries, among which ca. 120 are identifiable as verbs, we can see indeed how much more information had been made available.

The huge progress, obtained within one century of descriptive efforts, when going from a list of 120 verbs in HH's *Arte* to a list of 3,168 V-tagged VTCSP entries, which we can also call 'preliminary verb candidates', can be evaluated more precisely by providing more information on the grouping of items in 'verb families' which was first introduced inside section 5, and which was performed by means of the Under element, seen for instance in (4b) and (4c). More specifically, my current XML file of the VTCSP, contains 3475 occurrences of the Under element, pointing towards 1,421 distinct family head. This means that a family created in this manner contains one head accompanied by an average number of 2.45 items. To this must be added the fact that there are also currently 1478 entries which are tagged as V, but which are not related to other verbal forms, either as target or as source of an Under element. Each of those 1478 entries is the only member in a *singleton family*. This is summarized here in the form of a chart, which contains my current counts, inside the VTCSP edition *work in Progress*.

Although this table may appear complex, the position of the 6,374 entries which it describes is determined simply by answering three questions, the first one being purely morphological, and the other two requiring a greater command of grammar.

- does the entry end in -kkiratu or -kiratu or in -ratu? (if yes, it is labelled V)
- does the entry belong to the paradigm of another entry, considered as main citation form?<sup>59</sup>

<sup>58</sup>The sequence *ttaip* is the end of the preceding word *pirākattai*. The *p* is doubled because of sandhi. The following word starts with *cu*.

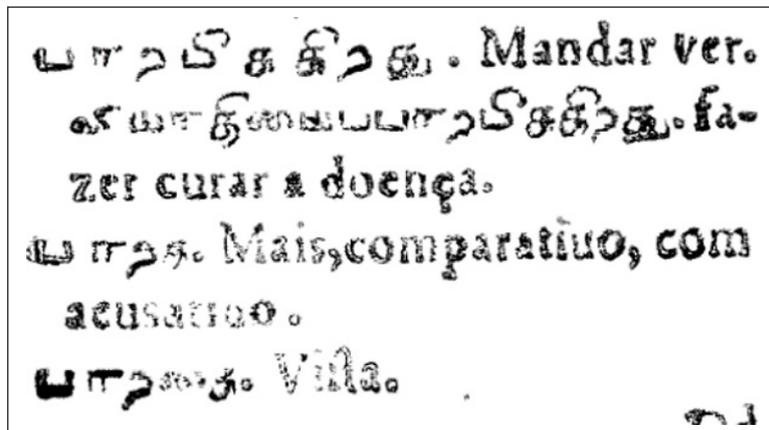
<sup>59</sup>Taking just one example, namely the situation involving (4c) and (4b), the person performing the tagging must have identified item 214\_L\_q as being the 'Relative participle' (alias PE, i.e. *peyar eccam*) of item 213\_R\_a. In terms of grammar to be mastered, that would mean that the person doing the tagging must have mastered the formation of the subtense β-J, discussed in section 5.

- is the entry the target of an Under element link contained in another entry?

As soon as we introduce other parameters however, things can become more complex. My purpose in this section is to illustrate that it would be premature on the basis of **Table 4** to conclude that the VTCSP contains 2,899 distinct verbs. I shall for that purpose provide basic information concerning the elements of a large subgroup which is contained inside the group of 2,899 heads of families VTCSP entries. This subgroup, which possesses 554 elements, contains all the VTCSP entries which I have labelled CAUSAT (for *causative*).<sup>60</sup> It will be briefly illustrated here, in a convenient manner, by the entry 213\_L\_k, alias *pārpikkīratu* ‘Mandar ver’ [‘to tell to see’] (See **Figure 17**). That entry is related, in a natural manner, with the entry 213\_R\_a (*pārkkīratu* ‘Olhar’ [‘to look at’]), which has been the focus of our observations in some of the preceding sections, starting with section 5.

**Table 4:** Preliminary repartition of the 6,374 entries belonging to the verbal paradigm in the VTCSP into 2,899 families.

Singleton entries	Primary Entries (which are the target of an <u>Under</u> element link)	Secondary entries (containing an <u>Under</u> element linking to another entry)		Total
1,478 items	1,421 items	3,475 items		6,374
2,899 heads of families (all having the V <u>CATEGORY</u> label)		with V <u>CATEGORY</u> label	without V <u>CATEGORY</u> label	items
		356 items	3,119 items, not having V among their <u>CATEGORY</u> labels	
3,255 items having the V <u>CATEGORY</u> label				



**Figure 17:** Entries 213\_L\_k to 213\_L\_m in the 1679 VTCSP.

Because a satisfactory treatment of the group of 554 causatives in the VTCSP would require a full-length article, I shall simply enumerate, in the rest of this section, a few basic facts, which have to do with the morphology of causatives and with the strategies which we can observe, inside the Portuguese glosses which explain the meaning of those entries. From the morphological point of view, it can be said that all the elements tagged by me as CAUSAT inside a CATEGORY element, end either in *-vikkīratu* or in *-pikkīratu*. The group

<sup>60</sup> Inside my XML database, the entry elements for a causative verb receive two CATEGORY element tags, one which labels them as V, and a second one which labels them as CAUSATIVE. There is also present in the entry a dedicated element, distinct from the Under element, linking them to the entry for the corresponding simple verb, by means of its coordinate. As an example, the entry element having coordinate 213\_L\_k (*pārpikkīratu* ‘Mandar ver’), contains a tag which is a pointer towards entry 213\_R\_a (*pārkkīratu* ‘Olhar’).

is already identified, morphologically and semantically, inside HH's Arte and we can conveniently refer to those forms as being the subtense  $\gamma$ -G (see **Table 1**), following the conventions introduced in Hein & Rajam (2013: 99). As far as the meaning is concerned, I shall simply reproduce, as a preliminary explanation, a small extract from from the two paragraphs devoted to the topic in HH's arte, where they take up the second half of folio 54v and the whole of folio 55r.

- (8a) Cheiyren quer dizer: fazer [sic], cheiyuiren: faço fazer; ariquiren: espamco, ariquiren: faço espamquar. (Vermeer 1982: 47)
- (8b) *Cheyren* means 'I do', *cheyvuiren* 'I cause to do', *ariquiren* 'I beat', *ariquiren* 'I cause to beat'. (Hein & Rajam 2013: 99)

When however we compare what looks like a simple and straightforward explanation, occupying two paragraphs in a grammar, with the more complex linguistic reality captured inside a set of 554 entries in the 1679 VTCSP, we find that several strategies have been used by the Portuguese translators for approximating the relationship between the base form and its causative, although there is very clearly one which is used in a majority of cases. More precisely, on the total of 554 pairs of VTCSP entries consisting of a base verb and its associated causative.

- in 494 cases, the causative is translated by a periphrastic form containing *fazer* combined with a Portuguese infinitive which has been used as one of the possible translations for the base verb. For instance 486\_R\_q (*cirippikkiratu*) is translated by *fazer rir* ['to cause to laugh'] whereas 486\_R\_s (*cirikkiratu*) is rendered by *rir* ['to laugh']. This corresponds to the strategy described in (8a–b) for *ariquiren* 'es-pamco' and *ariquiren* 'faço espamquar', although a more faithful transliteration of the corresponding entries in the VTCSP is *atippikkiratu* (36\_R\_l 'fazer espancar' ['to cause to beat']) and *atikkiratu* (36\_R\_p 'espancar, [...]') ['to beat']
- in 22 cases, we have a pair where the causative is translated by a periphrastic Portuguese form containing *fazer* while the base verb is a reflexive/pronominal form. For example, *kūnivikkiratu* (321\_L\_g) is explained by 'Fazer corcouar' ['to cause to bend'], whereas *kūnukiratu* (321\_L\_j) is explained by 'Corcouarse' ['to bend'].
- in 8 cases, instead of having a periphrastic form with *fazer*, we have a periphrastic form with *mandar* 'to order, to tell'. This is for instance the case with the example 213\_L\_k, alias *pārpikkiratu* 'Mandar ver' ['to tell to see'] already provided in **Figure 17**.
- in a few cases, unrelated verbs are used in the Portuguese translations for the causative and for the base verb. This is the case for instance for *paṭippikkiratu* (204\_R\_c, 'enfinar' ['to teach']) and *paṭikkiratu* (204\_R\_i, 'aprender, estudar' ['to learn, to study']). The translator may have thought that it would be too clumsy to use a periphrase in the translation.
- occasionally, additional grammatical technical terms (Active, Neutro, ...) are seen in some entries.

However, as a last observation before going to the next section, I shall remark that, if we leave the 17<sup>th</sup> century and move to an examination of the Madras Tamil Lexicon (MTL), we seem to find in it very few entries for causatives. A rapid search seems to locate only 19 causatives having a stem ending in *-pi* (such as *karpittal* 'to teach')<sup>61</sup> and 24 causatives

<sup>61</sup> *karpittal* 'to teach' is the causative of *karral* 'to study'.

having a stem ending in *-vi* (such as *terivittal* ‘to cause to know’).<sup>62</sup> Given the fact that the VTCSP has 16,215 entries whereas the MTL has more than 100,000 entries, the densities of causatives are respectively 3.4% and 0.043%. Before moving to the next section, I shall make two additional remarks in connection with this observation:

- the first remark is that we should probably not continue to count the causatives in the count of 2,899 verb candidates, which was given in **Table 4**. If we subtract from that total the 554 causatives, reclassified as another type of secondary entries,<sup>63</sup> we would now have a reduced total of 2,345 potential heads of families, whereas the total count of secondary entries would now increase to 4,029.
- the second remark is that it is possible that the high visibility of causative forms in the VTCSP is due to the fact that part of the corpus analysed by Portuguese missionaries was heavily sanskritized prose. Such a hypothesis can only be verified by a careful examination of early prose. I shall simply remark that HH declares, at the end of the second paragraph devoted to causatives *E vsaõ muito deste modo de cõpoer os que bẽ sabẽ falar* (Vermeer, 1982: 47).<sup>64</sup> Such a remark may mean that the missionaries were aware of the fact that causative forms were not used very spontaneously in ordinary language. This is a topic for further investigation.

We shall now move to a brief examination of another feature, which is probably more central in a typological characterization of the signature features of Tamil (and of other related languages).

## 10. How the VTCSP dealt with paired verbs

In this last section before the conclusion, I shall examine very briefly the elements belonging to two twin subsets of the set of primary V-tagged entries in the VTCSP, referred to hereafter as AV (i.e. set of ‘affective verbs’) and as EV (i.e. set of ‘effective verbs’). These sets both contain 167 entries and a one-to-one *natural* correspondence exists between them. When using the word *natural*, I am hereby postulating<sup>65</sup> that the correspondence is intuitively recognized by native speakers of Tamil but that it took time for the Portuguese-speaking missionaries to master it. In the modern period, the elements of these two sets have been described as ‘paired verbs’, each pair containing an affective verb and an effective verb. This topic has been studied for the first time in depth in Paramasivam (1979) and is also the topic of study chosen by Cristina Muru, in this issue of the JPL (See Muru 2021). In this section however, I shall limit myself to giving a brief overview of the visibility of those 167 pairs in the VTCSP, and shall also briefly give an account of HH’s strategy for coming to terms with their existence.

One of the difficulties in the initial evaluation of the situation is that all the pairs do not follow the same morphological pattern. In the largest subgroup which is observable inside the set of 167 pairs, both verbs in the pair have the same stem but they differ by the ending, whereas in other subgroups, the two verbs in the pair have different stems, which combine with the same endings.<sup>66</sup>

<sup>62</sup> *Terivittal* ‘to explain ...’ is the causative of *terital* ‘to know’.

<sup>63</sup> Those 554 causative entries would then be added to the 356 entries already contained in one of the cells of Table 4. Most of those 356 entries are what is called ‘Vector Constructions’ by many descriptors of Indian languages.

<sup>64</sup> Hein & Rajam (2013: 100) translate: ‘Those who know how to speak well use this form frequently’.

<sup>65</sup> Technically, in order to make visible the linking which I postulate to exist, I introduce inside the entry for an affective verb, a special tag which gives the coordinate of its effective counterpart in the pair, and symmetrically, inside the entry for an effective verb, I introduce another type of tag which gives the coordinate of its affective counterpart.

<sup>66</sup> Difference both in stem and ending also exists, but less frequently.

- the first situation is illustrated by the elements of the pair consisting of *kurai-kkiratu* [315\_R\_o] and *kurai-kiratu* [315\_R\_j], for which the Portuguese translations given in the VTCSP are ‘diminuir’ (effective verb) and ‘diminuirse’ (affective verb)
- the second situation is illustrated by the elements of the pair consisting of *kalakk-iratu* [255\_L\_g] and *kalañk-iratu* [254\_R\_n], translated respectively as ‘mifturar’ (effective verb) and ‘mifturarse’ (affective verb)

When HH first tried to describe the phenomenon in his *Arte*, in a section called ‘Dos verbos passiuos’ (Vermeer 1982: 107), he started his presentation by the following declaration:

- (9a) Para o que se saibe que nesta limgoa nõ ay passiuua que propiamente se possa dizer passiuua, mas ai modo de falar ã alguãs das conjugações que he quasi como passiuo. (Vermeer 1982: 107)
- (9b) Understand that in this language there is no passive properly so called. But in some of the conjugations there is a way of speaking that is almost like the passive. (Hein & Rajam 2013: 211)

I shall not however elaborate much further on the topic because this same issue of the JPL already contains an article by C. Muru dedicated to paired verbs. I shall simply observe that the current conclusion of several centuries of observations, as far as paired verbs are concerned, is that, unlike what is done in the treatment of causatives, it is currently considered preferable to treat the two elements of a pair as two distinct lexemes, rather than to consider the affective verb as a kind of passive form with respect to its corresponding paired effective verb. Finally, from a statistical point of view given the fact that my temporary conclusion (at the end of section 9) is that the number of (potential) heads of (verbal) families does not exceed 2,345, and given the fact that we have 167 pairs, associating one effective verb and one affective verb, we can conclude that the proportion of paired verbs recognized by the VTCSP is at least 14%. This figure is certainly lower than the estimation found in Paramasivam (1979: 5, fn.8). However, since the figure of 2,345 is an upper bound, which will decrease as the classification of entries is refined, the real percentage of paired verbs is certainly higher. As a final hypothesis, we can also consider the possibility that *paired verbs are under-represented* in the VTCSP because of the factor which I have already evoked at the end of section 9, when trying to explain why causatives are over-represented in the VTCSP, namely the fact that it is partly based on the analysis of heavily Sanskritized prose.<sup>67</sup> Additional research is necessary.

## 11. Conclusion & future challenges: from Figure 5 to Figure 18

I believe I have now reached the limit of what can be exposed on such a complex topic in a single article. In this concluding session, I would like to try to sketch the future challenges, which are implicit in the shuttling between several points of view which has

<sup>67</sup> See the remark contained in the VTCSP Preface, at the end of the section which starts by the title ‘Ao Leitor Pio e Zeloso’:

*Ponho porẽ muitas palauras puramente Grandonicas, affy pera quem ler os liuros do P. Roberto, aonde estaõ muitas, como por q̃ fã ordinarias no ordinario modo de fallar dos Bramanes, cujo fallar hẽ mais fubido, aquem os Tamuis, q̃ fe pressõ, de doutos, & querem fallar graues, & felecto, querem imitar.*

(Thani Nayagam 1966)

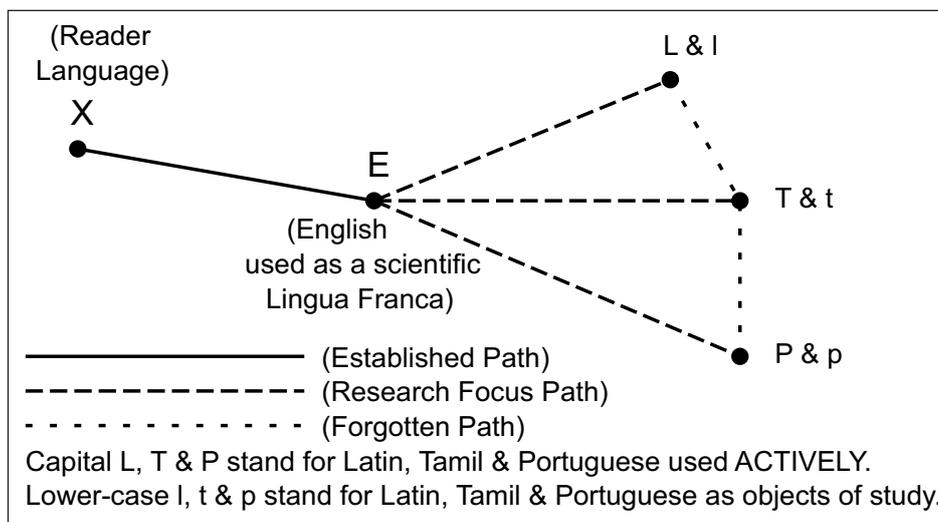
[‘But I include many Sanskritic words, both for the sake of those who read the books of Fr. Robert (Nobili) in which they are numerous, and also because they commonly occur in the ordinary conversation of the Brahmins, whose language is more elevated, and whom Tamilians, who consider themselves learned and wish to speak seriously and with care, try to imitate’

(Thani Nayagam, 1966: 13, Translation by E.C. Knowlton Jr & X.S. Thani Nayagam)].

taken up a great part of the substance of this article in which I am trying, using English as a scientific *Lingua Franca*, to explain to readers, who could have any language X as a mother tongue, why it is *relevant* to examine the content of several texts —and especially one, the VTCSP, in which I have invested a lot of efforts— which were created at a time when some human beings, who were using Portuguese as a *Lingua Franca*, were trying to learn, to teach and to use another language. This represented a huge effort, which should not be underestimated because, even after 200 years of successive attempts, they were producing books in which native Tamil teachers —being the exponents of the B-variety and the admirers of the C-variety of Tamil— would have found many faults, as should be clear from what is illustrated inside section 8, and notably by the group (7a)–(7c).

When using the word ‘relevant’ in the previous paragraph, I was having in mind several types of possible readers of this article, who are indirectly evoked by the elements of the small graph (or network) which is sketched inside **Figure 18**, above. I see primarily three possible main scenarios for the readers of this article, with an additional fourth one, depending on the language X which is a primary language for them, in addition to their compulsory command of E:

- If X refers to a language actively used by the reader of this article, and possibly a mother-tongue ...
- Scenario 1 (X = T). For those readers, who are already in control of the B-variety of Tamil and of at least one (A-variety) dialect of Tamil, the primary challenge consists in discovering other (A-variety) dialects, while learning enough archaic Portuguese for that purpose. Another possible challenge would be to make 17<sup>th</sup> century Portuguese into the primary target language.
- Scenario 2 (X = P). For those readers, who are already in command of Portuguese, the primary target will be Tamil and the main difficulty for them will be to understand what exactly the  Tamil triglossia is. They can also decide to invert the object of study and verify whether for instance the implicit comparison between the structures of Tamil and of Portuguese which is partly visible in many of the Portuguese translations given for Tamil examples, can decenter their perspective on their own language. They might also try to go into a deeper exploration of the phonetic side.



**Figure 18:** Knowledge is always mediated by at least one language, or by more ... [There is no language-independent knowledge about languages] (A Battle Plan for future explorations).

- Scenario 3 (X is neither T nor P). For those readers, everything is difficult, because there are two languages to study at the same time, namely T and P, and the object of study is a forgotten path between those two languages, as shown on **Figure 18**. However, I believe that even for these readers (who are like me), although the challenge is, in the absolute, greater than what it is for those belonging to the first two groups, the task is as beautiful.
- Scenario 3b (X = L). Finally, inside the group described in Scenario 3, there is a subgroup of researchers who are fluent users of Latin. Those users also have a unique perspective, as can be seen on **Figure 18**, but I shall not elaborate further.

I must of course not choose between all those potential readers and must now return to the task of preparing several entry points (and human interfaces) for the future consultation of the forthcoming database —*i.e.* the electronic edition of VTCSP— keeping in mind the various possible users which it may have, and remembering that, if *one has only one chance of making a good first impression*, the clarity of the interface navigation can hopefully go a long way in this new age of digital humanities, where content accessibility may have become the main challenge for the transmission of knowledge, if the impressive efforts made by AP's colleagues in 1679 are to be retrieved from the quasi-oblivion where they are now.

### Competing Interests

The author has no competing interests to declare.

### References

- Auroux, S.** (1994). *La révolution technologique de la grammatisation*. [The Technological Revolution of Grammatization]. Mardaga: Liège.
- Aussant, É., & Chevillard, J.-L.** (2020). Foreword. In É. Aussant & J.-L. Chevillard (Eds.), *Extended Grammars, Beiträge zur Geschichte der Sprachwissenschaft. Themenheft, 30(1)*, 3–8. Münster: Nodus Publikationen.
- Arden, A. H.** (1976). *A Progressive Grammar of the Tamil Language*. Revised by A. C. Clayton. Chennai: The Christian Literature Society.
- Beschi, C. J.** (1738 [ms. 1728]). *Grammatica Latino-Tamulica, ubi de Vulgari Tamulicæ Linguae Idiomate கொடுந்தமிழ் [KOṬUNTAMIL] dicto* [See translated title in following entries dated 1831 and 1848]. Tranquebar: Typis Miſſionis Danicæ [Danish mission press].
- Beschi, C. J.** (1831 [1806]). *A Grammar of the Common Dialect of the Tamulian Language, Called Koṭuntamiḷ*. [translated by Christopher Henry Horst]. Madras: Vepery Mission Press.
- 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 Mahon, George Wiliam*. Madras, Vepery: Christian Knowledge Society's Press.
- Chevillard, J.-L.** (1996). *Le commentaire de Cēṇāvaraiyar sur le Collatikāram du Tolkāppiyam. Sur la métalangue grammaticale des maîtres commentateurs tamouls médiévaux. Volume I*. [Cēṇāvaraiyar's commentary on the Collatikāram of the Tolkāppiyam. On the grammatical metalanguage of medieval scholarly commentators]. Publications du département d'Indologie — 84.1. Pondicherry: IFP/EFEO.
- Chevillard, J.-L.** (2009). Elementary Grammatical Devices in Classical Tamil Grammar. In P. Hummel (Ed.), *Epilanguages, Beyond Idioms and Languages* (pp. 203–216). Paris: Philologicum.

- Chevillard, J.-L.** (2015). The challenge of bi-directional translation as experienced by the first European missionary grammarians and lexicographers of Tamil. In É. Aussant (Ed.), *La Traduction dans l'Histoire des Idées Linguistiques* (pp. 107–126). Paris: Librairie Orientaliste Paul Geuthner.
- Chevillard, J.-L.** (2017). How Tamil was described once again: Towards an XML-encoding of the Grammatici Tamulici. *Histoire Epistémologie Langage*, 39(2), 103–127. Paris: SHESL/EDP Sciences. DOI: <https://doi.org/10.1051/hel/2017390206>
- Chevillard, J.-L.** (2018). The Tamil grammatical tradition: A long commute between theory and practice. *Indian Linguistics. Journal of the Linguistics Society of India*, 79(1–2), 25–38.
- Dhamotharan, A.** (1978). *Tamil Dictionaries. A Bibliography*. Beiträge Zur Südasien-Forschung Südasien-Institut Universität Heidelberg Bd. 50. Wiesbaden: Franz Steiner Verlag.
- Dominic Raj, M.** (2019). *Thembavani. A Garland of Unfading Honey-Sweet Verses*. Translation into English Verse with Sandhi-separate Tamil Text. India: Independently Published. ISBN: 978-1-073-06153-2.
- Ebeling, S., & Trento, M.** (2018). From Jesuit Missionary to Tamil Pulavar: Costanzo Gioseffo Beschi (1680–1747), the ‘Great Heroic Sage’. In T. Leucci & M. Fourcade (Eds.), *L’Inde et l’Italie. Rencontres intellectuelles, politiques et artistiques* (pp. 53–90) (Coll. Puruṣārtha 35). Paris: Éditions de l’École des hautes études en sciences sociales. DOI: <https://doi.org/10.4000/books.editionsehess.23276>
- Hein, J., & Rajam, V. S.** (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.
- Henrique, H.** (1967 [1586]). *Aṅṭirīkku aṭikaḷār iyarriya Flos Sanctorum enra aṭiyār varalāru*. [The *Flos Sanctorum* composed by the Venerable Henriques, also called History of the Saints]. Edited by Ca. Irācamānikkam. Tūttukkuṭi, India: Tamiḷ Ilakkiyak Kalakam.
- James, G.** (2000). *Colporuḷ, A History of Tamil Dictionaries*. Chennai: Cre-A.
- James, G.** (2007). The manuscript and printed versions of Antão de Proença’s Vocabulário tamulico (1679). no:kku, 1. (article accessed on the Academia page of the author).
- James, G.** (2009). Aspects of the structure of entries in the earliest missionary dictionary of Tamil. In O. Zwartjes, R. Arzápalo & T. Smith-Stark (Eds.), *Missionary Linguistics IV/Lingüística misionera IV* (pp. 273–301). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/sihols.114.15jam>
- Jeyaraj, D.** (2010). *Tamil Language for Europeans: Ziegenbalg’s Grammatica Damulica (1716)*. Translated from Latin and Tamil. [Annotated and commented by Daniel Jeyaraj]. Wiesbaden: Harrassowitz Verlag.
- Lehmann, A.** (1955). *Es begann in Tranquebar. Die Geschichte der ersten evangelischen Kirche in Indien*. [It started in Tranquebar. The History of the First Evangelical Church in India]. Berlin: Evangelische Verlagsanstalt.
- Lehmann, T.** (1989). *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture, Publication No. 1.
- Lindholm, J. M.** (1972). Cleft Sentences in Tamil and Malayalam. In V. I. Subramoniam & E. Valentine (Eds.), *Proceedings of the First All India Conference of Dravidian Linguists* (pp. 297–306). Trivandrum: University of Kerala, Kariavattam. Dravidian Linguistics Association of India.
- Mahon, G.** (1848). See Beschi[1848].
- Muru, C.** (2020). Grammaire Latine Étendue. Two Portuguese missionary Tamil Arte (17th cent.). In É. Aussant & J.-L. Chevillard (Eds.), *Extended Grammars*. Topical

- issue of *Beiträge zur Geschichte der Sprachwissenschaft*, 30(1), 59–73. Münster: Nodus Publikationen.
- Muru, C.** (forthc.). Verb conjugation and paired verbs in Tamil: How missionaries understood and classified them. To appear in *Journal of Portuguese Linguistics*.
- MTL: Madras Tamil Lexicon.** (1982 [1924–1939]). *Tamil Lexicon*. Published under the authority of the University of Madras, 6 volumes and 1 supplement. Madras: University of Madras.
- Paramasivam, K.** (1979). *Effectivity and Causativity in Tamil*. Trivandrum: Dravidian Linguistics Association.
- Pellin, T.** (2019). Introduction – Aspects of the grammatization of the Chinese language. In *Histoire Epistémologie Langage*, 41(1), 7–14. Paris: SHESL/EDP Sciences. DOI: <https://doi.org/10.1051/hel/2019010>
- Proença, A.** (de). (1679). [See Thani Nayagam(1966)]. [The original is now available online on the Vatican Library website at [https://digi.vatlib.it/view/MSS\\_Borg.ind.12](https://digi.vatlib.it/view/MSS_Borg.ind.12)]
- Schiffman, H.** (1979). *A Grammar of Spoken Tamil*. C.L.S. Books. Chennai: The Christian Literature Society.
- Steever, S. B.** (2005). *The Tamil Auxiliary Verb System*. Routledge Asian Linguistics Series. London and New York: Routledge.
- Sweetman, W., & Ilakkuvan, R.** (2012). *Bibliotheca Malabarica. Bartholomäus Ziegenbalg's Tamil Library. An Annotated Edition and Translation*. Collection Indologie – 119. Pondicherry: IFP/EFEO.
- Thani Nayagam, X. S.** (1966). *Antaõ de Proença's Tamil-Portuguese Dictionary A.D. 1679*, [University of Malaya, Kuala Lumpur]. Leiden, Netherlands: E. J. Brill.
- Vermeer, H. J.** (1982). *The first European Tamil Grammar. A Critical edition*. [English version by Angelika Morath]. Heidelberg: Julius Groos Verlag.
- Vinson, J.** (1910). Le Premier Vocabulaire Tamoul Imprimé. [The first printed Tamil vocabulary]. *Revue de Linguistique et de Philologie Comparée*, 74–82. Tome Quarante-troisième. France: Chalon-Sur-Saone.
- Walther, C. T.** (1739). *Observationes Grammaticae, quibus Linguae Tamulicae idioma Vulgare*, [Grammatical observations pertaining to the ordinary variety of Tamil]. Tranquebar: Typis Missionis Regiæ [Royal mission press].
- Wilden, E.** (2018). *A Grammar of Old Tamil for Students*. Collection Indologie n°137. NETamil Series n°3. Pondicherry: EFEO/IFP.
- Ziegenbalg, B.** (1716). *Grammatical Damulica*. [Tamil grammar]. Halae Saxonum, Halle: Orphanotrophei [Orphanage].
- Zwartjes, O.** (2011). *Portuguese Missionary Grammars in Asia, Africa and Brazil, 1550–1800* [Studies in the History of the Language Sciences 117]. Amsterdam/Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/sihols.117>

**How to cite this article:** Chevillard, J.-L. (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: 9, pp. 1–34.  
DOI: <https://doi.org/10.5334/jpl.269>

**Submitted:** 21 February 2021    **Accepted:** 24 May 2021    **Published:** 01 July 2021

**Copyright:** © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Journal of Portuguese Linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 