

Tamil Literature from Sangam to Modern Period: What does a search of the Tamil Electronic data reveal?



Prof. Dr. Vasu Renganathan

University of Pennsylvania, USA

Abstract

A systematic study of the Tamil language from Sangam to Modern period from a historical perspective may reveal that there does exist a continuum of changes that occurred from one stage to another in Tamil language. Without such a study, any synchronic description of Tamil would only reflect its complexity in an overwhelming way. In other words, The Tamil language, the way it is now with a museum of complex forms, expressions and grammatical constructions, both in written and spoken variety, demonstrates a vast number of linguistic characteristics at phonological, morphological and syntactic levels, that require a comprehensive diachronic study to fully understand them in a coherence way. In this respect an extensive electronic database of Tamil texts from all of the stages along with a powerful query tool to search texts from various dimensions is indispensable. This paper is an attempt to illustrate how such an electronic database for Tamil (<http://sangam.tamilnlp.com>) can be used extensively to study some of the morphological and

syntactic behaviors of Tamil from a historical point of view.¹

Introduction

Upon exploring the Tamil electronic database consisting of a variety of data ranging from the Sangam to Modern Tamil, especially by employing the principles of historical linguistics, one may immediately be able to notice that the changes that underwent throughout the history of Tamil language exhibit a systematic, regular and what one may attribute as a set of colorful changes in it. Phonological, morphological and syntactic changes that took place to this language one after another in a sequential manner contributed to the dearth of complexity as we see now as modern Tamil (both spoken and written) – a language that many have attempted to study it using many grammars and dictionaries in many different points of views!

What may one illustrate it in a minuscule is that when words or combination of words and suffixes undergo all possible phonological rules on them, either successively at one point

An earlier version of this paper was read in the tenth International Tamil internet conference of INFITT held in Philadelphia and subsequently published in the proceedings: <http://www.uttamam.org/papers/tic2011.pdf>.

of time or periodically at different stages, what results is a set of the most complex forms that can be understood in terms of many dichotomies such as social versus regional dialect; spoken versus written variety; high versus low register; casual versus platform speech and so forth. Thus, attempting to learn this language that contains such a complex set of shades of variations does indeed pose a greater level of difficulty than normal for any second language learner. Not only does it become a big challenge to any second language learner in having to comprehend and use these multiple facets of this language, but it also becomes an immense task for an instructor/evaluator as to how one can judge the competency of a learner who attempts to master it!

Thus, by not familiarizing oneself with the myriad of complexities within the Tamil language, either from a historical or purely from a synchronic point of view, one may tend to attribute each of these varieties as belonging to a separate language; and subsequently consider the variations therein as haphazard and random. Upon studying the Tamil language variations from a historical point of view, one may easily note that such variations are vibrant and quite regular, and notably they conform to a logical sequence of changes. In this respect, no form of Tamil, either it is spoken or written, is neither random nor spontaneous in nature. Not to mention the fact that any study of diaspora Tamil of any region, without such a systematic account from a historical point of view, would only result to provide an unscientific description of the language of the respective region.

Language change occurs as a result of both internal as well as external causes. Internal causes are a) application of more than one phonological rule on agglutinative words; b) undergoing many naturally occurring linguistic processes such as, grammaticalization, reanalysis, metaphorization etc., in the language (See Renganathan 2010). The external causes,

on the other hand, can be attributed to such factors like 'foreign language contact', 'bilingualism', 'language dominance', and so on to name a few. Not to mention the fact that over the period of a long history, Tamil language did undergo many changes both due to internal as well as external causes. Prakrit, Sanskrit, Persian, Portuguese, and more recently the English language contributed enormously to the development/distortion of Tamil language in a number of different ways.

Interestingly, many Tamil language movements, both conscious as well as unconscious, such as 'language purism', 'official language planning', 'language standardization', 'Tamilization', 'coining new vocabularies' and so on contributed to the retention of many of these variations within it without having to undergo any extinction in any subtlest manner possible. Many of the so called indigenous and historically relevant Tamil words and morphological and syntactic forms - although not all of them - from the Sangam era are still extant in modern Tamil in one way or another: in one dialect or another, in one speech form or another, or in one register or another.

This particular behavior of the Tamil language poses as a big threat not only for its continued consideration as an individual language, but also for its continued use of indigenous and historically significant forms under various labels as 'pure Tamil', 'Sangam Tamil', 'Chastised Tamil' and so on. Ironically, the major threat comes mostly from the judgments of second language learners for whom these historically relevant changes and existence of complex variations pose as a major hurdle in learning the language in a casual manner.

Delving into complexity – A case in point is the use of the verb en (என்) - 'to say':

Almost all of the grammatical categories in Tamil have a systematic history behind them, and accounting all of them may require enormous amount of time and

energy. An attempt is made in this section to trace the various use of the Tamil quotative marker *என்று* - *enru* - 'that' and its historical development, especially by making use of the electronic data extensively. Use of the verb *என்* - *en* 'to say' can be taken as one of the instances for the contribution of complex forms in Tamil. This verb has undergone a wide range of alterations throughout the history of the Tamil language, but yet, it is still in use in the modern language the way it was during the Sangam period - perhaps with a greater number of characteristics which were not prevalent at its earlier stages. Unlike any other verb, this verb exhibits many structural gaps in modern written variety, but, significantly, not in spoken Tamil.

Learning to master all of the uses of this verb, especially in spoken Tamil, is definitely one of the major challenges to any second language learner for the main reason that it not only underwent the process of grammaticalization, but also shows an agglutinative structure that is very difficult to comprehend and use by any non-native speaker of the language. This verb was used both as a regular lexical form as well as a grammatical form representing the 'complementizer' in the Tamil language. Both the forms of *என்று* - *enru* - 'that' and *என்பது/என்றல்/எனல்* - *enpatu/enral/enal* - 'the fact that' had their equivalents both in old and modern Tamil.

A search of the database using a number of combinations, including *என்*, *என்று*, *என்றிட*, *என்றுகொள்* and so on would reveal that besides the many of the finite forms of this verb, what underwent a significant change at a later period are the forms of negative adverbial (*என்னாது* - *ennātu* - 'without telling') and nominal derivative (*என்னாமை* - *ennāmai* - 'not saying'), which do not show any equivalent in modern literary Tamil.

*அரும் படர் அவல நோய் ஆற்றுவள்
என்னாது* (Kali. 28)

arum paṭar avala nōy āṟṟuvaḷ ennātu

'Without revealing the fact that she would experience the contagious love disease...'

அரிய ஆகும் என்னாமை (Aham. 191)

ariya ākum ennāmai ..

'Not saying that s.t. would be intricate to accomplish...'

Notably, the Modern Tamil equivalents of the suffixes - *ātu* and - *āmai* such as - *āmal* (eg. *collāmal* 'without saying' **ennāmal*) and -*ātatu* (*collātatu* 'that which was not said' **ennātatu*) respectively tend to occur with the verb *en* only in spoken Tamil but not in the corresponding literary variety. What turns out to be the crux of the issue here is the obscure nature of the spoken Tamil equivalents of the verb *en* 'say' in present, past and future forms, which normally occur as a single or clustered consonant: *ந்-்*' (*பாக்றேந்றேன்* - '*pākṛēṅṛēn* - 'I say that I see', *சொல்றாந்றான்* - '*colṛāṅṛān* - 'he says that he tells'); *ண்ண்-்*' (*பாக்றேண்ணேன்* - '*pākṛēṅṅēn* - 'I said that I see'); and *ம்ப்-*' (*கொடுப்பேம்பேன்* - '*koṭuppēmpēn* - 'I will say that I would give') respectively (Cf. Search: *ngr*, *என்கிற*). The obscure form of this suffix, its complex clause construction in an agglutinative form, along with the non-existence of some of the conjugations of this verb in written Tamil contribute enormously to the complexity of spoken Tamil.

- *வருவாங்காமெ* - *varu-vān-ṅ-āme* - 'without saying that he would arrive..'
- *வரமாட்டேங்காமெ* *vara-māṭṭ-ēn-ṅ-āmæ* 'without saying that I won't come...'
- *ஆகுங்காதது* - *āku(m)-ṅ-āta-tu* - 'saying that s.t. wouldn't happen'
- *கேப்பாங்காதது* - *kē(!)-pp-āṅ-āt-atu* - 'saying that he wouldn't ask'

Notably, these, supposedly, commonly occurring forms in spoken Tamil do not have any parallel in written Tamil, as a result it generates a structural gap in the corresponding written variety of Tamil.

ennæ māṭu kiṭu-ṇṇu-kiṭṭiru-kk-æ?

‘Why do you keep calling me a water buffalo?’

நீயே எடுத்துக்குவேண்ணுக்கோ!

nīyē eṭu-tt-u-kku-v-ēṇ-ṇu-kkō

‘Proclaim that you would take everything for yourself’

What do these exceptional forms imply is that ‘spoken Tamil’ and ‘written Tamil’ seem to have followed two different historical paths from Sangam to modern Tamil and in this respect the spoken Tamil seems to show a richer structure than the written Tamil, especially in terms of retaining more number of archaic forms than the corresponding written version. This is in opposition to those instances of modern Tamil where new structures evolved and no traces of which can be found either in Sangam or in medieval Tamil.

An example may be the case of experience subject construction, which is new to modern Tamil, but not in Sangam Tamil, as in yāṇ viyartṭaṇaṇ ‘I was sweat’ as opposed to eṇakku viyarkkiṛatu (cf. Murugaiyan 2004). Yet another feature from a historical point of view is loss of medieval and Sangam forms which do not have any trace in modern Tamil. A case in point is the use of imperative suffixes –min (kēlmin - ‘listen’) and –anmin (kūṛanmin - ‘do not utter’) etc., which do not have any occurrence in any identical forms in modern Tamil (cf. Renganathan 2010). Identifying the point of time in which these changes occurred is an endeavor that requires analyses of text of different genre in a thorough manner.

Yet another advantage of studying word forms that underwent many changes historically using electronic data is that it is possible for one to trace the trajectories of the cause of certain changes over the period of time in a systematic manner. One of such phenomena is authors’ handling of a

particular style causing the development of new categories. One of them that may be cited here is the formation of the modern Tamil modal auxiliary lām. It may be stated that various use of the combination of the infinitive suffix –al with the neuter future form of verb āku ‘become’ in ancient Tamil later caused the formation of lām. An extensive search using the keys such as லாகுமே, லாமே, லும் ஆமே, லும் ஆகும் etc., one may notice that the modal auxiliary lām came into existence in modern Tamil by the linguistic process of reanalysis due to various use of this structures by poet saints.

Consider for example the expression kēṭu uṇṭu eṇṛal tuṇintu colal ākum – Manimekalai and its modern Tamil equivalent kēṭu irukkum enkiṛ-atu tuṇintu collalām (Modern Tamil) ‘One may say for sure that there would be a devastation’, where the syntactic construction colal ākum ‘saying is possible’ is found to be occurring with many different combinations synonymously, as in colal ām - after phonological reduction of ākum to ām; colla lām with a reanalysis of verb forms and so on (see Renganathan 2010: pp. 171-73 for a detailed study of this change). By toggling between the selections of the bhakti, Sangam and modern literatures using the above search keys, one can notice the various use of this combinations more in bhakti texts than in Sangam texts.

Search techniques and need for a tagged Corpora:

Perhaps an advance search technique with many combinatory possibilities is needed to successfully derive all of the intended and unattested forms from all of the genres of Tamil language. Ideally, one may want to search text in many complex ways, like ‘words that end in particular suffix (-viṭu; kiṛ; koḷ etc.)’, sentences with a particular combination of words (dative subject and psychological verbs; subject with the suffix āl and modal verbs like -oṇṇ, -iyal etc.) and so on. Even though such sophisticated search

possibilities is yet to be made available for Tamil using any conceivable tagged corpus as discussed in detail in Renganathan(2001), Baskaran et al (2008) etc., with the current database, however, storing text in Unicode does offer some work around. For example, if one intends to retrieve all of the word forms with the suffix -iṭu, āl, -ukku and so on, one can use the Unicode glyphs of the initial vowels, as in ிடு, ால், ுக்கு respectively to accomplish this task.

This method can be considered as a substitute for any equivalent method of information retrieval using tagged corpus, which would normally contain all of the affixes parsed and stored separately in a more systematic manner. Absence of any such tagged corpus and an intelligent parser for all of the genres of Tamil texts from Sangam to Modern Tamil, one requires to use this kind of alternative search methods to accomplish the task. Among many others, the other significant advantages of using electronic data may be making dialect geographies from a historical point of view, attempting to find the chronology of authors and texts and so on.

References:

1. Baskaran Sankaran, Kalika Bali, Monojit Choudhury, Tanmoy Bhattacharya, Pushpak Bhattacharyya, Girish Nath Jha, S. Rajendran, K. Saravanan, L. Sobha, and K. V. Subbarao, A Common Parts-of-Speech Tagset Framework for Indian Languages, Proceedings of LREC 2008.
2. Murugaiyan, A (2004) Note sur les prédications expérientielles en tamoul classique, Bulletin de la Société de Linguistique de Paris, 99, p. 363-382.
3. Renganathan, Vasu. (2010). The Language of Tirumūlar's Tirumantiram, A Medieval Saiva Tamil Religious Text. Unpublished Ph.D. thesis, University of Pennsylvania, Philadelphia.

4. Renganathan, Vasu. (2001). Development of Morphological Tagger for Tamil, INFITT Conference, Kuala Lumpur.

About the Author:

Dr. Vasu Renganathan has been teaching Tamil language and literature at the University of Pennsylvania, USA since 1996. He earned his Masters from the University of Washington, Seattle and Ph.D. in medieval Tamil language and literature from the University of Pennsylvania. He also has another Ph.D. in linguistics from the Annamalai University, India.

Dr. Renganathan began his teaching career in 1989 and previously taught at the University of Michigan, Ann Arbor and the University of Washington, Seattle. In addition to teaching, he has also worked on various curriculum and instructional material development projects. Vasu also served as the Chairman of the International Forum For Informational Technology In Tamil (INFITT) and currently serving as one of the Executive Committee members of the South Asian Language Teachers Association (SALTA). Vasu is also in the International Advisory Board member for the Electronic Journal of Foreign Language Teaching, (e-FLT), Singapore.

His research interests include: History of Tamil language and literature, History of the Tamil Religious developments from medieval period, Language of Sangam Tamil literature and Tamil inscriptions, Teaching and learning of Tamil language and Impact of Technology in the pedagogical methods. Some of his Publications include: Tamil Language in Context: A comprehensive approach to learning Tamil and An English Dictionary of Tamil Verb (Associate Editor).

