

CORPUS ANALYSIS ON CONFLICTS

Ganesan Ambedkar

Dravidian University
ganesanambekar@gmail.com

Abstract

Modern Tamil language and society have been facing heavy pressure, predominantly- falling under the categories of visible-frequency-pressure - from the dominance of English language. Effects of visible-frequency - pressure-dominance on Tamil language may be found at varying degrees at different linguistic levels. The aim is to study outcomes of resultative sides of conflicts in Tamil. It is an observatory paper.

1. The Aim and organization

After the introduction of English, as a medium of instruction (for the details, see Annamalai, 2001 and 1999), more Tamil youngsters, of Tamil Nadu of the Republic of India, below thirty years of age, as on at the time of writing this research paper, are passive bilinguals: in their ecology of Language use, predominance of Tamil in day-to-day activities and, English in the selected domains of registers, can be found.

By virtue of the use of two different languages in two different spheres of Life for more than three decades, one must agree, as a matter of fact, that a conflict between Tamil and English for dominance in space or use exists.

The aim of this article is to analyze such conflicts from the corpus point of view. Conflicts are analyzed for lexical, morphological and phonological modules for its effect.

Hence, in the following sections, lexical conflict will be given, at first, along the side of its process. It is followed by the nature of mergers at morphological and phonological levels from the resultative side of conflicts.

To begin with, let us say that a word, after the borrowal, sailing across the internal structures of Tamil language viz. semantics, morphology, phonology and phonetics, gets on to attain its societal acceptance and nativity to varying degrees of convergence ranging from converged to being converged.

1.1 Lexical Conflict

Borrowed words undergo **four stages** that are **bold lettered**, as in the below order, on the basis of their usage, irrespective of diachronic and synchronic status:

- a) **arrival**: synchronic usage. `pas']_n bus, `rōṭu']_n road , `caikkil']_n cycle, `rēṭiyō']_n radio and `ṭōr']_n 'door'. Words of this category wait to meet one of the results under (b), (c) and (d).

- b) merging:** being dominant, more **frequency** and **visible** in synchronic usage, skūl]_n 'school', paip]_n 'pipe', naṭṭu]_n 'nut', pōlṭ]_n 'bolt' and suviṭcu]_n 'switch' despite availability of translated and native equivalents, if any.
- c) died:** become a diachronic word, as in the example of tāpatam]_n. Usage is no longer required and the word died. It becomes a part of used words in the History of Tamil Language. Plenty of Sanskrit and native words too fall under this category.
- d) merged:** can't distinguish its nativity or source language. Speakers do say that word belongs to one's own language without knowing its etymology aṅkikāram]_n 'recognition', niccayam]_n 'definite', kōpam]_n 'anger', tāpam]_n 'lust', lāpam]_n 'profit', nastam]_n 'loss' and kulam]_n 'group'

Based on the above four stages of resultative side on the processes of borrowed words, consider the data given below from the corpus:

English Token	English Token found through Tamil Script	Number of attestations	English Token found in English script	Number of attestations
Daddy	tāṭi	22037	Yes	3854
Mummy	mammi	29693	Yes	1911
Sister	cisṭar	10370	Yes	4329
Brother	pratar	26230	Yes	986

The above table consists of borrowed words from English. It has five columns. In the first column, English lexemes (tokens) found in the Tamil corpus are given. Transliterated ones, use of English-English Script in Tamil writings and their respective frequencies can be found in second, third and fifth columns, respectively.

These - despite dominance of native lexemes (*appā* 'daddy', *ammā* 'mummy' *akkā* 'elder-sister' *aṅṅaṅ* and *tampī* 'brother\elder\ younger') - exist in the spoken and written varieties. Through the above given examples, one may say that the penetration of English kinship terms on Tamil society had taken place, and it is found under the category of two, **merging**. In short, these words have phono-semantic⁸ status.

These tokens - **visible** to native speakers irrespective of age and other social variables - have higher **frequency** of usage among the younger generation and give **pressure-dominance** to the respective native equivalent words.

These are all evidence to show that one of the outcomes, under the resultative side of the dominance of English language, is the presence of non-native kinship terms in Tamil.

⁸ Contrast against Lexeme. A lexeme has 1) grammatical category properties, 2) phonological properties and 3) semantic properties. Where as phono-semantic-word has 1) phonological properties, 2) semantic properties, but lacks grammatical category properties. In short, borrowed words of other languages do have phono-semantic properties, but assignment of a grammatical category is always difficult. Take for instance, a phono-semantic word from English. fan]_{vjn}. In English, it has dual category. In Tamils' usage of the word, can one assign a noun category? If so, what are the properties of Nounhood in Tamil for a suffix to be recognized as a noun suffix?

Hence, a non-native speaker of Tamil, always comes across presence and usage of two different patterns of terms in kinship with wider and different social meanings in various contexts in Tamil Society. In short, conflict for the dominance of space between the borrowed words and native words has been existing as shown in the above examples.

1.2 Morphological Conflict

Consider the data:

1. valicciñ]_n 'paining'
2. tūkkiñ]_n 'lifting'
3. mukka]s]_n 'groans'
4. moṇakals]_n 'groans'

The above data is morphologically very significant. To a morphologist, whose aim is to study the formatives of Lexemes, significance exists.

The first significance is on the suffix-*ing*]_{suf}. It is found with native lexemes, through the form of *-iñ*]_{suf}, such as *vali*]_n]iv 'pain']_n or 'to feel pain']_{iv}, *tūkku*]_{iv} 'to lift' for the derivation of another lexeme found in the spoken and written variety of urban, educated, economically middle class, and English educated youths.

The second significance is on the suffix *-s*]_{pl.suf}. It is found with pluralization in Tamil. Deverbal nominalized nouns, such as *mukka*]_n and *moṇakal*]_n are its bases. Observe, "as it is condition" of pluralization of English suffixes in Tamil. Also, observe the position of English plural suffix after its suffixation: same in the order of pluralization of Tamil, i.e. immediate right next to the lexeme. In short, one of the outcomes of dominance of English on Tamil is the function of pluralization through the form of English Suffix in the same order of Tamil Pluralization. What a penetration and merge!

1.3 Phonological Conflicts

Before seeing the data for phonological sides on merger, recall and observe that mechanism of production of speech sounds, pulmonic ones and vowels, remains singular and have oneness, irrespective of ecology of bilingualism, (read as, any two languages, here, for instance, Tamil and English) and multilingualism (read as, any three languages, such as Tamil, Kannada, French or Telugu).

To a set of phonemes, produced from the single mechanism of speech through speech organs, manner of articulation and place of articulation remain one and same, despite languages. For instance, the sound 'p', a common bilabial stop between Tamil and English, comes from a single manner and place of articulation from the speech mechanism in pronunciation of words that consist of /p/, as in English /pin/ and Tamil *pinpu* 'behind'. In short, a single mechanism to pronounce a single phoneme is available for two different languages. Another instance can be given, too: the sound 'm', a nasal bilabial, comes from a single manner and place of articulation from the speech mechanism in pronunciation of

words that consist of the phoneme 'm', as in, English 'make]_v]_n, Tamil /manam]_n/ 'soul', Telugu /manam]_n/ 'ours', and French /macon]_n/ 'smoked and salted mutton'.

As for a set of not-found or uncommon sounds between two languages, Tamil has a very straightforward approach: Borrow through more or less perceived nearness to the original sound. For instance, a bilabial stop /b/ as in baal]_n 'ball', bandh]_n 'strike', bandaari]_n 'personal name' and a vowel /æ/ as in /bænk⁹/ 'bank]_n'.

Thus, on account of the presence of singular phonetic mechanism for the production of Tamil phonemic units, there exists eight phonemic contexts irrespective of native or borrowed sounds:

1) a single manner of articulation, through a single place of articulation, generates a single phonemic sound, as in the examples of Tamil trill of alveolar /r/ as in /raudi/, English lateral approximant of alveolar /l/ as in /lav/.

2) a single manner of articulation, through a single place of articulation, generates more than a single sound, as in the examples of bilabial stops Tamil /p/ as in /paal/]_n 'milk' and /b/ as in /baal/]_n 'ball'.

3) a single manner of articulation, through many places of articulation, generates a single sound, as in fricatives of labio-dental, dental, alveolar, post-alveolar and glottal examples of English 1) 'f' as in five]_n /faiv/, 2) 'v' as in. very]_n /veri/, 3) 'θ' as in thanks]_n /θænkz/, 4) 'dʒ' as in zero]_n /dʒi:ro/, and 5) 'h' as in hen]_n /hen/

4) a single manner of articulation, through many places of articulation, generates more than a single sound. Examples are not found.

5) Bi or multi manners of articulation, through a single place of articulation, generates a single sound. Examples are not found.

6) Bi or multi manners of articulation, through a single place of articulation, generate more than a single sound. Examples are not found.

7) Bi or multi manners of articulation, through many places of articulation, generates a single sound. Examples are not found.

8) Bi or multi manners of articulation, through many places of articulation, generates more than a single sound. Examples are not found.

As a consequence and result to the above said first-three phonemic realities, combinations, found at synchronic Tamil, are in below and, the corpus analysis of sounds of written Tamil shows :

1. Speakers of Tamil may have added more sounds through existing places of articulation to pronounce sounds of English and other languages.

⁹ In Tamils' speech, the lexeme 'bank' is never pronounced with the front, open long vowel /e:/ as */beenk/.

2. On account of (1) merger must have taken place (for the details, see Keane, Elinor, 2004).
3. Due to (2), mérger is between two phonemes that are in nearness in place of articulation.
4. On account of (3), i.e. mergedness, words - that consist of merged sounds - have orthography representation only.
5. On account of (4) due to nearness in place of articulation, phonemic inventory must have less phonemes.
6. On account of (5), only certain phonemes from the manner of articulation will be visible in certain geographical areas
7. On account of phonetic facts from (2) to (5), manner of articulation may not have undergone any significant changes, i.e. no new manner of articulation is found or introduced.

The above observations are illustrated through data from our corpus. Consider the examples below to study the observation (1), which states that “Speakers of Tamil have added more sounds through their own existing places of articulation to pronounce sounds of English”.

Evidence for the claim comes from the below examples of English fricatives through Tamils’ manner of articulation of, as in below:

fricatives of English are found to be used are: 1) ‘f’ as in five]_n /faiv/, 2) ‘v’ as in very]_n /veri/, 3) ‘θ’ as in thanks]_n /θɛnks/ , 4) ‘dZ’ as in zero]_n /dZiro/, and 5) ‘h’ as in hen]_n /hen/. Also, a vowel /æ/ as in /bænk/ ‘bank]_n’.

From the data, it is found that the Tamil speakers had added fricatives in their spoken forms to pronounce English lexemes and, as a direct consequence, there must be a loss, too. To begin with, see the examples below:

- | | |
|------------------------|-----------------|
| 1) pārai | `rock’ |
| 2) karai | `dissolve’ |
| 3) palam | `strength’ |
| 4) pa _l am | `fruit’ |
| 5) pa _l lam | `pit’ |
| 6) nāy | ‘dog’ |
| 7) kaṇṇaṇ | `name of a God’ |
| 8) kaṇru | `calf’ |

The above examples are illustrative ones to the phonemes that are very nearer in place of articulation. Thus, you find, as in the below table, the presence of two r̥ trills, three l̥ approximants and three n̥ nasals.

Before going into the details, analyze the table below. The below table lists all phonemes that are very nearer in place of articulation. The columns consist of four major sections. In the first section, phonetic informations are given for the phonemes that have nearness in place of articulation. These phonemes, for instance, r̥ and r can be found in written Tamil. Their given phonetic values, for instance, r̥ and r as Trills. These trills are of alveolar and dental.

In the second section, social information of these sounds are given. Presence and absence in the spoken form of standard Tamil, its geographical restriction besides its status of merger are given in the next columns.

In the third section, it shows the status of the merger. In the last section, it lists examples, too. In short, the table answers a fundamental question: which phoneme merged with which type.

Phonetic Details			Social Details			Status of Merger	Information
Phoneme	Phonetic Value	Type	Spoken Standard Tamil	Geographic restriction	Merged	Merged with	Examples
r̥	Trill	Alveolar Dental	No	Yes	Yes	Dental Trill	pāraī is produced as pāraī.
r	Trill	Dental	Yes	No	No	No	karai is produced as karai.
l̥	Approximant	Retroflex	No	Yes	Yes	Retroflex	paḷam is produced as paḷam
l̥	Approximant	Retroflex	Yes	No	No	Retroflex	paḷlam is produced as paḷlam
l	Approximant	Alveolar	Yes	No	No	Retroflex	paḷam is produced as paḷam
n̥	Nasal	Retroflex	No	Yes	Yes	Alveolar	kaṇṇan is produced as kaṇṇan
n̥	Nasal	Alveolar	No	No	No	Alveolar	kaṇṇu is produced as kaṇṇu
n	Nasal	Dental	No	No	No	Alveolar	nāy is produced as nāy

Observe that alveolar trill consisting sound words got merged with dental trill sound words, as *pārai* 'rock' is produced as *pārai*, but *karai* 'sting' is produced as *karai*. In short, alveolar trill is merged with a dental trill.

Similarly, observe the loss of one approximant, i.e. three approximants becoming two approximants, as in the examples of *paḷam* 'fruit' is produced as *paḷam*. In a similar way, three nasal sounds became two in number, as in the examples of *kaṇṇaṇ* is produced as *kaṇṇan*.

On account of these mergedness, Tamil native speakers could find only an orthographic representation than that of psychological based phonemic realization of these merged sounds. In short, there is no psychological realization to these merged sounds, and due to this fact, i.e. psychological unrealisation of merged sounds, i.e. absence of letter-to-sound-realization at psychological levels, these sounds can be found either in writing or text-alone-phonemes, or in spelling errors such as *paṇṇai* as *pannai*, *aṇṇam* as *aṇṇam* or in different pronunciation in spoken forms, for instance, *kārru* as *kāttu*.

Due to these variations, it is worth, here, to mention that there exists correlation between these two gaps, i.e. unrealized and merged speech sounds against or to the presence of borrowed words: the borrowed words are found in against those words that have one of these merged sounds. This feature is found, predominantly, on the spoken forms, too. See the samples of quantification direction on the words that have phonemic mergedness:

No	Native Item	Gloss	Total Frequency of Native Items	Borrowed Item in transliteration	Total Frequency of Borrowed item in transliteration	Borrowed item in English Script	Total Frequency of Borrowed item in English script
	A				B		C
1	<i>pārai</i>	Rock	453733	rāk	74733	Rock	62784
2	<i>paḷam</i>	Fruit	734345	pruṭ	140675	Fruit	7549
3	<i>aṇṇam</i>	food	2193	mīḷs	22473	Meals	2473

The above table has three representative words. In each of these three sample words, merged sounds were found. In the data one, numerical combination of borrowed word (B+C), representing 30.30% to the total native word (453733), shows usage direction of mergedness of the merged phoneme alveolar trill found in '*pārai*'. A similar fate may hang on other sounds, too.

In order to verify whether "borrowal-against-mergedness" is of isolated one or "found-only-in-those-above-example-words," a brutal run is given against *ten other* high frequency different words consisting of one of those merged sounds.

In the table below, results are shown against a single variable: a nasal retroflex. The examples demonstrate borrowedness of words that do not have any features of mergedness. For other merged sounds, the work requires statistical verification that is being taken place. The inconclusive impression from the early results of data, the overall

trend, i.e. being emerged, is more or less same. The results are below for the above said nasal retroflex:

No	Native Item	Gloss	Total Frequency of Native Items	Borrowed Item in transliteration	Total Frequency of Borrowed item in transliteration	Borrowed item in English Script	Total Frequency of Borrowed item in English script
	A				B		C
1	kaṇṇāṭi	Glass	453733	mirar	8765432	Glass	87333
2	taṇṇīr	Water	267444	vāṭṭar	5432126	water	98282828
3	aṇṇam	Food	18000	puṭ	543210	food	5643211
4	aṇṇā	Elder brother	773233	pratar	8632155	brother	76543
5	uṇṇi	Parasite	9000	parācaṭ	976532	parasite	5463211
6	aṇṇi	Sister-in-law	237653	cis-iṇ-lā	09	sister-in-law	1802202
7	vaṇṇam	Color	9987	kalar	9234561	color	65432129
8	paṇṇai	Farm	78632	pārm	3246785	farm	282822
9	kaṇṇīr	Cry	4532111	krai	87	cry	92020
10	kaṇṇiyam	Decent	74532	ṭicaṇṭ	65321763	decent	6435325151
11	mannennai	Fuel	32101	keraciṇ	70928282	kerosin	563535353

In a similar way, more or less same directional results are expected for other merged sounds.

Conclusion

All these show that the mergedness, seen through corpus, gives dominance conflict between Tamil and English.

REFERENCE

- Annamalai, E. (1999). Language Choice in Education Conflict resolution in Indian courts. Great Britain: Language Science
- Annamalai, E. (2001). The Colonial language in Multilingualism and the Process of Modernization. In E. Annamalai (ed). Managing Multilingualism in India: Political and Linguistic Manifestations. New Delhi: Sage.
- Keane, Elinor. (2004). Illustrations of the IPA: Tamil. London : Journal of the International Phonetic Association.
- Ravishankar, G. (2019). Personal Communication
- Ravishankar, G. (2019). Intonation. Pondicherry: PILC