**Written to Spoken Tamil : Conversion Rules**
*(A Study in Speech Synthesis)*

K. Karunakaran and R. Krishnan  
*University of Malaya, Kuala Lumpur*

**Abstract:**
Tamil is a diglossic language for centuries. The two varieties which participate in the diglossic situation namely the written and spoken varieties show considerable change both in language structure and language use. These two varieties are taught at different levels for different purposes. There are learners interested in learning both the varieties, either one of the varieties, depending upon their needs.

When we try and used to introduce these two varieties in the teaching of Tamil as a second language/foreign language, there arises the need to convert one variety from the other. So, this paper focuses on written to spoken conversion by formulating a set of rules called conversion rules. The conversions start right from the phonetic level itself and extended up to the level of discourse. There are phonetic and phonological conversions of which some are significant; morphological conversions which are conditioned in many environments; there are morphophonemic rules which are exclusively found in each of these varieties and there are phrase and sentence formations found to be little different in them. So, the written to spoken conversions as found in the above mentioned levels are considerably ordered and systematically explainable. Hence, the aim paper is to demonstrate the way the conversions take place along with occurrences as well as conditioning factors. Though there are differences in the occurrences of the final phones-phonemes-morphemes, their occurrences show similarities also. The conversions show a pattern also, like word-final occurrences (before #), when followed by a suffix, phonetic and phonological significance, the phonetic realization and so on. The conversion rules include all such patterns in word formation, inflectional, derivational processes vice-versa process can also be explained, but needs…..

*(Key words: diglossia, conversion, varieties, conversion rules)*

1.0 **Introduction**

In the modern period dialectal words have found their way into the vocabulary domain of standard Tamil either by the virtue of their intrinsic worth or by the authority of the poets who pressed them into service (Sethupillai, 1953). The two distinct varieties had parallel development and functional significance since then and continue to be so even today. The two varieties not only have structural variation at different linguistic levels but also functional variation in the use of language in different socio cultural domains such as education, administration, mass media science and technology, home language and so on. That is each variety has different sets of social functions in formal and informal levels. These varieties do not overlap or merge much in the modern use, but for the movie language and dialogues and conversations among characters in novels and short stories in modern literature as well as some of the socio cultural usages in writings.

So, a study pertaining to the conversion of written to spoken and the vice versa is needed in order to make appropriate selection of usages for different purposes. With reference to the conversion written to spoken Tamil seems to be more systematic based on rule formation.

2.0 **Aims of the Study**

The following are the aims of the present study:

2.1 To generalize the regular written to spoken conversion at the word level.

2.2 To explain the conditions in the form of conversion rules at the word final and medial positions.

2.3 To identify and explain all those changes found in the morphological structures of Tamil in the form of conversation rules.

3.0 **Research Questions**

The following questions are put forward:
3.1 What are all the different word final and medial conversions to be generalized at the word level?
3.2 What type of conditions are needed to explain the word level conversions?
3.3 What are all the conversions found in the morphological structures and how to explain the conversions in the form of rules?

4.0 Methodology

Qualitative methodology is adapted in this study in order to present the exact variations and presenting the variations in the form of conversion rules from written to spoken Tamil. The rules are presented under two headings (i) Word Level Conversion Rules and (ii) Grammatical Level Conversion Rules.

The data for analysis includes words /lexical items found in modern spoken Tamil in the Malaysian contexts. However, those distinct variations found in the Tamilnadu spoken Tamil also have been taken into account, as this would help to generalize the conversion rules as far as possible. So, the frame work has two segments the use of spoken Tamil in different contexts and their functional significance on the one hand and that of making rules for systematic conversions.

5.0 Data Analysis

5.1 Word Level Variation and Conversions

5.1.1 Variations at the word final positions
5.1.2 Variations at the word medial positions
5.1.3 Some of the explainable variations at word initial position

5.2 Sample analysis

CR1 - ai # ➔ -e#
CR2 - am / -a:m / -a:n / -um / -o:m etc... ➔ nasalized vowels followed by #
CR3 - C # ➔ Cu # (phonetically i)
CR4 - y # ➔ yi # / -yyi # (the addition of –y or doubling of –yy has to be explained by a morphophonemic rule in the process of conversion.
CR5 - this conversion rule takes care of word medial conversion especially -rr- ➔ -tt- (spoken Tamil does not show use of R except in Kanyakumari Tamil but in the case of words with -rr- they are pronounced as -tt- with or without case suffix)
CR6 - this conversion rule takes care of word medial conversion especially –nr ➔ nn /-n (conditioned by initial syllable CV /CV)

Conversions at the Grammatical Level

5.3.1 Conversions in the Noun Structures

The conversions include : (i) Number Suffixes (ii) Gender Number Suffixes (iii) Link morphemes (iv) Case suffixes (v) Derivative Suffixes such as adjectival suffix, adverbial suffix, common human noun suffixes and so on.

There are numeral forms both cardinal and ordinal which show variations and they have to be explained in the form of conversion rules. The cardinal bases also show variations especially from 1-9.

So, at least there would be 15 conversion rules with reference to noun structures.

5.3.2 Conversions in the Verb Structures
The conversion rule includes all those variations found in the finite and non-finite verb forms in spoken Tamil.

5.3.2.1 Finite Verb Conversions

There are at least three types of finite verb forms namely (i) with tense suffixes, (ii) without tense suffixes, (iii) imperative permissive, hortative type of finite forms.

5.3.2.2 Non Finite Verb Conversions

The following non finite verb forms have to be explained in the form of conversion rules.

(i) Infinitive Forms
(ii) Verbal Participle Forms
(iii) Adjective Participle Forms
(iv) Conditional Participle Forms
(v) Temporal Participle Forms

5.3.2.3 Person – Gender – Number Suffixes which occur in finite Verb Forms after the tense suffixes

5.3.2.4 Negative suffixes, Negative words and Affirmative Negative differences in spoken Tamil

5.3.2.5 Participial Nouns and Verbal Noun Formations

5.3.2.6 Verbal Derivative Nouns (Verb based + Link Morpheme + Derivative Suffix)

There are complex verb forms as well as compound verb forms in spoken Tamil which show distinctive variations. So, there is a need to explain all such variations in the form of conversion rules.

Type 1 : Complex Verb Forms

There are two divisions under this namely aspectual forms and modal forms which show distinct conversion.

Type 2 : Compound Verb Forms (Verb form 1 + verb form 2 : Noun form + verb form etc.)

There are dependent parts of speech like adjectives, adverbs and particles which occur with noun forms, verb forms, noun and verb forms respectively. All these formations have to be explained in the form of conversion rules. So, all together there would be around 30-35 rules needed. The speech synthesis study with reference to modern Tamil have to be modernized by explaining conversions which are needed when one goes from written to spoken Tamil. The NLP study should try to formulate a series of ordered set of rules (conversions or correspondences).

This kind of standardization could be done with the help of data using modern methodology in NLP studies.