

In A Statistical Approach For Tamil Pulavarkal From The 3rd Bc To The 3rd Ad, The Grammars Mellinam And Idaiyinam Were Used

Arumugam K¹, Ramesh Ganesan^{2*}, Saviour Prakash Gnana Prakasam Louis Raja³, Kalimuthu S⁴, Sivasamy R⁵

¹Department of Computer Applications, Bharathiar University, Coimbatore, Tamilnadu, India.

^{2*}Department of Tamil, School of Social Science and Humanities, Central University of Tamil Nadu, India.

³Department of Computing Technology, SRM IST, Kattankulathur, Chengalpattu-03, Tamilnadu, India.

⁴Department of Tamil, Government Arts College (Autonomous), Kumbakonam-612002, Tamilnadu, India.

⁵Department of English, Government Arts and Science College, Kangeyam, Tamilnadu, India.

¹aaruk.dpm@gmail.com ^{2*}rameshg@cutn.ac.in

Abstract: Grammaticalization is a key notion that may be used to make sense of the study of different grammatical categories. The primary focus of this research is on the grammatical categories that can be found in a topologically generalised language. They have conducted in-depth study on Mellinam as well as Idaiyinam. The utilisation of this grammatical structure in both spoken and written linguistic interaction serves as the major focus of attention here. In addition to this, we have conducted extensive research on the multiplicity of factors that play a role in the development of a language. Researchers working with managed languages have resorted to quantitative approaches in order to arrive at credible results and accomplish their objectives. We investigated the many ways in which Tamil lyricists use Mellinam and Idaiyinam by employing the Latin Square Design methodology. Because of the way the data was structured, we were able to arrive at some reasonable inferences.

Keywords: ANOVA, Latin Square Design, Lyricist, TamilLyrics, Tamil Illakanam.

INTRODUCTION

There are 12 vowels (Uyireuttu) and 18 consonants (Meyyeluttukal) in the Tamil script, as well as one vowel (Aythaeuttu) (special character). This is known as "akku" in Tamil writing and is neither a Meyyeluttukal nor a vowel. In the vowel list, however, it is stated towards the end of the alphabet. Rather of an alphabetic script, it's a syllabic one instead. For a total of 247 combinations of a Meyyeluttukal and a vowel, a mute Meyyeluttukal, or a vowel alone (Uyirmeyyuttu, "soul body letters"), the script has 31 independent letters and an additional 216 combined letters (Uyirmeyyuttu, "soul body letters"). A vowel marker is added to the meyyeluttukal in order to

unite the letters. The meyyeluttukal bass form for some vowels requires a modification particular to that vowel. Meyyeluttukal may need a vowel-specific suffix in certain cases, while in others, a prefix and a suffix are both required. The vowel marker is distinct from the vowel characteristic in both cases. From left to right, Tamil letters are spelled.

LITERATURE STUDY

Many statisticians from all over the world have used statistical approaches to analyse languages over the last 120 years.

Historically, items (1), (2), (3), (4), (5), (6), and (7) have all contributed significantly to

the field of linguistic statistics. (Subbarayan 1995) compared the Warrens-Herdan distribution to a sample of Peyarccolkautilised by three authors of the MUVAR TEVARAM, a famous Tamil epic. Eighth, numerical techniques were emphasised as crucial to the study of languages. To get around the problems, special letters called consonant-vowel ligatures are utilised. Every other vowel is glued onto the consonant letter. Other vowels are attached to the consonant letter via unique letters called consonant-vowel ligatures (8). The use of enjambed punctuation and previously underutilised consonant clusters has helped both written and spoken communication. The existing Tamil grammar form's grammaticalization and development have been extensively explored, and new alphabets (9),(10), (11) are being examined. The compound grammar structure of the Dravidian language family was studied(12). Lyricists who contributed to the composition of Tamil lyrics were uncovered throughout the course of the study. Academics should be recognised as some of the earliest advocates for and contributors to the widespread use of statistical methods(13). According to, researchers interested in linguistic structures can use either the t-test or the analysis of variance (14). The ANCOVA and MANCOVA tests for analysing mean differences in grammar use were extensively discussed (15).

This is an excellent description of how to investigate language structure using analytical statistics. The debate based on the analytical method covers model formation, model selection, and other topics for a more in-depth investigation.

The newly constructed alphabets that are being used in the Tamil language have been examined in this research paper. There are three aspects to this.

1. Extraction and categorization of features
2. Feature extraction and character segmentation

3. Validation of results comparison with enhancement.

Authors for the Study Selection

3.1. Tamil as a language and literature

The majority of people who speak Tamil are located in India. Tamil is a Dravidian language. It is the official language of Tamil Nadu, which is a state in India, as well as of Pondicherry, which is a Union Territory (Pondicherry). Additionally, a sizeable population of people in Sri Lanka, Singapore, Malaysia, Mauritius, Fiji, and South Africa also speak it as their native language. Tamil was given the status of a classical language in India in 2004, which indicates that it meets the requirements for this category, which are as follows: it has ancient beginnings, an independent tradition, and a significant body of ancient literature.

At the turn of the twenty first century, there were 74 million native speakers of Tamil. Writing in Tamil is done horizontally, from left to right, and uses a basic set of 24 letters, 18 of which are consonants and 12 of which are vowels. Each consonant contains an inherent vowel, and the syllabic alphabet is used to write all of the letters in the alphabet.

3.2. The lyricist contributed significantly to Tamil's grammatical structure.

1. Paranar

It is only natural for people to want to lead their following by using their favourite celebrities' names as their nicknames. Paranar is one of the names that is kept in this category. In the ninth century, during the Sangam era, a talented poet by the name of Paranar produced a Patiyal book. 34 lines from his novel were included in the Panirupatiyal book.

2. Kapilar

5	Nalvelliyār	A5
6	Nakkīrar	A6

3.3. Dataset of the lyricist

Historians are not entirely in agreement over the order in which the literary works of Sangam were composed. According to several literary, archaeological, and numismatic sources, Sangam literature goes back to the time period of BC. It was constructed in the third century after the common era. There is a possibility that it was constructed in the third century. The

literary genre known as Sangam Literature may be traced back to its roots in India. In Sangam era society, you may find information about the political climate as well as the social situations. There are presently 2,381 songs from the Sangam catalogue that may be downloaded. There are a total of 2279 songs, however only 2279 of them list the poets who sang them. The identity of the poets who sung the other songs are now unknown.

Statistics about the lyric's characteristics	
Years range	3 rd BC - 3 rd AD
Lyrics	60
Words	850
Lyricists	6
Average songs/ Lyricists	10
Average words / Lyrics	15
Minimum words / Lyrics	12
Maximum words / Songs	30

Figure 1: At a fundamental level, statistics on the lyric's features.

A Summary of the Primary Data Structure

4.1. A description of the data structures of Vallinam, Mellinam, and Idaiyinam

To discover the differences in word usage

(Meyyeluttukal) Consonants are divided into three categories.

1. Vallinam (க்ச்ட்த்பற்) - MARBU (CHEST)
2. Mellinam (ங்ஞண்நம்ன்) - MOOKU (NOSE)
3. Idaiyinam (யர்ல்வழள்) - KALUTHU (NECK)

4.2. Meyyeluttukal- 18

GN,ங் - Both of these verbs, k and n, are born by touching the first part of the tongue, the palate.

NG,ஞ் - Both of these verbs ch and n are born by touching the middle tongue (middle tongue) middle palate.

ண் - T and N are born when the tip of the tongue and the tip of the chin are touched.

N,ந் - Matching the base of the upper tooth to the tip of the tongue produces the letter T.

M,ம் - Both p and m, which correspond to the top and lower lips, are born.

ய் - Y, By connecting the base of the tongue to the base of the upper lip, the letter Y is formed.

ர, ற், ல, ழ் - The verbs r and z are created by rubbing the tip of the tongue on the roof of the mouth.

ல, ள் - The thickening of the tongue's borders at the base of the upper tooth gives birth to the letter L.

ல, ள் - rubbing the top lip and the margins of the tongue produces the letter L.

வ, வ் - The letter V is formed by the upper and lower teeth coming together.

ந, ன் - Because the tip of the tongue is so close to the top lip, both of these verbs, r and n, are born.

Data structure algorithms for Mellinam, and Idaiyinam

The method of algorithm is utilised to count the letters from the words, which lyricists written lyrics were saved in a text file referred to as a string array as read file and entitled Tamilsong.txt. In addition, the grammatical structure of the sentence is evaluated according to the criteria that we have established.

```

Procedure Meyyeluttukal Count:
readFile ← FileRead ("TamilSong.txt", "r")
Song [] ← StringArray(readFile)
S ← Song []
Mellinam[] ← ['ங்', 'ஞ்', 'ண்', 'ந்', 'ம்', 'ன்']
Idaiyinam[] ← ['ய்', 'ர்', 'ல்', 'வ்', 'ழ்', 'ள்']
CharecterCount(Song, Mellinam)
CharecterCount(Song, Idaiyinam)
CharacterCount(song, c)
    fori in range(length(song))
        if (song[i] == c)
            count = count + 1
print( c → count)

```

Algorithm 1: Identifying Mellinam, and Idaiyinam

Data Structure Construction

5.1. Latin Square Data Representation and Design

This module creates Latin Square designs. There are designs ranging from three to 10

treatments that may be chosen from. Latin Square patterns are very comparable to randomised block patterns; however, rather than eliminating a single blocking variable, these patterns are meticulously created to enable the removal of two blocking variables. They are able to achieve this while also

reducing the number of experimental units that are required to carry out the investigation.

When utilising the Latin Square layout, it is essential to have a solid understanding of the assumptions that are being made. As a result of making the assumption that the magnitudes of the interaction terms are sufficiently modest to be ignored in this design, the number of experimental units that are necessary has been significantly reduced. As an illustration, the Latin Square design is focused mostly on primary impacts. The treatments, the row factor, and the column factor are all factors that can have an effect on the response. This is another way of saying it.

It is imperative that the randomization procedure be adhered to stringently whenever probability statements are being formulated during the analysis of experimental data.

Randomization is accomplished by the use of the following method:

1. Pick an orthogonal layout at random from the possibilities that are given to you.
2. Using a randomization method, assign different levels of the row component to each row.
3. Assign the levels of the column factor to the columns using a randomization method.
4. Distribute the treatments according to the treatment letters in a random order (or numbers as the case may be).

Each design may have as many as six different treatments. The amount of treatments that are

offered is directly proportional to the number of squares that are available. The number of orthogonal squares that were recorded throughout this method is displayed in the table below.

In the table that is seen above, the letters A1, A2, A3, A4, A5, and A6 indicate the six different treatments that are available. As a result of the way the letters are ordered, there is only one instance of each letter inside each row and column. It's important to note that a straightforward random design would need for 216 different experimental units (6 x 6 x 6). This Latin Square needs just 36 experimental units, which is a 75 percent decrease from the previous need!

A sixth component of the design may also be eliminated by using a different set of letters, but this time with the case changed to lower case. The Latin Square is the name given to this particular design. Our technique only requires 36 experimental units, which represents a savings of around 94 percent in terms of experimental units compared to what would ordinarily be required to test six variables over six levels. Combining together two Latin Squares that face in opposite directions from one another. Latin Squares may be performed on any treatment numbers other than the number six.

The following is an experiment with Latin Square with six different treatments. The following is how the experimental design is laid out: There is a diagram below here.

GS	C1	C2	C3	C4	C5	C6
R1	A1	A2	A3	A4	A5	A6
R2	A2	A3	A4	A5	A6	A1
R3	A3	A4	A5	A6	A1	A2
R4	A4	A5	A6	A1	A2	A3
R5	A5	A6	A1	A2	A3	A4

R6	A6	A1	A2	A3	A4	A5
----	----	----	----	----	----	----

Figure 2: Latin Square with six treatments.

The grammatical structure of the above features is denoted by GS.

5.2. Data Structure: Mellinam, and Idaiyinam Authorship and Usage

Two-directional variations are researched in a variety of circumstances in order to acquire accurate estimates when comparing things. This is accomplished through the use of a mathematical framework known as the Latin Square Design.

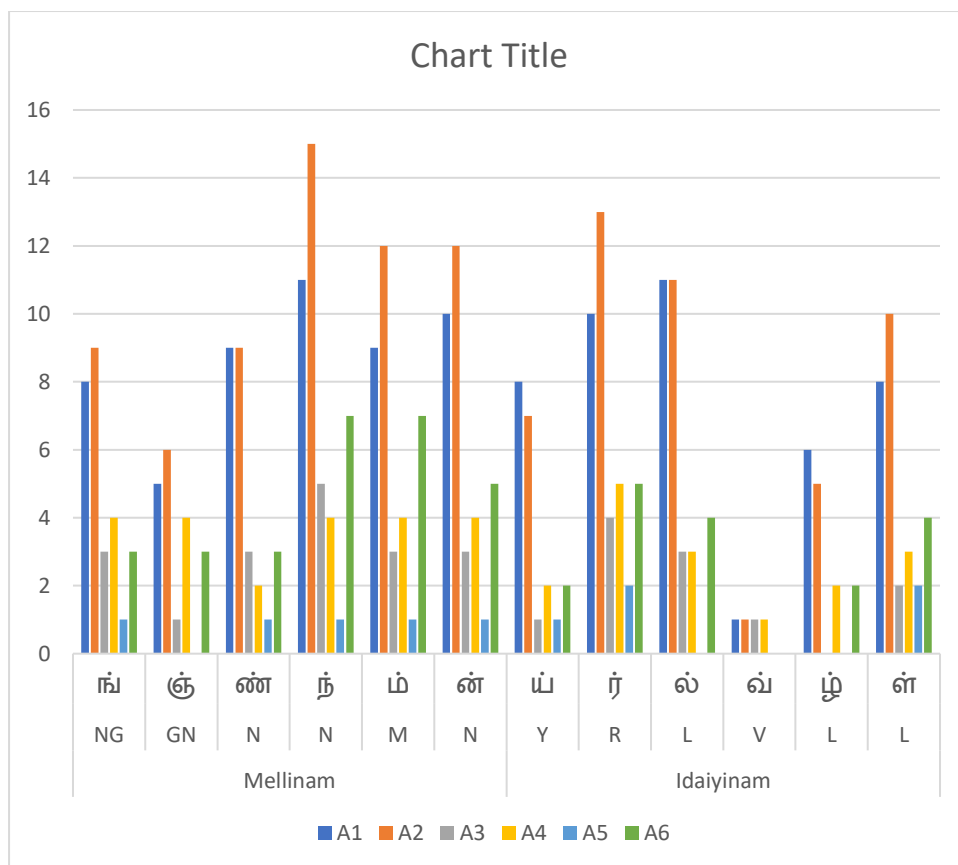
Table 2: The following table shows the aggregate data structure for Mellinam, and Idaiyinam.

AUTHORS	Mellinam						Idaiyinam					
	NG	GN	N	N	M	N	Y	R	L	V	L	L
	ங்	ஞ்	ண்	ந்	ம்	ன்	ய்	ர்	ல்	வ்	ழ்	ள்
A1	8	5	9	11	9	10	8	10	11	1	6	8
A2	9	6	9	15	12	12	7	13	11	1	5	10
A3	3	1	3	5	3	3	1	4	3	1	0	2
A4	4	4	2	4	4	4	2	5	3	1	2	3
A5	1	0	1	1	1	1	1	2	0	0	0	2
A6	3	3	3	7	7	5	2	5	4	0	2	4

Table 2 shows the author's utilisation of the grammar structures for the above-mentioned reference.

Figure 3: Use of Grammar Structures by the Author

GRAPH



This design is used to investigate the differences between:

1. Mellinam and Idaiyinam

5.2.1. Mellinam and Idaiyinam

The information is shown in the table below.

1. Row-by-row data is related to the Mellinam.

2. The information in the columns is related to the Idaiyinam.

3. Mellinam and Idaiyinam, the writers' names, are used with their permission.

Mellinam and Idaiyinam's data structures are shown in Table 7.

The following table shows the data structure for Vallinam and Mellinam..

AUTHORS	Mellinam						Idaiyinam					
	NG	GN	N	N	M	N	Y	R	L	V	L	L
	ங்	ஞ்	ண்	ந்	ம்	ன்	ய்	ர்	ல்	வ்	ழ்	ள்
A1	8	5	9	11	9	10	8	10	11	1	6	8
A2	9	6	9	15	12	12	7	13	11	1	5	10
A3	3	1	3	5	3	3	1	4	3	1	0	2
A4	4	4	2	4	4	4	2	5	3	1	2	3
A5	1	0	1	1	1	1	1	2	0	0	0	2
A6	3	3	3	7	7	5	2	5	4	0	2	4

Mellinam and Idaiyinam's average data structure is shown in Table 8.

The following table shows the average data structure for Mellinam and Idaiyinam.

GS	I1	I2	I3	I4	I5	I6
M1	8	11	3	3	1	4
M2	7	3	4	0	3	7
M3	2	4	1	2	8	10
M4	3	2	6	4	10	4
M5	1	5	10	7	2	4
M6	4	10	12	2	3	2

GRAPH

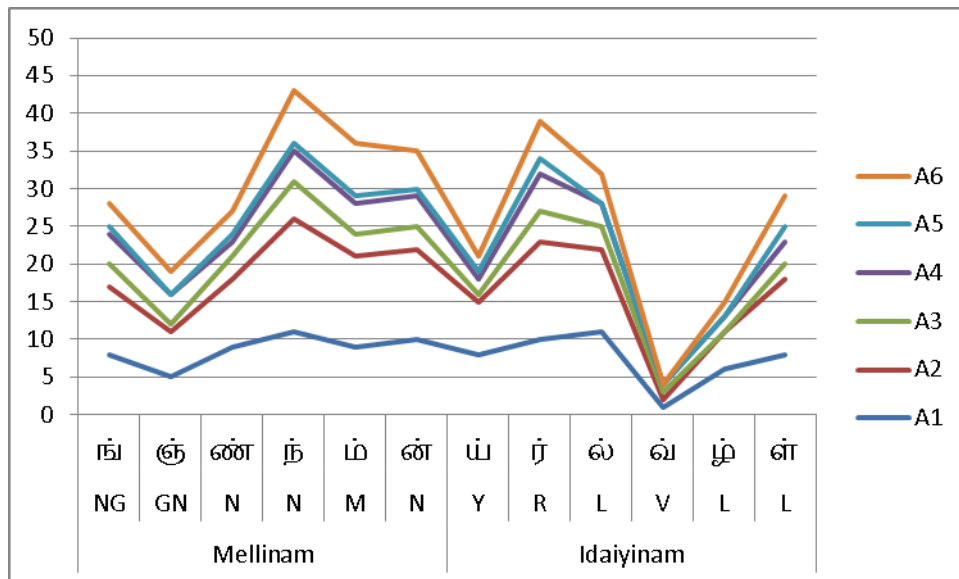


Figure 4: Parent data Mellinam and Idaiyinam

The following are the steps involved in the analysis:

N=No. of Observations = 36

G= Grant total of Observations = 172

Correction Factors = $(172)^2/36$

Total sum of squares = 372.2222

Between Rows sum of squares (Mellinam) = 7.5555

Between Columns sum of squares (Idaiyinam) = 38.2222

Between Authors sum of squares= 308.2222

Error Sum of squares = 18.2222

Data structure-based empirical analysis

The data was organised into the aforementioned six columns. The value of the Vallinam (column) is stored in the first Mellinam (column). The value of the column may be found in the second column of the table. The

letter pertaining to therapy may be found in the third column. Use the laws of randomness that were discussed before when designing the architecture of your experiment so that you may put this theory into practice.

After you have completed the experiment, you will need to replace the random values in cell A6 with the findings from the experiment. After that, the LM approach would be utilised to perform an analysis on the data. The first factor, which represents the row, would be fixed (or random, depending on the circumstances), the second factor, which represents the column,

would also be fixed (or random), and the treatment, which represents the column, would be fixed. A6 would be the response variable in this case. Alter the Which Model Terms drop-down menu in the LM ANOVA technique's Model window to read "Up to 1-Way."

As a consequence of this, the tool is compelled to generate an error term based on all of the interaction terms. In this fashion, the results would be open to inspection.

The following table presents the findings from doing an analysis of variance on 5.2.1.

Table 3: Analysis of Variance Table (based on 5.2.3)

Sources of variation	Degrees of freedom	Sum of squares	Mean Sum of squares	F ₀	F _e
Mellinam	5	7.5555	1.5111	1.658536585	2.71
Idaiyinam	5	38.2222	7.6444	8.390243902	2.71
Authors	5	308.2222	61.6444	67.65853659	2.71
Error	20	18.2222	0.91111		
Total	35	372.2222			

The data was organised into the aforementioned six columns. The Mellinam (row) value is located in the first Idaiyinam of the table (column). The value of the column may be found in the second column of the table. The letter pertaining to therapy may be found in the third column. Use the laws of randomness that were discussed before when designing the architecture of your experiment so that you may put this theory into practise.

After you have completed the experiment, you will need to replace the random values in cell A6 with the findings from the experiment. After that, the LM approach would be utilised to

perform an analysis on the data. The first factor, which represents the row, would be fixed (or random, depending on the circumstances), the second factor, which represents the column, would also be fixed (or random), and the treatment, which represents the column, would be fixed. A6 would be the response variable in this case. Alter the Which Model Terms drop-down menu in the LM ANOVA technique's Model window to read "Up to 1-Way."

As a consequence of this, the tool is compelled to generate an error term based on all of the interaction terms. In this fashion, the results would be open to inspection.

GRAPH

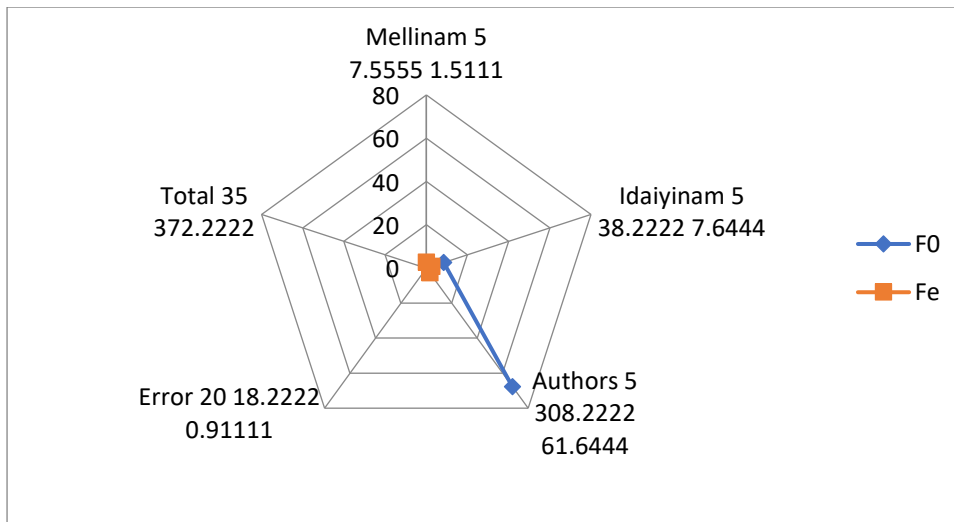


Figure 5: ANOVA Table of Mellinam, Idaiyinam and Authors.

There will be a difference between the F-ratios, the total of squares, and the mean squares that are shown below. However, there will be no difference in the amount of degrees of freedom. It is important to keep in mind that the error term has a total of just six degrees of freedom (S). This is a significant challenge for a Latin Square layout that consists of only six levels. In order to maximise the potential for error and increase the number of degrees of freedom, you would repeatedly carry out the experiment. It is also important to point out that in order to compute the Expected Mean Square values, each and every assumption made in the model must be used. These numbers are inaccurate due to the fact that the Latin Square is not finished (it does not include all row-by-column-by-treatment combinations).

Result and Discussion

6.1. Mellinam and Idaiyinam Inference

For drawing inferences based on the results reported in the ANOVA table, we see the following sources of variation.

1. For the space between the rows, we get the values $F_0 = 1.51$ and $F_e = 2.71$ (Mellinam). Due

to the fact that F_0 is equivalent to F_e , we are able to draw the conclusion that there is no distinction in Mellinam word use.

2. For the space between the columns, we get the values 7.64 and 2.71 for F_0 and F_e respectively (Idaiyinam). Since F_0 is more common than F_e , we may conclude that there is a significant difference in the way that Idaiyinam words are used.

3. In terms of the authors, the values $F_0 = 61.64$ and $F_e = 2.71$ stand out as significant. Because F_0 is greater than F_e , we may assume that the way Mellinam and Idaiyinam are used is very distinct from one another.

Conclusion

The conclusions of the study, which are fascinating, are based on an analysis of the ways in which Tamil lyricists have utilised various grammatical structures throughout the course of the past fifty years. Mellinam and Idaiyinam are the linguistic foundations upon which the lyricist's construction is based.

The formation of the words is assisted by these many form variables. This strategy involved analysing the compositions of six lyricists who have made significant contributions to the field over the course of time.

A one-of-a-kind configuration has been chosen for the data system. In the recent years, just a handful of research on quantitative grammatical structure have been carried out, which is something that need to be brought to people's attention. The Latin square design paradigm was used as the basis for a comprehensive investigation of the organisation of the grammatical language.

The research makes it abundantly evident that lyricists writing in the Tamil language utilise grammatical structure in a variety of quite distinct ways.

References

1. Greg WW, Yule GU. The Statistical Study of Literary Vocabulary. *Mod Lang Rev.* 1944;39:291.
2. Mosteller F, Wallace D. Inference and disputed authorship: The Federalist.(1964). 1964;
3. Posner R, Herdan G. Quantitative Linguistics. Vol. 129, *Journal of the Royal Statistical Society. Series A (General)*. 1966. 138 p.
4. Doležel L. A framework for the statistical analysis of style. *Stat style.* 1969;10–35.
5. Sinclair JM. CB Williams," Style and Vocabulary: Numerical Studies"(Book Review). *Yearb English Stud.* 1977;7:182.
6. Mephram M. Introduction to the Mathematics of Language Study, by Barron Brainerd. (Mathematical Linguistics and Automatic Language Processing, 8). New York: American Elsevier, 1971. Pp. ix + 313. *Can J Linguist Can Linguist [Internet]*. 2016/06/27. 1973;18(2):181–3. Available from: <https://www.cambridge.org/core/article/introduction-to-the-mathematics-of-language-study-by-barron-brainerd-mathematical-linguistics-and-automatic-language-processing-8-new-york-american-elsevier-1971-pp-ix-313/1B07C033C15D707192C1D92543D07017>
7. BRAINERD B. Weighting Evidence in Language and Literature [Internet]. University of Toronto Press; 1974. Available from: <http://www.jstor.org/stable/10.3138/j.ctt15jjcf0>
8. Steever SB. Verb+ verb sequences in Dravidian. In: *Verb-Verb Complexes in Asian Languages*. Oxford University Press; p. 327–53.
9. Fienberg SE, Hoaglin DC, Tanur JM. Frederick Mosteller 1916-2006. 2013;
10. Annamalai E. Lexical anaphors and pronouns in Tamil. In: *Lexical Anaphors and Pronouns in Selected South Asian Languages: De Gruyter Mouton*; 2011. p. 169–216.
11. Mohanlal S, Sharada BA, Fatihi AR, Gusain L, Karunakaran K, Marie Bayer J, et al. LANGUAGE IN INDIA Strength for Today and Bright Hope for Tomorrow PARSING IN TAMIL – PRESENT STATE OF ART PARSING IN TAMIL: PRESENT STATE OF ART. *LanguageinindiaCom.* 2006;6(August):1–14.
12. Sankaravelayuthan R. “izations” of Noun Modifying Expressions in Tamil. 2020.
13. Saviour Prakash Gnana Prakasam Louis Raja; Ramalingam Viswanathan Venkatesan. TAMIL LYRICISTS' USE OF GRAMMATICAL STRUCTURE: A NON - PARAMETRIC STATISTICAL APPROACH USING THE FRIEDMAN MODEL [Internet]. Vol. 20. 2022. Report No.: ISSN: 669-2481 / eISSN: 2669-249x. Available from: <https://zenodo.org/record/7068241#.YyQGgKRBzIV>
14. Renganathan V. Expressives in Sangam, Medieval and Modern Tamil. In:

- Expressives in the South Asian Linguistic Area. Brill; 2020. p. 125–53.
15. Gries ST. Statistics for Linguistics with R: A Practical Introduction [Internet]. De Gruyter Mouton; 2021. Available from:
<https://doi.org/10.1515/9783110718256>
- .