Comparative Constructions in Tamil and English: A Comparative Study with the existing Machine Translation systems

ABSTRACT

The concept of comparative construction is a linguistic technique used to demonstrate the similarities and differences between two or more items. Typically, a comparative construction is composed of a predicate and two noun phrases. One noun phrase serves as the "criterion" of the comparison, while the other functions as the object of comparison (the comparee NP). Sentences like "Raja is taller than John" exemplify comparative constructions, wherein the noun phrase following the word "than" functions as the standard NP. A prototypical comparative construction includes three primary elements: the participants of comparison (comparee and standard of comparison), the property (parameter of comparison), and the index of the comparison. However, it's important to note that this paper does not explore into the typological study of comparative constructions. Instead, it focuses on the comparative constructions in Tamil as compared to English, and seeks to formulate rules for Machine Translation. The paper also involves testing the translation of comparative construction sentences from Tamil to English through existing Machine Translation systems, and discusses the obtained results.

Keywords:*Comparative Construction, Tamil language, comparee, Standard of comparison,Degree Marker,Comparison of Inequality,Comparison of Equality,Machine Translation*

1. INTRODUCTION

A prototypical comparative construction involves a quality or property whose extent is compared, the entity being compared, and the standard of comparison. Comparison is a mental act by which two or more items are examined in order to assess similarities or differences between them. The comparison can be made with regard to a certain gradable, one-dimensional property, and the items are then assigned a position on a predicative scale. This mental act of comparison finds its linguistic encoding in comparison constructions, especially comparative constructions for the expression of comparison of inequality or equative constructions for the expression of comparison of equality. The study of comparative constructions across different languages reveals both universal patterns and language-specific variations, offering insights into the principles governing human language. In this research, we have tested and compared the grammatical structures found in the Tamil language with those present in English. This extensive analysis has enabled us to establish a set of rules that can be applied to facilitate Machine Translation between the two languages. To validate the effectiveness of our findings, we conducted thorough testing of the translation process for comparative construction sentences in both Tamil and English using an established Machine Translation system. The outcomes of these tests have been meticulously analyzed and are presented for discussion and evaluation.

2. LITERATURE REVIEW

The foundation work for understanding the structure of comparatives was established by Bresnan (1973), suggesting that comparatives are created through transformations from underlying deep structures. Government and Binding Theory by Chomsky (1977) further investigated the syntactic properties of comparative constructions by emphasizing the role of movement and empty categories. Recent studies, such as Kennedy and Merchant's work in 2000, have concentrated on the connection between syntax and semantics in comparatives. Extensive research has been conducted on the semantics of comparatives. Heim (1985) introduced the degree-based approach, which analyzes comparatives in terms of degrees, proposing that adjectives have an implicit degree argument that gets bound by the comparative morpheme. This viewpoint was expanded by Kennedy (1999), who proposed a scalar semantics approach, suggesting that adjectives map entities to points on a scale. Comparative constructions involve not only syntactic or semantic phenomena but also pragmatic aspects. Klein (1980) discussed the role of context in interpreting comparatives, introducing the concept of "standard of comparison," which is contextually determined. The role of focus and presupposition in interpreting comparatives was further explored by Büring (2007), demonstrating how different focus structures can give different comparative readings. Comparative constructions are not exclusive to English and display intriguing cross-linguistic variations. Beck et al. (2004) compared English comparatives with those in German and other languages, uncovering significant syntactic and semantic differences. McCawley (1988) analyzed Comparative correlative structures from a transformational perspective, while recent work by Culicover and Jackendoff (2005) examined them from a construction grammar viewpoint, emphasizing their fixed form and meaning pairings. Psycholinguistic studies have contributed to understanding comparatives. Fults and Phillips (2004) investigated real-time processing of comparatives, providing evidence for the cognitive mechanisms involved in their comprehension.

Several theoretical frameworks have been employed to analyze comparative constructions cross-linguistically. Functional and typological approaches, as seen in the work of Croft (2003), focus on the roles and functions of these constructions across languages. Comparative constructions exhibit diverse syntactic structures across languages. Kennedy (2005) illustrates the structural differences and highlights the parameter settings in Universal Grammar. The use of comparatives in discourse varies across languages and cultures. Pragmatic factors, such as politeness and emphasis, influence the choice of comparative structures. Wierzbicka (1991) discussed how different cultures use comparatives to convey subtle social meanings, and how these pragmatic uses are reflected in linguistic forms. Bisang (1992) analyzed Comparative correlative structures in East Asian languages, noting that while the function is similar, the syntactic realization differs significantly. Research by Jindal and Liu (2006) explored opinion mining using comparatives to extract sentiment information. Comparative constructions present challenges for translation systems as accurately capturing the structural and semantic differences is necessary to maintain meaning across languages.Typological studies, such as those by Dixon (2008), have cataloged the variations in comparative constructions across languages. Dixon (2005) makes use of the terms comparee, index, parameter, mark and standard. The traditional grammar of English has four distinguished degrees of comparison: Positive degree, Equative degree, Comparative degree, and Superlative degree as discussed by Yvonne Treis (2018).

Tamil language does not use of degree marker or parameter marker; it makes use of only parameter of comparison. In Tamil, there is no morphological distinction between positive degree, comparative degree and superlative degree. Rajendran (1976) elaborately studied comparison of inequality and equality in Tamil.

3. COMPARISON OF INEQUALITY

Yvonne Treis (2018) has discussed that the linguistic literature has especially been concerned with comparison of inequality and comparative constructions as found in the following English sentences.

(1) Mary is tall-er than Peter -Comparee - Parameter - Parameter/Degree Marker - Standard Marker - Standard

(2) Mary is more intelligent than Peter - Comparee - Parameter/Degree Marker - Parameter - Standard Marker - Standard

In a prototypical comparative construction in Tamil and English, the comparee occupies the subject position, and the standard of comparison occupies the predicate position. In English, the standard of comparison occupies a position at the end of the comparative construction after the parameter of comparison whereas in Tamil, the standard of comparison occupies the predicative position before the parameter of comparison. The standard of comparison is marked for accusative case in Tamil. The parameter of comparison does not make use of a comparative degree marker; it is pronominalized to agree with the subject NP. English makes use of than as parameter of comparison and Tamil makes use of viTa or kaaTTilum as parameter of comparison.

(3) raaNiraataiy-ai vita/kaaTTilumazahkaana-vaL

Rani Radha-ACC POC ADJ-PN

' Rani is more beautiful than Radha '

The mapping rules between Tamil and English can be given as follows:

Mapping rule 1:

NP_{COM} + NP-ai + vita/kaaTTilum + ADJ-PN = NP_{COM} + BE + more + ADJ + than + NP_{SOC}

In the superlative comparative construction, the parameter of comparison is marked for superlative degree in English. In Tamil, the parameter of comparison is not marked for superlative degree. The standard of comparison has to be an inclusive NP of superlative nature meaning 'of all', 'among all' and so on. In Tamil too, the standard of comparison in superlative comparative construction must be inclusive nature: *avarkaLelloorilum* 'among all.

(4) raaNiavarkaLelloor-ai-yum viTaahakaana-vaL

Rani they all-ACC-EMP than ADJ-PN

Rani is most beautiful among all

Mapping rule 2:

NP+ NP-ai + viTa/kaaTTilum + ADJ-PN = NP_{COM} + BE + most + ADJ + among all.

In English, certain adjectives inflect for comparative degree is marked with -er instead of *more*.

(5) raaNiraataiy-aiviTa/kaaTTilumuyaramaanavaL

Rani Radha-ACC POC ADJ-PN

Rani is taller than Radha

Mapping rule 3:

NP_{COM} + NP-ai + vita/kaaTTilum + ADJ-PN = NP_{COM} + BE + ADJ-er + than + NP_{SOC}

Some adjectives in English, inflect for comparative marker -est instead of *most* .

(6) raaNiavarkaLelloor-ai-yum viTauyaramaana-vaL

Rani they all-ACC-EMP than tall-PN

Rani is tallest among all

Mapping rule 4:

NP+ NP-ai + viTa/kaaTTilum + ADJ-PN = NP_{COM} + BE + ADJ-est + among all.

In the place of ADJ-PN Tamil can makes use of noun denoting quality +adverbial marker when followed by the be-verb *iru*. English makes use of be-verb and adjective combination only. The following sentence will exemplify this statement.

(7) raaNiraataiyaiviTa/kaaTTilumazhak-aakairu-kkiR-aaL

Rani Radha-ACC than beauty-ADVP be-PRE-3FS

Rani is more beautiful than Radha

Mapping rule 5

NP_{COM} + NP-ai + vita/kaaTTilum + N-ADVP iru-TEN-PNG= NP_{COM} + BE + more + ADJ + than + NP_{SOC}

Similar to adjectives, adverbs too can make comparison of inequality. Consider the following example.

(8) raaNiraataiy-aiviTa/kaaTTilumveekamaakaooT-in-aaL

Rani Radha-ACC than fast run-PAS-PNG

Rani ran faster than Radha

In Tamil, the standard of comparison is marked for accusative marker; the standard marker *viTa*/*kaaTTilum* comes next and the parameter of comparison which is an adverb comes after standard marker and before the verb. In English, the verb comes before the parameter of comparison which is an adverb which is inflected for the comparative marker -er; the

standard of comparison comes at the end and the standard marker *than* comes before the standard of comparison.

Mapping rule 6:

NP_{COM} + NP-ai + viTa/kaaTTilum + ADV +V-TEN-PNG = NP_{COM} + V-TEN + ADV-er + than + NP_{SOC}

4. COMPARISON OF EQUALITY

If two or more items are found to be similar quantitatively or qualitatively the can be subjected to the comparison of equality. Consider the following sentence.

(9) raaNiraataiy-aippoolaazhakaana-vaL

Rani Radha-ACC like ADJ-PN

'Rani is as beautiful as Radha'

In the above sentence, Rani is the compare, i.e. item compared; Radha is the standard of comparison; *poola* is the standard marker; and*azhakaana* 'beautiful' is the parameter of comparison. In English, the standard of comparison comes at the end; the parameter of comparison comes in-between the standard marker as-----as.

Mapping rule 7:

NP + NP-ai + poola + ADJ-PN = NP + BE + as-ADJ-as + NP_{SOC}

Similar to adjectives, adverbs also undergo comparison of similarity. Consider the following sentence.

(10) raaNiraataiy-aippoolaveekamaakandaTa-kkiR-aaL

Rani Radha-ACC like walk-PRE-3FS

'Rani walks as fast as Radha'

In Tamil, the subject function as the comparee; the standard of comparison marked for accusative case follows it; the standard marker *poola* follows next; the parameter of comparison which is an adverb follows it and the verb which inflect for tense and person-number-gender (PNG) occupies the final position of the construction. In English, the subject NP function as the compare. The verb which is inflected for tense comes next. The parameter of comparison comes in between the standard marker '**as---as'**.

Mapping rule 8:

NP_{COM} + NP-ai + poola + ADV + V-TEN-PNG = NP_{COM} + V-TEN + as-ADV-as +NP_{SOC}

Comparison can be made without the explicit expression of adverb. In that context, the comparative construction become ambiguous. Consider the following example:

(11) raaNiraataiy-aippoolaooTu-kiR-aaL

Rani Radha-ACC like run-PRE-3FS

Rani runs like Radha

The Tamil sentence is ambiguous as it can be interpreted in a number of ways: 'Rani, runs (instead of walking) like Radha', 'Rani runs in the same speed like Radha', Rani runs in the same style or manner like Radha' and so on. Consider the following example,

(13) raaNi-kkuraataiy-aippoolaceelaiiru-kkiR-atu

Rani-DAT Radha-ACC like sari be-PRE-3NS

'Rani has sari like Radha'

The Tamil sentence is ambiguous inviting different interpretations: Rani has similar sari like Radha, Rani has sari of same colour like Radha, Rani has sari of same texture like Radha, and so on.

Comparison can be made without explicitly expressing the parameter of comparison. Consider the following example:

(14) raaNiraataiy-aippoolairu-kkiR-aaL

Rani Radha-ACC like be-PRE-3FS

` Rani resembles Radha.

The not-expression of parameter of comparison makes this sentence ambiguous allowing different interpretation from the point of view of quality and quantity.

Mapping rule 9:

NP_{COM} + NP-ai + poola + iru-TEN-PNG = NP_{COM} + resemble-TEN + NP_{SOC}

iru can be replaced by *toonRu* 'appear' in the above construction of equality.

(15) raaNiraatay-aippoolatoonRu-kiR-aaL

Rani Radha-ACC like appear-PRE-3FS

Rani appears like Radha'

The addition of emphatic -ee can make the resemblance more closer.

(16) raaNiraataiy-aippoolav-eeiru-kkiR-aaL

Rani Radha-ACC like-EMP be-PRE-3FS

` Rani resembles Radha very much.

Instead of *poola* 'like', *maatiri* 'like', aLavukku' as much', attanai 'that many' can be made use of as standard marker.

(17) raaNikku raataiy-ai maatiri pasi.

Rani-DAT Radha-ACC like hunger

'Rani is hungry like Radha'

(18) raaNikkuraataiy-aiaLavukkupasi.

Rani-DAT Radha-ACC that much hunger

'Rani is that much hungry like Radha'

(19) raaNikkuraataiy-aiattanaipasi.

Rani-DAT Radha-ACC that much hunger

'Rani is as much hungry as Radha'

aLavu and *attanai* specifies quantity. Another way of expressing quantity for the sake of comparison is using *etttanai* 'how much' and *attanai* as exemplified in the following comparative construction.

(20) raaNi-kkuettanaiceelaiiru-kkiR-at-ooattanaiceelairaatai-kk-um iru-kkiR-atu

Rani-DAT how-much be-PRE-3NS-Q that-much saree Radha-DAT-EM be-PRE-3NS

'Radha has as many as saris Rani'

ettanai---attanai, *evvaLavu--- avvaLavu* can be equated with *as many as* and *as much as* respectively. Similarly eppaTi---appaTi and *evvaaRu--avvaaRu* can be equated with English *what manner--that manner* as exemplified in the following example.

(21) raaNieppaTiooT-in-aaL-ooappaTiraat-aiy-um ooT-in-aaL

Rani how run-PAS-3FS-Q that-manner Radha-ACC-EMP

Rani waked in the same manner like Radha'

Exact resemblance can be expressed by making use of the emphatic markers -ee and taan as exemplified by the following sentence.

(22) raaNiraatai-eetaan

Rani Radha-EMP EMP

'Rani is exactly like Radha'

camamaaka 'equally', iNaiyaaka 'equally' can be used to specify the exactness in the resemblance.

(23) raaNi raatai-kku samamaaka/iNaiyaaka paaT-in-aaL

Rani Radha-DAT equally sing-PAS-3FS

Rani sang as equally as Radha

One can notice similarity expressed in the following comparative phrases. In these constructions *poonRa* is used as standard marker.

(24) matipoonRamukam

moon like face

'moon like face'

(25) taamaraipoonRamukam

lotus like face

'lotus like face'

Comparison of equality can be studied elaborately, but such an elaborative method is not adopted here.

5. COMPARATIVE STUDY WITH THE EXISTING MACHINE TRANSLATION

Comparative construction Sentences in Tamil are translated to English using the existing Translation Systems. The table given below shows the translation rules for the Comparative construction Sentences between Tamil and English and also the results of the existing Translation Systems.

Table 1. Comparative Study with the existing Machine Translation

Rule No	**Tamil sentence & Rule**	**English Translation & Rule**	**Google Correct [1] / Wrong [0]**	**Bing - Microsoft Correct [1] / Wrong [0]**	**Systran Correct [1] / Wrong [0]**
1.	ராணிராதாவைவிடஅழகானவள் **(NP_{COM} + NP-ai + vita/ kaaTTilum + ADJ-PN)**	Rani is **more** beautiful than Radha	Rani is **more** beautiful than	Rani is **more** beautiful than	Queen is **better** than Radh

		(NP_{COM} + BE + more + ADJ + than + NP_{SOC})	Radha [1]	Radha [1]	a [0]
2.	ராணிஅவர்கள்எல்லோரையும்விடஅழகானவள் **(NP+ NP-ai + viTa/kaaTTilum + ADJ-PN)**	Rani is the **most** beautiful among all **(NP_{COM} + BE + most + ADJ + among all)** [0]	Rani is **more** beautiful than all of them [0]	Rani is **more** beautiful than all of them. [0]	The Queen is **more** beautiful than them all [0]
3.	ராணிராதாவைவிடஉயரமானவள் **(NP_{COM} + NP-ai + vita/kaaTTilum + ADJ-PN)**	Rani is **taller** than Radha **(NP_{COM} + BE + ADJ-er + than + NP_{SOC})** [1]	Rani is **taller** than Radha [1]	Rani is **taller** than Radha [1]	Queen is **taller** than Radha [1]
4.	ராணிஅவர்கள்எல்லோரையும்விடஉயரமானவள் **(NP+ NP-ai + viTa/kaaTTilum + ADJ-PN)**	Rani is the **tallest** among all **(NP_{COM} + BE + ADJ-est + among all)** [0]	Rani is **taller** than all of them [0]	The Queen is **taller** than all of them [0]	The Queen is **taller** than them all [0]

5.	ராணிராதையைக்காட்டிலும்அழகாகஇருக்கிறாள் (**NP_{com} + NP-ai + vita/ kaaTTilum + N-ADVP iru-TEN-PNG**)	Rani is **more** beautiful than Radha (**NP_{com} + BE + more + ADJ + than+ NP_{soc}**)	Rani is **more** beautiful than Radha [1]	Rani is **more** beautiful than Radha [1]	Queen is **more** beautiful than queen [1]
6.	ராணிராதாவைவிடவேகமாகஓடினாள் **(NP_{com} + NP-ai +viTa/ kaaTTilum + ADV +V-TEN-PNG)**	Rani ran **faster** than Radha (**NP_{com} + V-TEN+ ADV-er +than + NP_{soc}**)	Rani ran **faster** than Radha [1]	Rani ran **faster** than Radha [1]	**Queen** ran **faster** than Radha [1]
7.	ராணிராதாவைப்போல்அழகானவள் **(NP + NP-ai + pool + ADJ-PN)**	Rani is **as beautiful as**Radha (**NP + BE + as-ADJ-as + NP_{soc}**)	Beautiful **like** Rani Radha [0]	Rani is **as beautiful as**Radha [1]	She is **as beautiful as** queen Radha [1]
8.	ராணிராதாவைப்போல்வேகமாகநடக்கிறாள்	Rani walks **as fast as**Radh	Rani walks **as fast**	Rani walks **as fast**	Queen **acts swiftl**

	(NP_{COM} + NP-ai + poola + ADV + V-TEN-PNG)	a (**NP_{COM} + V-TEN + as-ADV-as +NP_{soc}**)	**as** Radha [1]	**as**Radha [1]	**y as**Radha [0]
9.	ராணிராதாவைப்போல்இருக்கிறாள் **(NP_{COM} + NP-ai + poola + iru-TEN-PNG)**	Rani looks/ resembles Radha. **(NP_{COM} + Look/ resemble-TEN + NP_{soc})**	**Looks like** Rani Radha [0]	Rani **looks like** Radha [0]	She is like Rani Raddha [0]
10	ராணிராதையேதான்	'Rani is exactly like Radha'	Rani is Radha [0]	It's Rani Radha [0]	Rani Raddha [0]
	Correct output percentage		60% [6/10]	60% [6/10]	40% [4/10]

6. CONCLUSION

The Comparison of equality and inequality needs elaborate study. Only certain important aspects of comparison are studied here from the point of view of machine translation. We try to map comparative constructions of equality and inequality in Tamil and English by positing mapping rules. The Comparative construction Sentences in Tamil are translated to English using the existing Translation Systems and the results indicate that the existing Translation system needs to concentrate on the specific pattern of comparative construction Sentences between Tamil and English. The Proposed Mapping rules will enhance the results of the existing Machine Translation systems.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

REFERENCES

1) Beck, S., Oda, T., &Sugisaki, K. (2004). Parametric variation in the semantics of comparison: Japanese vs. English. Journal of East Asian Linguistics, 13(4), 289-344.
2) Bisang, W. (1992). The 'more... the more' construction in Chinese. In Proceedings of the 18th Annual Meeting of the Berkeley Linguistics Society.
3) Bresnan, J. (1973). Syntax of the comparative clause construction in English. Linguistic Inquiry, 4(3), 275-343.
4) Büring, D. (2007). Cross-polar nomalies. Linguistics and Philosophy, 30(3), 353-367.
5) Chomsky, N. (1977). On WH-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), Formal syntax (pp. 71-132). Academic Press.
6) Croft, W. (2003). Typology and Universals. Cambridge University Press.
7) Culicover, P. W., &Jackendoff, R. (2005). Simpler syntax. Oxford University Press.
8) Dixon R.M.W. (2008). Comparative constructions: A cross linguistic typology, Studies in Languages 32,4:787-817.
9) Dixon R.M.W. (2012). Basic Linguistic Theory, Volume 3, Further Grammatical Topics. Oxford: Oxford University Press.
10) Dixon R.M.W.(2005). Comparative constructions in English, StudiaAnglicaPosnaniensia.
11) Dixon, R. M. W. (2008). Comparative constructions in English. In R. M. W. Dixon (Ed.), Adjective classes: A cross-linguistic typology (pp. 1-45). Oxford University Press.
12) Fults, S., & Phillips, C. (2004). The effects of syntactic complexity on comprehension: Comparatives and relative clauses. Journal of Memory and Language, 51(1), 43-61.
13) Heim, I. (1985). Notes on comparatives and related matters. Ms., University of Texas, Austin.

14) Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. In Proceedings of the National Conference on Artificial Intelligence (Vol. 21, No. 2, p. 1331). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press.
15) Kennedy, C. (1999). Projecting the adjective: The syntax and semantics of gradability and comparison. Routledge.
16) Kennedy, C. (2005). Comparative deletion and optimality in syntax. Natural Language and Linguistic Theory, 23(1), 137-205.
17) Kennedy, C., & Merchant, J. (2000). Attributive comparative deletion. Natural Language & Linguistic Theory, 18(1), 89-146.
18) Klein, E. (1980). A semantics for positive and comparative adjectives. Linguistics and Philosophy, 4(1), 1-45.
19) McCawley, J. D. (1988). The comparative conditional construction in English, German, and Chinese. In Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society (pp. 176-187).
20) Rajendran S.(1976). Comparative constructions in Tamil, Bulletin of the Deccan College Research Institute, volume xxxvi nos 1-4.
21) Wierzbicka, A. (1991). Cross-cultural Pragmatics: The Semantics of Human Interaction. Mouton de Gruyter.
22) Yvonne Treis, (2018).Comparative constructions: An Introduction. Linguistic Discovery volume 16, issue 1.