

Study Morphological Complexity of Non-Related Languages to Build a Universal Morphological Model for Machine Translation

MAST Goonatilleke
Faculty of Computing
General Sir John Kotelawala
Defence University
Ratmalana, Sri Lanka

B Hettige
Faculty of Computing
General Sir John Kotelawala
Defence University
Ratmalana, Sri Lanka

AMRR Bandara
Faculty of Applied Sciences
University of Sri
Jayewardenepura
Nugegoda, Sri Lanka

Faculty of Applied Sciences
University of Sri
Jayewardenepura
Nugegoda, Sri Lanka

Abstract: Language is the primary medium used by human beings to convey their thoughts, ideas, feelings, and information. There are many languages in the world each with its own unique complexities. Therefore, language barrier among people is rapidly increased, and the language complexity has become an unsolved problem in linguistics. However, this language complexity can be reduced using the technology named “Machine Translation” which is one of the areas in Natural Language Processing. It is a computer-aided machine that can translate one language into another language without any intervention of humans. Although, there are many machine translation systems in the world to translate different language pairs, this process still remained as a complex process due to various reasons. The main problem behind this situation is there is no any universal language interlingua model for machine translation to represent and model language information of a particular language that could use for any translation. As the solution, it is recommended to design and develop a universal language model that could facilitate machine translation. As the first step of this research, developing a universal morphological model for English language is proposed that can be used to generate appropriate target morphological model for any language. The main aim of this article is to study and compare the morphological differences and complexities of two non-related language pairs namely English and Sinhala to design and develop this universal morphological model. Therefore, these two selected languages were deeply studied and analyzed in morphological point of view and many promising differences have been identified in respect to grammar structures, parts-of-speech, inflectional categories and etc.

Keywords: machine translation; morphology; language complexity; English language; Sinhala language

1. INTRODUCTION

Languages are considered as the most flexible tool of human adaptation. More specifically, natural language is a structured system, mainly used for communication. It is also named as ordinary language that can exist in various forms such as speech, sign and written. Also, according to linguistics, a natural language is any language which has evolved naturally by humans through use and repetition without conscious planning or premeditation [1]. There are 5000 – 7000 unique languages that belong to different countries and cultures in the world [2][3]. In addition, the English term “language” is derived from the Latin word that means tongue and the scientific study of language is named as linguistics. The history of languages is stepped back about 1.5 million years ago, specifically to the period which Homo erectus lived. Nevertheless, the formal study of languages has been started during 5th century in B.C in India with Panini who was a famous grammarian [4].

The main problem behind languages is language complexity as these languages in the world are unique and complex. Language complexity or linguistic complexity refers to different number of attributes that make a language intricate or challenging to learn and use. Moreover, it can be examined from multiple perspectives such as phonological complexity, morphological complexity, syntactic complexity, semantic

complexity, pragmatic complexity, lexical complexity, orthographic complexity and sociolinguistic complexity [5][6]. In addition, language complexity has become a topic in linguistics that has gained the lowest attention comparing other areas in linguistics. At present, using the power of Artificial Intelligence, more specifically using Natural Language Processing technology this language complexity can be minimized and reduced. Therefore, as the solution, machine translation systems can be used to solve this problem effectively and helps to minimize and reduce the language barrier between different languages in the world.

Machine translation is one of the branches in Natural Language Processing and one of the sub-areas in Artificial Intelligence which can be explained as an automatic process to translate one natural language into another natural language without any human involvement [7]. Also, machine translation has a great history that is stepped back into 1940s and since many machine translation systems have been developed to translate related or non-related language pairs by following various machine translation approaches such as rule-based approach, corpus-based approach, hybrid approach and neural approach. Although, there are many machine translation systems that have been developed through different machine translation approaches, still the machine translation process is remained as a very complex mechanism due to several challenges. In addition, neural machine

translation approach which is the latest machine translation approach also has different challenges in terms of technical, ethical and practical issues. Furthermore, these issues are related to data collection and preparation, user privacy, data credibility and representation, cultural representation, explainability and validity [8]. The main reason behind such problems is that there is no any universal language interlingua model for machine translation to represent and model language information that could use for machine translation. As the solution, it is recommended to design and develop a universal language model that could facilitate machine translation. As the first step of this research, developing a universal morphological model for English language is proposed that can be used to generate appropriate target morphological model for any language. The aim of this study is to explore the morphology and morphological complexity of non-related language pairs to design and develop a universal morphological model for English language. Therefore, to achieve objectives of this research, we have chosen two non-related language pairs that are commonly used in Sri Lanka such as English language and Sinhala language.

The rest of the paper is structured as follows. Chapter 2 explains the methodology of the research and chapter 3 gives a detail explanation about language complexity by addressing the different stages of the machine translation pyramid. Chapter 4 has been reserved to give a detailed overview about morphology of languages. Moreover, Chapter 5 aims to discuss about the English language morphology while Chapter 6 discusses the Sinhala language morphology. Then, chapter 7 gives a deep comparison between these two languages and finally, the paper concludes with highlighting the future works of this research.

2. METHODOLOGY

To achieve the aim with its respective objectives successfully, the following steps can be performed. In this research, the main concern is to model the existing interlingua to represent language knowledge accurately by designing and developing a markup language for knowledge representation in machine translation. Whilst, for this proposed research and the final development, design science research methodology will be applied. The figure 1 shows the design science research methodology process model and the methodology of the proposed research and development will be described with respect to this design science research methodology process model.

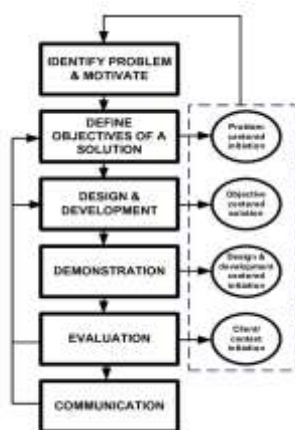


Figure 1. Design Science Research Methodology Process Model.

1. At first, the real-world problem and the proper motivation have been discovered which exists in the machine translation. It is, there is no proper mechanism or methodology to store language knowledge that can be read and identify easily for machine translation. However, when there is a proper mechanism or methodology exists, can do analysis to the source language text and easily generate the appropriate target language text.
2. As the second step, an effective solution should be proposed with a clear and suitable set of objectives. According to the proposed research, the solution is to design and develop a knowledge representation model that stores language knowledge for machine translation. As the initial step of this master research, it is focused to design and develop a universal morphological model to represent and store morphological details of a word in the language after morphologically analyzing the source language word. Therefore, as the objectives, this paper focuses to study the language complexity with respect to the morphological complexity, morphology of languages, morphology of English language and morphology of Sinhala language. As the final objective, the paper gives a detail comparison of these studied languages.
3. Thirdly, the design and development stage will be addressed to design and develop a universal model that could be read and used by anyone. Accordingly, in this stage, it is very important to identify input, output and the process. The input of this system will be source language words, phrases or sentences and the output will be a UNL model that will be generated by analyzing the input. According to this paper's objectives, as the further work, the authors will design a model for words to represent morphological elements and information.
4. The next stage of the methodology is demonstration. Accordingly, the working mechanism and the principle of the developed model will be demonstrated with the aid of a machine translation system. Accordingly, the proposed machine translation system will be designed to translate English language words, phrases or sentences into Sinhala language through the usage of the developed universal language knowledge intermediary representation model.
5. After developing the universal language knowledge intermediary representation model and the proposed machine translation system, the system can be tested and evaluated by human evaluators including experts in the languages and natural language processing area.
6. Finally, the developed system can be deployed for the usage and also can distribute with other people to modify the existing machine translation system to translate different language pairs mainly using the developed universal readable language knowledge intermediary representation model.

Despite the main methodology of the master research, a small methodology for this study can be explained using the Figure 2.



Figure 2. Methodology Process for the Proposed Study.

3. COMPLEXITY OF LANGUAGES

Language complexity is one of the major problems in modern linguistics that can be divided into four sub-topics such as phonological complexity, morphological complexity, syntactic complexity and semantic complexity [9][10][11]. Moreover, language complexity can be dependent upon variety of elements and inter-relational structure and therefore, again the language complexity can be divided into four areas such as syntagmatic complexity, paradigmatic complexity, organizational complexity and hierarchic complexity. Syntagmatic complexity refers to the number of parts such as word length while paradigmatic complexity refers to the types of parts such as number of distinctions in a grammatical category. Thirdly, organizational complexity refers to the order of components and recursion or lexical-semantic hierarchies are examples for hierarchic complexity [12]. Besides, language complexity can be further clearly understand using the levels of the machine translation pyramid that have been divided into four layers such as lexical or word level, morphology, syntax, and semantics. The complexity of each layer in the machine translation pyramid has been described below. Figure 3 shows the machine translation pyramid.

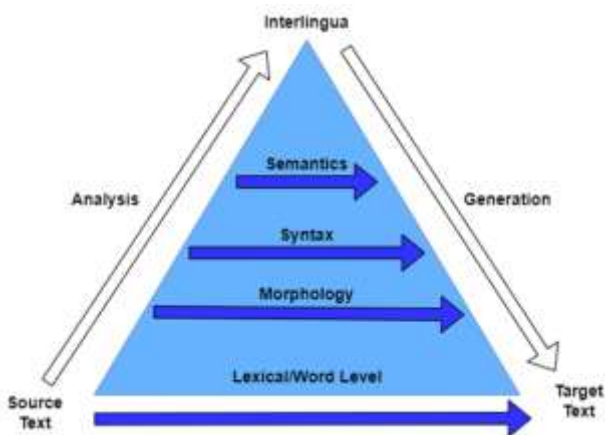


Figure 3. Machine Translation Pyramid.

3.1 Lexical Complexity

The bottom layer of the pyramid comprised of words or lexical components. In this level, lexical ambiguity mainly causes lexical complexity. Lexical ambiguity occurs when a single word contains two or more meanings and lexical ambiguity can be classified into two types such as polysemy and homonymy. Polysemy refers to a single word that can be either noun or verb with many related meanings while homonymy refers to words which are either nouns or verbs that are spelt same in speech but their meanings are different [13].

3.2 Morphological Complexity

Morphology refers to the scientific study of the internal structure or construction of words [14]. Accordingly, in a language, there are a massive number of words, and these words are comprised of different forms. These different forms have different meanings, and this will lead to have complexity. This complexity is referred to as morphological complexity [15].

3.3 Syntactic Complexity

The word “syntax” originated from Ancient Greek roots that stands for coordination or ordering together. In linguistics, syntax is one of the main branches that refers to the protocols which governs to form phrases, clauses and sentences using combinations of words. Accordingly, syntax mainly studies the structure and formation of sentences by explaining how these words are arranged to form a correct sentence. Whilst the topic “syntax” covers different topics such as grammar rules and word order. Therefore, the main aim of syntax is to form proper grammatical correct phrases and sentences [16]. The complexity of syntax in a language is known as syntagmatic complexity that is mainly concerned with words, sentences and discourse. Syntagmatic complexity mainly arises due to two main reasons such as grammatical complexity and syntactic ambiguity. A brief description about each syntagmatic complexity is given below.

3.3.1 Grammatical Complexity

Grammatical complexity is one of the challenges meets when learning languages that is mainly based on set of grammatical rules. Basically, every sentence in any language mainly contains three elements such as subject (S), object (O) and verb (V). The order of these three elements will be different from one language to another language and this will result to have grammatical complexity in different languages. The English language has SVO structure while Tamil and Sinhala languages have SOV structure. Accordingly, Tamil and Sinhala languages are considered as related language pairs while Sinhala and English languages are considered as non-related language pairs. Moreover, number of cases and tenses also lead to have grammatical complexity in languages. [17].

3.3.2 Syntactic Ambiguity

Syntactic ambiguity is also known as structural ambiguity, grammatical ambiguity, amphiboly or amphibology when a single sentence depicts more than one meaning. This situation mainly arises due to the poor word choice [18].

3.4 Semantic Complexity

The fourth layer of the machine translation pyramid has been reserved for the semantics that plays a valuable part in our daily communication, language learning and understanding. Semantics is the study of the meanings of words and sentences in a language [19]. Furthermore, semantics in a

language can be broadly divided into three subcategories such as formal semantics, conceptual semantics, and lexical semantics. Formal semantics refers to the study of grammatical meanings of words and sentences in a language. Secondly, conceptual semantics refers to the study of words at their core while lexical semantics is the study of meaning of words. There are different factors which affect to increase the semantic complexity of a language. Among them, the nature of the vocabulary list took an important place. The vocabulary list may be comprised of simple and complex vocabulary terms. Accordingly, if a text contains more complex vocabularies the semantic complexity will be very high. More-over, syntactic and surface features such as sentence length, letter count, syllable count, sentence structure are also involved to increase the semantic complexity.

4. MORPHOLOGY OF LANGUAGES

The research tradition and the theory of morphology is very much different from the other areas in linguistics and the history of morphological analysis dates back to the period of the ancient Indian linguist named Panini. Besides, Greco-Roman grammatical tradition and studies in Arabic morphology were also considered as pioneers that have engaged in morphological analysis. Whilst the formal term “morphology” was coined in 1859 by August Schleicher which is a Greek term that is makeup of two words such as “morph” and “ology” that means form and the study of something respectively. However, the modern study of morphological analysis was started during the early 1970s [20].

The term “morphology” is the scientific study of words considering the theories of how they have been formed as well as their relationships with other words in the same language. Furthermore, morphology analyzes the structure and parts of the words including stems, root words, prefixes and suffixes. In addition, these words have been built up from morphemes that can be considered as smaller meaning-bearing units or minimal meaning-bearing units in a language. [21][22]. Morphemes can be broadly classified into two groups as free morphemes and bound morphemes. Moreover, free morphemes can be further classified into two parts as lexical morphemes and grammatical morphemes. Secondly, bound morphemes are also can be classified into two groups such as inflectional morphemes and derivational morphemes [23][24].

5. MORPHOLOGY OF ENGLISH LANGUAGE

English is a West Germanic language that is mainly falls under Indo-European language family that is originated from the Anglo-Frisian and Old Saxon dialects brought to Britain. It is the most popular and widely spoken language in the world. In addition, fifty-nine (59) states are using English as their official language and there are more than three (3) billion speakers in all over the world. Furthermore, English language is considered as the international communication language. The modern English alphabet is comprised of twenty-six (26) letters including five (5) vowels. Moreover, this language has eight (8) part of speech such as noun, verb, pronoun, adjective, adverb, prepositions, and conjunctions. English language contains few morphological rules, and the changes of these rules are made through the involvement of suffix and apex into the root word [25][26]. Table 1 shows the common suffixes available for the English language.

Table 1. Common Suffixes for English Language

Affix	Grammatical Category	Mark	Part of Speech
-s	Number	Plural	Nouns
-'s/'s	Case	Genitive	Nouns, Noun phrases, Pronouns
-self	Case	reflexive	Pronouns
-ing	Aspect	progressive	Verbs
-en/ -ed	Aspect	perfect non- progressive	Verbs
-ed	Tense	past(simple)	Verbs
-s	Person, Number, Aspect, Tense	3 rd person singular present	Verbs
-er	Degree of Comparison	comparative	Adjectives
-est	Degree of Comparison	superlative	Adjectives

Morphology in English language can be further classified into four groups such as noun morphology, adjective morphology, verb morphology and adverb morphology and a brief description about each morphology type is given below.

5.1 English Noun Morphology

A “noun” in a language can be described as a word that is used to name something such as a person, a place, an object and an idea. Accordingly, nouns can play different roles such as subject, direct object, indirect object and etc. in a sentence. In the morphology point of view, English noun can be considered as the main morphological category that is mainly participating in inflection and derivation. Inflection refers to the modification of a word to express different grammatical categories such as tense, case, voice, aspect, person, number gender and mood. Accordingly, nouns in the English language participate in number, gender and case inflections. Table 2 shows some morphological rules for English noun inflections.

Table 2. Morphological Rules for English Noun Inflections

Grammar	Morphology			Example
	Base form	Add	Remove	
Singular	Noun	-	-	Student
Plural	Noun	s	-	Students
Plural	Noun	es	-	Dishes
Plural	Noun	ies	y	Ladies
Plural	Noun	ves	f	Knives
Singular Possessive	Noun	's	-	Book's
Plural Possessive	Noun	s'	-	Girls'
Singular	Verb	er	-	Reader
Plural	Verb	ers	-	Readers
Singular	Verb	ment	-	Achievement
Plural	Verb	ments	-	Achievements

Furthermore, considering the English noun inflection, a noun can be divided into two categories such as regular noun or an

irregular noun. Table 3 shows some regular and irregular noun forms.

Table 3. Regular and Irregular Noun Forms

Inflexion form	Regular	Irregular
Singular	Student	corpus
Plural	Students	corpora
Singular Possessive	Student's	corpus's
Plural Possessive	Students'	corpora's

5.2 English Verb Morphology

Verbs in a language are words that are used to describe an action or something that happens. English language contains verbs and they have few morphological forms such as simple present tense, third-person singular, simple past tense, present participle, and past participle. Moreover, considering verbs, conjugated English verbs can be categorized into regular and irregular forms. Table 4 shows the regular and irregular English verb forms while table 5 shows morphological rules for verb conjugation.

Table 4. Regular and Irregular Verb Forms

Inflexion form	Regular	Irregular
Infinitive	Cook	Eat
Simple present	Cooks	Eates
Present Participle	Cooking	Eating
Past	Cooked	Ate
past Participle	Cooked	Eaten

Table 5. Morphological Forms for Verb Conjunctions

Grammar	Morphology			Example
	Base form	Add	Remove	
Infinitive	Verb	-	-	Cook
Simple present	Verb	s	-	Cooks
Present Participle	Verb	ing	-	Cooking
Past	Verb	ed	-	Cooked
past Participle	Verb	ed	-	Cooked

5.3 English Adjective Morphology

Adjectives in a language are words that are used to describe quantities, qualities and states of a noun in other words. Moreover, adjectives are used to modify the noun and also can act as a complement to linking verbs or verbs. Not only that but also, considering the degrees of comparison of an adjective, there are three forms are available such as absolute adjectives, comparative adjectives and superlative adjectives. Table 6 shows the morphological rules for English adjective.

Table 6. Morphological Rules for English Adjective

Grammar	Morphology			Example
	Base form	Add	Remove	
(Positive) Adjective	Base	-	-	Bad
(Positive) Adjective	Noun Base	ish	-	Boyish
(Positive) Adjective	Noun Base	ful	-	Useful
(Positive) Adjective	Noun Base	less	-	Shameless
(Positive) Adjective	Noun Base	en	-	Golden
(Positive) Adjective	Noun Base	active	-	Talkative
(Positive) Adjective	Noun Base	able	-	Moveable
(Comparative) Adjective	Adjective	er	-	Colder
(Comparative) Adjective	Adjective	r	-	Larger
(Comparative) Adjective	Adjective	ier	y	Dirtier
(Superlative) Adjective	Adjective	est	-	Cleverest
(Superlative) Adjective	Adjective	st	-	Simplest
(Superlative) Adjective	Adjective	iest	y	Dirtiest

5.4 English Adverb Morphology

Adverbs are words that are used to describe a verb, an adjective, another adverb or sometimes the whole sentence. Generally, adverbs are ended with the suffix such as 'ly'. Furthermore, adverbs can be used to give a full description of "how something happens", using words such as when, how, where, in what, way and "to what extent". Table 7 shows some relationships between verb and adverb.

Table 7. Relationships between Verb and Adverb

Verb	Adverb	Example
When?	early	He always arrives early.
How?	carefully	He drives carefully.
Where?	everywhere	They go everywhere together.
In what way?	slowly	He eats slowly.
To what extent?	slowly	It is slowly hot

6. MORPHOLOGY OF SINHALA LANGUAGE

Sinhala or Sinhalese is an Indo-Aryan language. It is mainly spoken by Sinhala people in Sri Lanka who belong to the largest ethnic group of the island, approximately sixteen million. Moreover, Sinhala is one of the official languages in Sri Lanka and it is written using the Sinhala script that is mainly derived from Brahmic scripts and Grantha script. The

history of Sinhala language dates to 3rd century BCE and the evolution and development of the language has been divided into four epochs such as Sinhala Prakrit, Proto-Sinhala, Medieval Sinhala and Modern Sinhala [27] [28] [29].

Sinhala (සිංහල) language contains its own alphabet with few versions, including the Unicode version (18 vowels and 45 consonants) and alphabet of the “Sedath sagara” (10 vowels and 20 consonants). Nevertheless, Sinhala mix alphabet comprised of 18 vowels and 36 consonants. More importantly, Sinhala contains only four parts-of-speech (භේද) Namely Naama (Noun), Kriya (Verb), Nipath and Upsarga (indeclinable particle). This four parts-of-speech embrace the eight parts-of-speech specified in English. Figure 4 shows the part of speech mapping between English and Sinhala languages [30] [31].

As one of the rich and old languages in the world, Sinhala language participates in inflection, derivation, and conjugation of nouns and verbs. Brief description about noun and verb morphology of the Sinhala language is given below.

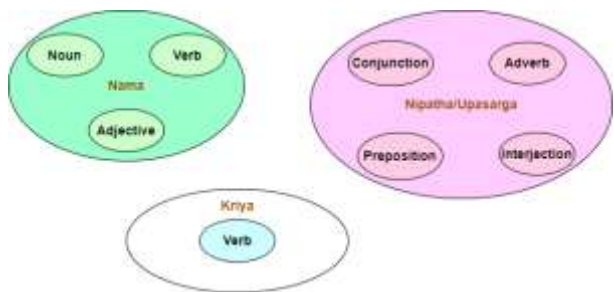


Figure 4. Parts of Speech Mapping between Sinhala and English Languages.

6.1 Sinhala Noun Morphology

Nouns in Sinhala language are named as ‘Nama’. Compared with the English language, these Sinhala language nouns represent not only nouns but also pronouns and adjectives in the English language. Further, Sinhala nouns show gender, number, person and case base inflection. Not only that but also, in morphological point of view, Sinhala nouns can be broadly classified into three groups such as simple nouns, complex nouns and compound nouns. Besides, they also have nine cases namely “nominative, accusative, instrumental, auxiliary, dative, ablative, genitive, locative, and vocative”. According to the above grammar, more than 27 forms of nouns can be generated by inflecting a single root word (base form of a Sinhala noun). Table 8 shows some Sinhala noun inflection forms considering an example.

Table 8. Sinhala Noun Inflection forms for Base Word “ළමයා” (Child)

Case	Singular direct	Singular indirect	Plural
Nominative	ළමයා	ළමයෙක්	ළමයි
Accusative	ළමයා	ළමයෙකු	ළමුන්

Instrumental	ළමයා විසින්	ළමයෙක් විසින්	ළමයි විසින්
Auxiliary	ළමයාගෙන්	ළමයෙක්ගෙන්	ළමයින්ගෙන්
Dative	ළමයාට	ළමයෙකුට	ළමුන්ට
Ablative	ළමයාගෙන්	ළමයෙකුගෙන්	ළමුන්ගෙන්
Genitive	ළමයාගේ	ළමයෙකුගේ	ළමුන්ගේ
Locative	ළමයා කෙරෙහි	ළමයෙකු කෙරෙහි	ළමුන් කෙරෙහි
Vocative	ළමයා		ළමුනේ

6.2 Sinhala Verb Morphology

Sinhala verb is the action word that is available in the Sinhala sentence that can be divided into two groups such as transitive and intransitive. Moreover, Sinhala verbs are inflected from four categories such as voice, mood, tense, number, and person. Not only that but also, when compared with the English verb, the Sinhala verb takes only three tenses such as the present, past, and future. Nevertheless, Sinhala verb shows more inflexion forms (verb conjugation) than the Sinhala noun. Therefore, it shows more than 36 inflexion forms, including active, passive, optative mood, imperative mood, and conditional mood. Table 9 shows some inflexion forms for the Sinhala verb “කරනවා” (doing).

Table 9. Sinhala Verb Inflection forms for the Word “කරනවා” (Doing)

කාලය	උත්තම ඒක	උත්තම බහු	මධ්‍යම ඒක	මධ්‍යම බහු	ප්‍රථම ඒක	ප්‍රථම බහු
කතෘ වර්තමාන	කරමි	කරමු	කරහි	කරහු	කරයි	කරති
කතෘ අනාගත	කරන්නේමි	කරන්නේමු	කරන්නේහි	කරන්නේහු	කරන්නේයි	කරන්නේති
කතෘ කාරක අතීත	කළෙමි	කළෙමු	කළෙහි	කළෙහු	කර	කෙරු
කර්ම කාර වර්තමාන	කෙරෙමි	කෙරෙමු	කෙරෙහි	කෙරෙහු	කෙරෙයි	කෙරෙති
කර්ම කාර අතීත	කෙරෙන්නේමි	කෙරෙන්නේමු	කෙරෙන්නේහි	කෙරෙන්නේහු	කෙරෙන්නේයි	කෙරෙන්නේති

අනාගත	මි	මු	හි	හු		න්නෝ
කර්මකාරක අතීත	කෙරෙහි	කෙරෙහිමු	කෙරෙහි	කෙරෙහි	කෙරෙහි	කෙරෙහි

7. ENGLISH LANGUAGE VS. SINHALA LANGUAGE

English language and the Sinhala language are considered as non-related language pairs. This is mainly due to the formation of the grammar order such as the order of subject (S), verb (V) and the object (O). The English language sentences are mainly built up following the SVO structure while Sinhala language sentences are mainly built up following the SOV structure. Besides, many of the differences between these two languages can be examined considering various factors. A comparison between these two language has been shown in the Table 10.

Table 10. Comparison between English Language and Sinhala Language

Category	English Language	Sinhala Language
Alphabet	26 Letters	54 Letters
Number of vowels	5	18
Number of consonants	21	36
Parts of Speech	8 Noun, Verb, Pronoun, Adjective, Adverb, Prepositions, Conjunctions	4 Naama (Noun), Kriya (Verb), Nipath, Upsarga (indeclinable particle)
Grammar Structure	Subject-Verb-Object (SVO)	Subject-Object-Verb (SOV)
Types of Morphology	Inflection and Derivation	Inflection and Derivation
Inflectional Categories	8 Tense, Case, Voice, Aspect, Person, Number, Gender, Mood	4 Case, Person, Number, Gender
Types of Nouns	Simple, Complex, Compound	Simple, Complex, Compound
Cases of Nouns	7 Nominative, Vocative, Accusative, Dative, Genitive, Instrumental, Locative	9 Nominative, Vocative, Accusative, Dative, Genitive, Instrumental, Locative, Auxiliary, Ablative
Noun Forms	8	27
Tenses	12 Present Simple tense, Present Continuous tense,	3 Simple Present tense, Simple Past tense, Future tense

	Present Perfect tense, Present Perfect Continuous tense, Past Simple tense, Past Continuous tense, Past Perfect tense, Past Perfect Continuous tense, Future Simple tense, Future Continuous tense, Future Perfect tense, Future Perfect Continuous tense	
Language Complexity	Medium	High

8. CONCLUSION AND FUTURE WORKS

The paper mainly discussed about the morphological difference between two non-related language pairs such as Sinhala Language and English language. The objective of selecting these languages is that to convince the huge difference among these language pairs as these languages are fell under two different language roots of families. English language is fell under the Indo-European language/west Germanic language family while Sinhala is fell under the Indo-Aryan language family. In there, it was clearly identified many differences between these two languages morphological point of view. Nevertheless, the study has been done as the initial step to design and develop a universal morphological model that can be used for any language to perform machine translation very accurately. Moreover, the future work of this research aims to design and develop a universal morphological model to help machine translation and also to achieve the main aim of the master research that is designing a universal model for machine translation.

9. REFERENCES

- [1] P. Ramat, "John Lyons, Natural language and universal grammar. Essays in linguistic theory, volume 1. Cambridge: Cambridge University Press, 1991. Pp. xv + 290.," J. Linguist., vol. 28, no. 2, pp. 554–558, 1992, doi: 10.1017/s0022226700015474.
- [2] T. Kamusella, "The History of the Normative Opposition of 'Language versus Dialect': From Its Graeco-Latin Origin to Central Europe's Ethnolinguistic Nation-States," Colloq. Humanist., vol. 5, no. 5, pp. 164–188, 2016, doi: 10.11649/ch.2016.011.
- [3] M. Miestamo, K. Sinnemäki, F. Karlsson, and L. Typology, "Reviews," vol. 62, pp. 241–242, 2009.
- [4] P. Kiparsky, "Päinian Linguistics," Concise History of the Language Sciences. pp. 59–65, 1995, doi: 10.1016/b978-0-08-042580-1.50012-x.
- [5] Q. Ma and X. Wang, "What Is Language Complexity?," Macrolinguistics, vol. 7, no. 11, pp. 1–29, 2019, doi: 10.26478/ja2019.7.11.1.
- [6] M. Miestamo, K. Sinnemäki, and F. Karlsson (eds.), Language Complexity: Typology, Contact, Change, vol. 94. Amsterdam: John Benjamins Publishing Company, 2008.

- [7] H. Mercan, Y. Akgün, and M. C. Odacıoğlu, “The Evolution of Machine Translation: A Review Study* The Evolution of Machine Translation: A Review Study,” *Int. J. Lang. Transl. Stud.*, vol. 4, no. 1, pp. 104–116, 2024, [Online]. Available: <https://dergipark.org.tr/tr/pub/lotus>.
- [8] R. Kimera, Y.-S. Kim, and H. Choi, “Advancing AI with Integrity: Ethical Challenges and Solutions in Neural Machine Translation,” 2024, [Online]. Available: <http://arxiv.org/abs/2404.01070%0Ahttp://dx.doi.org/10.22724/LMR.2024.22.1.171>.
- [9] M. Baerman, D. Brown, and G. G. Corbett, “Morphological complexity,” p. 188.
- [10] X. Lin, M. Afzaal, and H. S. Aldayel, “Syntactic complexity in legal translated texts and the use of plain English: a corpus-based study,” *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, pp. 1–9, 2023, doi: 10.1057/s41599-022-01485-x.
- [11] S. Pollard and A. W. Biermann, “A Measure of Semantic Complexity for Natural Language Systems Shannon Pollard * and Alan W. Biermann Department of Computer Science, Duke University,” *Area*, no. 919, pp. 42–46.
- [12] S. S. Mufwene, “The emergence of complexity in language: An evolutionary perspective,” *Underst. Complex Syst.*, pp. 197–218, 2013, doi: 10.1007/978-3-642-32817-6_13.
- [13] J. Maciuszek, “Lexical Access in the Processing of Word Boundary Ambiguity,” *Soc. Psychol. Bull.*, vol. 13, no. 4, pp. 1–11, 2018, doi: 10.32872/spb.v13i4.28690.
- [14] M. Haspelmath and A. D. Sims, *Understanding Morphology*, no. May 2010. 2013.
- [15] V. Brezina and G. Pallotti, “Morphological complexity in written L2 texts,” *Second Lang. Res.*, vol. 35, no. 1, pp. 99–119, 2019, doi: 10.1177/0267658316643125.
- [16] R. Carnap, “Logical syntax of language,” *Logical Syntax of Language*. pp. 1–352, 2014, doi: 10.4324/9781315823010.
- [17] M. J. Owren and G. C. Cardillo, “The relative roles of vowels and consonants in discriminating talker identity versus word meaning,” *J. Acoust. Soc. Am.*, vol. 119, no. 3, pp. 1727–1739, 2006, doi: 10.1121/1.2161431.
- [18] D. D. Oaks, “Structural Ambiguity in English: An Applied Grammatical Inventory,” *Struct. Ambiguity English An Appl. Gramm. Invent.*, pp. 1–539, Jan. 2010, doi: 10.5040/9781474211956.
- [19] “Semantics of Natural Language,” *Semant. Nat. Lang.*, 1972, doi: 10.1007/978-94-010-2557-7.
- [20] M. Aronoff and K. Fudeman, “What is morphology?,” Accessed: Jun. 16, 2024. [Online]. Available: https://books.google.com/books/about/What_is_Morphology.html?id=v-uKEAAQBAJ.
- [21] R. Lieber, *Introducing Morphology*. 2015.
- [22] R. Wiese, “Phonology: Overview,” *Encycl. Lang. Linguist.*, pp. 562–564, 2006, doi: 10.1016/B0-08-044854-2/00035-3.
- [23] A. D. Friederici and W. Singer, “Grounding language processing on basic neurophysiological principles,” *Trends Cogn. Sci.*, vol. 19, no. 6, pp. 329–338, Jun. 2015, doi: 10.1016/j.tics.2015.03.012.
- [24] A. Spencer, “Morphology in Linguistics,” *Int. Encycl. Soc. Behav. Sci.*, pp. 10047–10053, 2001, doi: 10.1016/B0-08-043076-7/02988-0.
- [25] R. Hogg and D. Denison, *A history of the english language*. 2006.
- [26] “Words in English: Latin and Greek Morphology.” <https://www.ruf.rice.edu/~kemmer/Words/classmorph> (accessed Jun. 19, 2024).
- [27] F. Strik Lievers, M. Bolognesi, and B. Winter, “The linguistic dimensions of concrete and abstract concepts: Lexical category, morphological structure, countability, and etymology,” *Cogn. Linguist.*, vol. 32, no. 4, pp. 641–670, Nov. 2021, doi: 10.1515/COG-2021-0007.
- [28] “The Evolution Of The Sinhala Language: An Important Reference - Colombo Telegraph.” <https://www.colombotelegraph.com/index.php/the-evolution-of-the-sinhala-language-an-important-reference/> (accessed Jun. 19, 2024).
- [29] S. Coperahewa, “Origin of Sinhala and Kalinga – Lanka Linguistic Affinities,” no. November, 2015, doi: 10.13140/RG.2.1.2417.2885.
- [30] J. W. R. W. K. Jayaweera, “Buddhism and the Sinhala Writing Tradition,” *Adv. Soc. Sci. Res. J.*, vol. 6, no. 1, pp. 559–569, Feb. 2019, doi: 10.14738/assrj.61.6132.
- [31] “Sinhalese Language, A Comprehensive Grammar of (Gunasekara).pdf.” .