Journal of Computer Science

Original Research Paper

A Context-Free Grammar for Parsing Manipuri Language

Yumnam Nirmal and Utpal Sharma

Department of Computer Science and Engineering, Tezpur University, Tezpur, India

Article history
Received: 23-03-2021
Revised: 08-06-2021
Accepted: 09-07-2021

Corresponding Author:
Yumnam Nirmal
Department of Computer Science and Engineering, Tezpur University, Tezpur, India
Email: ynirmal@tezu.ernet.in

Abstract: Parsing, i.e., identifying the underlying hierarchical structure of natural language expressions is important for several natural language processing applications. In recent times Machine Learning (ML) approaches have been developed for this study for many languages. Most of the effective techniques require an annotated corpus of the language for training and validation. For the Manipuri language of the Tibeto-Burman family, neither such a corpus nor a grammar framework to automatically analyse and represent the structure of sentences exists yet. This study proposes a Context-Free Grammar (CFG) that provides the framework to represent the structure of Manipuri sentences. This paves the way for parsing Manipuri sentences using CFG-based parsers for various applications and to conveniently build a Treebank for developing ML-based parsers for Manipuri. The rules of the proposed CFG are handcrafted after extensive analysis of the structure of Manipuri sentences. The grammar covers simple, compound, complex and compound-complex sentences. For evaluation, we induce an Earley's parser with the proposed CFG and test it over a collection of sentences that covers the possible varieties of structure. A recognition rate of 83.20% achieved in these experiments indicates the effectiveness of the proposed grammar.

Keywords: Context-Free Grammar, Parsing, Manipuri, Tibeto-Burman

Introduction

Syntactic parsing is one of the important aspects of natural language processing that involves the analysis and establishment of syntactic relations among the constituents of a sentence. The result is a parse tree or trees that indicate the syntactic relationship between the constituents. Each constituent plays a distinct role and is hierarchically related to the others. The grammar checking feature in word processors is a common example of syntactic parsing. A sentence cannot be syntactically parsed if it contains grammatical errors or if it is too hard to read (Jurafsky and Martin, 2000).

Automatic parsing of sentences has been successfully done only for a small fraction of all the languages (Makwana and Vegda, 2015; Ammar *et al*., 2016; Han *et al*., 2019; Yang *et al*., 2021). Supervised data-driven approaches require Treebanks and unsupervised data-driven approaches require other resources such as Parts-of-Speech (PoS) tagged corpora. For many languages, neither adequate Treebanks or the other resources required for data-driven parsing, nor any formal computational grammar is available. For Manipuri, Treebanks are nonexistent but a small amount of POS tagged corpus is available. Even though a majority of the work on unsupervised data-driven parsing is based on well-established corpora and covers multiple domains, their performances are inadequate as compared to supervised data-driven approaches (Le and Zuidema, 2015; Han *et al*., 2019; Yang *et al*., 2020).

For Manipuri, data-driven approaches are not attractive at this stage since developing the required resources takes considerable work and time. Hence, to start work for this language we focus on developing a computational constituent grammar or framework by analysing the language structure, based on which syntactic structures of sentences can be represented. Such grammar can be used with known parsing methods. Sentences automatically parsed in this way can be accumulated into a Treebank that in turn will pave the way for adopting data-driven approaches.

DOI: 10.3844/jcssp.2021.855.869

Manipuri Language

Manipuri (also known as Meitei-lon) is a Tibeto-Burman language (Matisoff *et al*., 1996) mainly spoken in the northeastern Indian state of Manipur. It is a scheduled language under the Indian constitution and the lingua franca (trade language) among different communities residing in Manipur. It is also spoken in parts of Assam, Tripura, Bangladesh and Myanmar. It is the only native language with its unique script know as the Meitei Mayek. Manipuri has been currently classified as vulnerable by UNESCO (Blackburn and Opgenort, 2010).

An interesting feature of this language is its highly agglutinative nature. A Manipuri root can take as many as ten suffixes (Singh, 1987). Another feature is its tonal nature where a high number of Manipuri words has a low and a high tone (Sharma, 1987; Bhat and Ningomba, 1997; Chelliah, 2011).

Initial literary works on Manipuri grammar can be seen in the works of Primrose (1995) and Pettigrew (1912). These works provided a useful list of words, phrases and idioms, but are not grammatically exhaustive. A few of the notable Manipuri modern grammar is that of Thoudam (1991), Bhat and Ningomba (1997), Singh (2000) and Chelliah (2011).

Additionally, works on Manipuri to English dictionary are that of Imoba (2004) and Sharma (2006).

A few illustrations of Manipuri sentences with English translations have been given in examples 1, 2 and 3:

Example 1.
ꯃꯅꯤꯄꯨꯔ ꯚꯥꯔꯠꯀꯤ ꯈꯥ-ꯅꯣꯡꯄꯣꯛꯇ ꯂꯩꯕ ꯑꯄꯤꯛꯄ ꯂꯩꯕꯥꯛ ꯑꯃꯅꯤ
mənipuɹ bʰaɹətki kʰa-noŋpoktə ləibə əpikpə ləibak əməni
Manipur is a small state in the northeastern part of Bharat (India).

Example 2.
ꯃꯅꯤꯄꯨꯔ ꯆꯤꯡꯂꯣꯟ ꯃꯄꯟꯅ ꯀꯣꯏꯅ ꯄꯟꯁꯕ ꯂꯝꯗꯝꯅꯤ
mənipuɹ ciŋlon mapənnə koinə pənsabə ləmdəmni
Manipur is a land surrounded by nine layers of mountains.

Example 3.
ꯃꯅꯤꯄꯨꯔ ꯁꯒꯣꯜ-ꯀꯥꯡꯖꯩ ꯃꯁꯅꯒꯤ ꯍꯧꯔꯛꯐꯝꯅꯤ
mənipuɹ səgol-kaŋɟəi məsanəgi həurəkpʰəmni
Manipur is the birthplace of the game Sagol Kangjei (modern polo).

Related Works

Syntactic parsing can be achieved by using handcrafted grammar rules or through data-driven approaches. Manually handcrafting language syntax into grammar rules requires in-depth knowledge of a language and extensive labour. It may be difficult to cover an entire language structure as natural languages are complex but may serve as a starting point for languages where no Treebanks are available (Ababou *et al*., 2017; Korzeniowski and Mazurkiewicz, 2017; Þorsteinsson *et al*., 2019; Sharipbay *et al*., 2019; Kapanadze, 2019; Dorđević and Stojković, 2020).

In data-driven approaches, grammar rules are induced using either statistical or machine learning algorithms. Such work require adequately sized Treebanks or POS annotated corpora. The majority of the state-of-the-art supervised and unsupervised parsers are based on major languages and uses well-established datasets such as the Penn Treebank (Le and Zuidema, 2015; Han *et al*., 2019; Kim *et al*., 2019; Mrini *et al*., 2019; Zhou and Zhao, 2019; Yang *et al*., 2020; Yang and Deng, 2020).

For Manipuri, resources are scarce and computational tools such as a POS tagger are unavailable. In such situations, it is not possible to induce a data-driven parser for the language. Additionally, the grammar of Manipuri so far given by linguists is not computation ready. To overcome this gap, a viable solution would be to manually handcraft a computational grammar such as CFG or Tree-Adjoining Grammar (TAG) and to induce a rule-based parser.

Even though CFG is the most widely used grammar formalism, a CFG specifically designed for a particular language is hardly applicable to another. The reason behind this is the different structures across different languages. As an example, English follows Subject-Verb-Object (SVO) pattern, whereas Manipuri, a Tibeto-Burman language, is verb-final (Singh, 2000) and follows SOV and, when we talk about the phrase level difference, a determiner in English always precede a head noun, whereas, in Manipuri, a determiner always succeed a head noun.

Unlike Indo-Aryan Indian languages such as Hindi and Bangla, Manipuri belongs to the Tibeto-Burman language family and follows a different structure. Similarly, its structure is also different from the Dravidian languages. An example of such a difference has been illustrated in example 4 where the position of Quantifiers (Qtf) in relevance to a head noun (Noun) is shown for both the languages. Quantifiers follow a head noun in Manipuri, whereas it precedes a head noun in Hindi:

Example 4.
(a) Manipuri: ꯑꯉꯥꯡ ꯃꯌꯥꯝ (child many)

əŋaŋ\Noun mjam\Qtf

(b) Hindi: बहुत बच्चे (many children)
bəhut\Qtf bəcce\Noun

Example 5.
(a) Manipuri: ꯑꯊꯣꯠꯄ ꯑꯉꯥꯡꯕ ꯂꯩ (soft red flower)
ətʰotpə\Adj əŋaŋbə\Adj ləi\Noun

or, ꯂꯩ ꯑꯊꯣꯠꯄ ꯑꯉꯥꯡꯕꯦ (flower soft red)
ləi\Noun ətʰotpə\Adj əŋaŋbə\Adj

(b) Hindi: नरम लाल फुल (soft red flower)
nəɹəm\Adj lal\Adj pʰul\Noun

*फुल नरम लाल (flower soft red)
pʰul\Noun nəɹəm\Adj lal\Adj

Another example is that in Manipuri adjectives (Adj) either succeed or precede a head noun, whereas, in the case of Hindi, it always precedes a head noun. The semantics may change if we restrict adjectives to succeed a head noun in Hindi.

As illustrated in example 5(a), the change in position of adjectives about the head noun does not affect the semantic information. The constituents still collectively act as a noun phrase. But, for Hindi, as illustrated in example 5(b), changing the position of adjectives (sentence with * symbol) results in loss of the original semantic information. The adjectives no longer modify the head noun, but rather they become two separate phrases.

We do not follow existing CFGs or rule-based grammars of other languages as syntactic structures are different across languages. Even though some languages may exhibit similar features, it becomes necessary to develop a rule-based grammar for the individual languages. Additionally, for Manipuri, at present, it is not viable to learn syntax rules using data-driven approaches as Treebank is non-existent for this language.

There has been minimal work done on the parsing of the Manipuri language. Sarangthem and Singh (2014) tried to develop an abstract data structure for Manipuri noun phrases. The work focuses on few instances of noun phrases and does not consider constituents that can occur along with a head noun. Instead, it is focused on identifying head nouns by analyzing the affixes that occur along with a noun. Additionally, the authors did not define any concrete formalism for syntactic parsing.

Singh and Sharma (2012) attempted a bottom-up parsing approach for the Manipuri language. This study is preliminary and focuses only on few instances of simple and compound sentences. The authors did not cover complex and compound-complex sentences as well. Additionally, the work lacks the generalisation of Manipuri phrases. Out of the possible constituents of a phrase, only a few constituents and their ordering have been discussed.

Nirmal and Sharma (2018) discussed issues and their possible solutions that may arise during the syntactic parsing of the Manipuri language.

Recent work on Manipuri syntactic parsing is that of Nirmal and Sharma (2019) where the authors present a grammar-driven approach using Earley's algorithm covering simple, compound and complex sentences. The authors presented a general structure for Manipuri sentences, but all the possible instances of the compound and complex sentences have not been covered. Additionally, the work does not present the general structure of compound-complex sentences.

Structure of Manipuri Language

This section presents an analysis of Manipuri phrases and sentences. It is based on existing literature, the basic principles of linguistics and our knowledge of the Manipuri language.

Phrase Structure

In Manipuri, phrases can be categorized into noun phrases, verb phrases and modificative phrases. Modificative phrases can be further divided into adjective and adverb phrases.

Noun Phrase

Sentences in Manipuri are mostly populated with Noun Phrases (NPs) as compared to Verb Phrases (VPs). There can be zero or more NPs in a sentence as arguments of VP. The NPs can be omitted in Manipuri (Chelliah, 2011). The structural constitution of NPs and VPs in a sentence will be discussed in the upcoming sections.

A head noun forms the basic and obligatory component of an NP in Manipuri. It may optionally contain adjectives, quantifiers, Demonstratives (Dmn) and post-positions as its constituents (Madhubala, 1979; Chelliah, 2011; Singh, 2000). If the head noun lacks any of these constituents, it generally occurs with a case suffix. The head noun can be either a noun or pronoun in its pure form without any suffixes, but not verbal nouns. Verbal Nouns (V Noun) cannot act as a

head by itself. Additionally, Pronoun (Pron) or proper nouns can form an NP by itself without any of the optional constituents (Sarangthem and Singh, 2014).

Adjectives

A head noun may be either preceded or succeeded by an adjective or adjectives as its modifier. In theory, any number of adjectives can modify a head noun. Example 6 illustrates such a case where a total of five adjectives modify a head noun in Manipuri.

Example 6.
ꯑꯆꯧꯕ ꯐꯖꯕ ꯑꯉꯥꯡꯕ ꯑꯊꯨꯝꯕ ꯑꯍꯥꯎꯕ ꯍꯩꯅꯧ
əcəubə\Adj pʰəɟəbə\Adj əŋaŋbə\Adj ətʰumbə\Adj əhaubə\Adj həinəu\Noun
A mango that is big, beautiful, red, sweet and tasty.

Quantifiers

If a quantifier is present in an NP, it always succeeds the head noun. But, if an adjective or more is already succeeding the head noun, then the quantifier should succeed the adjectives. In other words, adjectives should always be the immediate neighbor of a head noun.

Example 7.
(a) ꯍꯩꯅꯧ ꯈꯔ
həinəu\Noun kʰəɹə\Qtf
Some mango.

(b) ꯑꯊꯨꯝꯕ ꯑꯍꯥꯎꯕ ꯍꯩꯅꯧ ꯈꯔ
ətʰumbə\Adj əhaubə\Adj həinəu\Noun kʰəɹə\Qtf
Some sweet tasty mango.

(c) ꯍꯩꯅꯧ ꯑꯊꯨꯝꯕ ꯑꯍꯥꯎꯕ ꯈꯔ
həinəu\Noun ətʰumbə\Adj əhaubə\Adj kʰəɹə\Qtf
Some sweet tasty mango.

Demonstratives

Demonstratives, succeed a head noun in Manipuri. Similar to quantifiers, if adjectives are succeeding the head noun, it will succeed the adjectives.

Example 8.
(a) ꯍꯩꯅꯧ ꯑꯗꯨ
həinəu\Noun ədu\Dmn
That mango.

(b) ꯑꯊꯨꯝꯕ ꯑꯍꯥꯎꯕ ꯍꯩꯅꯧ ꯑꯗꯨ
ətʰumbə\Adj əhaubə\Adj həinəu\Noun ədu\Dmn
That sweet tasty mango.

(c) ꯍꯩꯅꯧ ꯑꯊꯨꯝꯕ ꯑꯍꯥꯎꯕ ꯑꯗꯨ
həinəu\Noun ətʰumbə\Adj əhaubə\Adj ədu\Dmn
That sweet tasty mango.

Locative Nouns

Manipuri is a post-positional language. The post-positions are generally directional and indicate temporal dimensions within a syntactic relation (Singh, 2000). It always succeeds a noun or a noun phrase. These directional post-positions occur as locative nouns (Noun-loc) and with case markers suffixed to them. Example 9 illustrates such cases.

Example 9.
(a) ꯎꯄꯨ ꯃꯊꯛꯇ
upu\Noun mətʰktə\Noun-loc
Over the cupboard.

(b) ꯊꯥꯑꯁꯤꯒꯤ ꯃꯅꯨꯡꯗ
[tʰa asigi]NP mənuŋdə\Noun-loc
Within month this.

NPs with Coordinate Conjunctions

An NP, apart from being formed by a head noun along with its constituents, can also be formed with the help of coordinate conjunctions (Conj). In such a case, the coordinate conjunction joins two surrounding nouns or NPs.

Example 10.
(a) ꯇꯣꯝꯕ ꯑꯃꯁꯨꯡ ꯆꯥꯎꯕ
tombə\Noun əməsuŋ\Conj caubə\Noun
Tomba and Chaoba.

(b) ꯑꯛꯅꯕ ꯅꯤꯄꯥ ꯑꯃꯁꯨꯡ ꯑꯄꯤꯛꯄ ꯅꯤꯄꯥ
[əknbə nipa]NP əməsuŋ\Conj [əpikpə nipa]NP
Strong man and tiny man.

Role of Affixes in NP Formation

Manipuri being highly agglutinative, the majority of words occur with affixes and these affixes play an important role in language construction. Suffixes in Manipuri are generally morphemes and adds additional information to a word. Some of these morphemes also help in determining the role of a word in a sentence. It is this agglutinative nature that has allowed some word categories to behave as NP by themselves.

One such case is that of demonstratives where they can occur as the suffix -ꯗꯨ (-du) attached to a head noun

DOI: 10.3844/jcssp.2021.855.869

(Noun-dmn: general nouns with demonstrative suffix). In such a case, adjectives modifying a head noun, if any, will always precede the head noun. Thus, we can restructure examples 8(a) and 8(b) as shown in examples 11(a) and 11(b) respectively:

Example 11.
(a) ꯍꯩꯅꯧꯗꯨ
həinəu-du\Noun-dmn
That mango.

(b) ꯑꯊꯨꯝꯕ ꯑꯍꯥꯎꯕ ꯍꯩꯅꯧꯗꯨ
ətʰumbə\Adj əhaubə\Adj həinəu-du\Noun-dmn
That sweet tasty mango.

We have already stated that pronouns and proper nouns can form an NP by themselves, without any optional constituents. Additionally, these word categories still behave as NP by themselves even after being marked by case suffixes. On the other hand, common nouns generally behave as an NP by themselves only after being marked by a case suffix. The case markers available in Manipuri are nominative -ꯅ (-nə), accusative -ꯄꯨ~ꯕꯨ (-pu~bu), instrumental -ꯅ (-nə), locative -ꯇ~ꯗ (-tə~də), associative -ꯀ~ꯒ (-kə~gə) and genitive -ꯀꯤ~ꯒꯤ (-ki~gi) (Singh, 1987). These case markers appear at the end of a word and indicate the syntactic role of the word in a sentence (Noun-case: General nouns with case suffix).

Example 12.
ꯆꯥꯎꯕꯅ ꯇꯣꯝꯕꯕꯨ ꯐꯨꯢ
caubə-nə\Noun-case tombə-bu\Noun-case pʰui\Verb
Chaoba beat Tomba.

Verb Phrase

Verb phrases are minimal in Manipuri (Chelliah, 2011). They are formed with a single verb (example 13(a)) or a verb preceded by its modifier (example 13(b)). The modifier is in the form of an adverb (Adv) or multiple adverbs (theoretically infinite). Alternatively, the verb may also be preceded by an antecedent in the form of a noun (illustrated in example 13(c)).

Example 13.
(a) ꯆꯥꯔꯤ
caɹi\Verb
Eating.

(b) ꯀꯟꯅ ꯆꯥꯔꯤ
kənnə\Adv caɹi\Verb
Eating seriously.

(c) ꯀꯅꯥꯒꯨꯝꯕ ꯑꯃꯠꯇ ꯂꯥꯛꯄ ꯎꯗꯦ
[kənagumbə əmtə]$_{NP}$ lakpə\VNoun ude\Verb
No one is seen coming.

Structure of Manipuri Sentences

Structurally, Manipuri sentences can be broadly categorized into simple, compound and complex. Additionally, a compound-complex sentence can also be derived out of these three structures and is quite common.

Simple Sentences

Simple Sentences (SSim) in Manipuri consist of at least a verb phrase, optionally preceded by a single or multiple NPs (Chelliah, 2011). These sentences neither accommodate a complex nor a compound construction.

Like the majority of the Tibeto-Burman language, Manipuri follows subject-object-verb word order and has a verb as the final occupant of a sentence (Bhat, 2002). Instead of a verb as the final occupant, a sentence may also have a copula (Cop). In such a case, the verb is replaced by a copula which functions similarly to a verb. Copulas generally appear in the form of a suffix by attaching themselves to a noun.

Depending on whether a sentence consists of a verb or copula, simple sentences can be categorized into verbal and nominal (Singh, 2013).

Nominal sentences generally consist of two different NPs linked by the copula '-ꯅꯤ' (-ni).

Example 14.
ꯅꯤꯄꯤꯃꯆꯥ ꯑꯗꯨ ꯇꯣꯝꯕꯒꯤ ꯃꯆꯥꯅꯤ
[nipiməca ədu]$_{NP}$ [tombə-gi məca]$_{NP}$ -ni\Cop
That girl is Tomba's child.

In example 14, the NPs "ꯅꯤꯄꯤꯃꯆꯥ ꯑꯗꯨ" (nipiməca ədu) and "ꯇꯣꯝꯕꯒꯤ ꯃꯆꯥ" (tombə-gi məca) are connected by the copula '-ꯅꯤ' (-ni).

Verbal sentences constitute a verb or VP as predicate and one or more NP occurring as arguments. A verbal sentence can also constitute only a VP without any NP (Chelliah, 2011).

Example 15.
ꯅꯤꯄꯤꯃꯆꯥ ꯑꯗꯨ ꯆꯍꯤ ꯱꯴ ꯆꯪꯂꯦ
[nipiməca ədu]$_{NP}$ [cəhi təɹaməɹi]$_{NP}$ cəŋle\VP
The small girl is approaching 14 years.

Compound Sentences

In Manipuri, Compound Sentences (SCpd) are formed by conjoining two or more simple sentences using lexical coordinators (coordinate conjunctions). Example 16 illustrates a compound sentence formed by conjoining three simple sentences using the coordinate conjunction (Conj) ꯑꯗꯨꯒ (ədugə).

Example 16.
ꯑꯍꯟꯕ ꯖꯨ-ꯂꯤꯛꯂꯤ ꯑꯃ ꯁꯠꯂꯤ ꯑꯗꯨꯒ ꯃꯌꯥꯅ ꯃꯈꯨꯝ ꯂꯩꯈꯥꯏ ꯑꯗꯨꯒ ꯌꯥꯎꯝ ꯃꯔꯤ-ꯃꯉꯥ ꯊꯛꯆꯤꯜꯂꯤ
[əhənbə ju-likli əmə sətli]$_{SSim}$ ədugə\Conj [məjanə məkʰum ləikʰai]$_{SSim}$ ədugə\Conj [jəum məɹi-məŋa tʰəkcilli]$_{SSim}$
A new wine bottle is pulled out and with teeth the lid is opened and four-five mouthful is drank.

Complex Sentences

In Manipuri, Complex Sentences (SCplx) are formed by embedding one or more sentences within another sentence (Thoudam, 1980). While embedding, only one of the sentences acts as the main Clause (ClMain) and the remaining become subordinate Clauses (ClSub). Main clauses (can also be considered as a simple sentence) can stand alone and act as complete sentences, while subordinate clauses cannot. Subordinate clauses precede the main clause (Singh, 2013) and are dependent on the main clause. They can be classified as *nominal, adverbial*, *sentential* and *coordinate* clauses.

Nominal Clause

Nominal Clauses (NClause) are formed by nominalizing the verb of a sentence that is to be embedded (Chelliah, 2011) using nominalizer -ꯄ~ꯕ(-pə~bə) as a suffix of the verb. The verb, thus nominalized, takes the form of a verbal noun and may also occur with a case marker suffixed to it. Example 17 is one such complex sentence where the embedded sentence has a nominalized verb.

If a nominal clause helps in clarifying the noun of the main clause that follows, then it becomes a relative clause. The noun of the main clause, thus clarified, becomes the head of the relative clause and is known as the relativized argument. The relativized argument may occur in the form of a noun or noun phrase.

Relative clauses in Manipuri are generally found to be *externally headed* (Chelliah, 2011), but *internally headed* and *headless* relative clauses exist as well.

Example 17.
ꯍꯩꯅꯧ ꯃꯈꯣꯡꯗ ꯇꯨꯝꯂꯤꯕ ꯅꯤꯄꯥ ꯑꯗꯨ ꯂꯤꯜꯅ ꯆꯤꯛꯂꯦ
[həinəu məkʰoŋdə tumli-bə\VNoun]$_{ClSub}$ [nipa ədu lilnə cikle]$_{ClMain}$
The man sleeping below the mango tree is bitten by a snake.

Adverbial Clause

Adverbial clauses (Adv Clause) behaves like an adverb and modifies the main clause. Clauses of this type are formed by adding the adverbial suffix -ꯅ (-nə) to the verb of subordinating clause (Singh, 2013) (Verb-adv: General verbs with adverbial suffix). To accommodate the adverbial suffix, the aspectual markers of the verb may also be modified accordingly. Example 18 is one such complex sentence where the embedded sentence has a verb attached with the adverbial suffix -ꯅ (-nə).

Example 18.
ꯂꯩꯔꯕ ꯄꯔꯖꯥ ꯂꯤꯁꯤꯡ-ꯂꯤꯁꯤꯡꯅ ꯆꯅꯤꯡꯉꯥꯏ ꯂꯩꯇꯗꯨꯅ ꯃꯄꯨꯛ ꯄꯥꯏꯗꯨꯅ ꯈꯥꯡꯂꯤ
[ləiɹəbə pɹəɟa lisiŋ-lisiŋnə caniŋŋai ləitədunə\Verb-adv]$_{AdvClause}$ [məpuk paidunə kʰaŋli]$_{ClMain}$
Thousands of poor people are suffering with empty stomachs as they have nothing to eat.

Sentential Clause

Sentential Clauses (SClause) are formed by adding complementizers such as ꯍꯥꯏꯕ (haibə) and ꯍꯥꯏꯅ (hainə) after the verb of the clause being subordinated (Singh, 2013). They are also known as Sentential Complements (SCompl) since they form subordinate clauses with a full-fledged sentence (Bhat and Ningomba, 1997). Example 19 and 20 are two complex sentences formed by using the complementizers ꯍꯥꯏꯕ (haibə) and ꯍꯥꯏꯅ (hainə) respectively.

Example 19.
ꯍꯩꯅꯧ ꯃꯈꯣꯡꯗ ꯇꯨꯝꯂꯤ ꯍꯥꯏꯕ ꯑꯩ ꯈꯉꯤ
[[həinəu məkʰoŋdə tumli]$_{SSim}$ (haibə)$_{SCompl}$]$_{ClSub}$ [əi kʰəŋi]$_{SSim}$
I know that he is sleeping below the mango tree.

Example 20.
ꯍꯩꯅꯧ ꯃꯈꯣꯡꯗ ꯇꯨꯝꯂꯤ ꯍꯥꯏꯅ ꯑꯩꯅ ꯇꯥꯏ
[[həinəu məkʰoŋdə tumli]$_{SSim}$ (hainə)$_{SCompl}$]$_{ClSub}$ [əinə təi]$_{SSim}$
I heard that he is sleeping below the mango tree.

Coordinate Clause (Clause)

Due to the agglutinative nature of Manipuri, lexical coordinators can also appear as a suffix of the verb preceding it (Singh, 2013). In such a case, the aspectual marker of the verb is either replaced by the suffix or modified to accommodate the suffix (Verb-cc: General verbs with suffix coordinator). The lexical coordinator ꯑꯗꯨꯒ (ədugə) of example 16 has the same syntactic behavior to its respective suffix coordinator -ꯒ (-gə).

Example 21.
ꯑꯍꯟꯕ ꯖꯨ-ꯂꯤꯛꯂꯤ ꯑꯃ ꯁꯠꯂꯒ ꯃꯌꯥꯅ ꯃꯈꯨꯝ ꯂꯩꯈꯥꯏꯗꯨꯒ ꯌꯥꯎꯝ ꯃꯔꯤ-ꯃꯉꯥ ꯊꯛꯆꯤꯜꯂꯤ

DOI: 10.3844/jcssp.2021.855.869

[əhənbə ju-likli əmə sətləgə\Verb-cc]$_{ClSub}$ [məjanə məkʰum ləikʰaiɹəgə\Verb-cc]$_{ClSub}$ [jəum məɹi-məŋa tʰəkcilli]$_{ClMain}$
A new wine bottle is pulled out and with teeth, the lid is opened and four-five mouthful is drunk.

Compound-Complex Sentences

Though the basic structure of Manipuri sentences can be *simple, compound and complex,* it is observed that Manipuri accommodates *compound-complex* structure as well and are quite common. Compound-complex sentences (SCpdCplx) are defined as those which have at least one dependent clause and at least two independent clauses. It can be said that compound-complex sentences are a mix of simple, compound and complex sentences. One such sentence has been illustrated in example 22, where a simple and a complex sentence is conjoined using coordinate conjunction thus forming a compound-complex sentence.

Example 22.
ꯃꯥꯂꯦꯝꯁꯤꯗ ꯃꯔꯨꯑꯣꯏꯕ ꯃꯍꯩ ꯃꯔꯣꯡꯒꯤ ꯄꯨꯛꯀꯩ ꯍꯟꯊꯔꯛꯂꯤ ꯑꯗꯨꯒ ꯃꯃꯜꯗꯤ ꯍꯦꯡꯒꯠꯂꯛꯂꯤ ꯍꯥꯢꯅ ꯃꯥꯂꯦꯃꯒꯤ
ꯆꯥꯅ-ꯊꯛꯅꯕ ꯀꯥꯡꯂꯨꯞꯀꯤ ꯑꯈꯟ-ꯑꯍꯩꯁꯤꯡꯅ ꯍꯥꯢꯔꯤ
[malemsidə məɹuoibə məhəi məroŋgi pukkəi hənthəɹəkli]$_{SSim}$ ədugə\Conj [[məməldi hengətləkli həinə]$_{ClSub}$ [malemgi cənə-tʰəknəbə kaŋlupki əkʰəŋ-əhəisiŋnə həiɹi]$_{ClMain}$]$_{SCplx}$
World's important crops stock is decreasing and the price is increasing said the world's food experts.

Constituency Structure

The constituency structure that we present is an abstract idea of how we group words behaving as single units (also called chunks). This structure can be adapted into proper grammar formalisms such as CFG and TAG.

From the language structure, we have discussed in the previous sections, we formulate a constituency structure for Manipuri phrases, clauses and sentences as given in Table 2, 3 and 4 respectively. The left column of these tables lists the chunks, while the right column gives their possible constituents and their ordering as regular expressions. Abbreviations used in the constituency structure have been given in Table 1. Symbols within the regular expression are defined as follows:

( ) for obligatory constituents;
( )? for constituents that can occur once, or not at all;
( )+ for constituents occurring at least once;
( )* for constituents occurring zero or more times; and
| for either of the constituents.

The Manipuri CFG

A Context-Free Grammar (CFG), also called phrase-structure grammar is one of the most widely used formalisms for modeling the constituent structure of a language. It can be used as a tool for sentence generation, or as a tool for analyzing and assigning a structure to a given sentence (Jurafsky and Martin, 2000).

We propose a CFG for the Manipuri language by adapting the constituency structure we have presented. The proposed CFG has been given in Appendix: Table 7 and has a total of 151 production rules. The left column of the table lists non-terminals (chunks), while the right column gives their respective production rules. The grammar covers simple, compound, complex and compound-complex sentences. A fully expanded version of our grammar can be found at http://www.tezu.ernet.in/~nlp/Manipuri/ManipuriCFG_ExpandedV_1.0.txt. The meaning of non-terminals of our grammar has been given in Appendix: Table 6. For terminal symbols, we use an extended version of the BIS tagset(http://tdil-dc.in/tdildcMain/articles/134692Draft%20POS%20Tag%20standard.pdf) for Indian languages. The tagset has been extended to accommodate morphological information to the original tagset. The extended POS tags used in the proposed CFG have been given in Appendix: Table 5. Using our proposed CFG, the parse tree for example 21 has been illustrated in Fig. 1.

Issues with the Grammar

Even though we try to cover the overall structure of the Manipuri language, there are few limitations of the proposed CFG. Since natural languages are a lot more complex than expected, some word categories exhibit exceptional behavior and defy the constituency structure we have defined.

One such word category is a common noun that occasionally behaves as an NP by itself thus defying the constituency structure we have defined. In such as case, neither do they accompany any of the optional constituents of an NP, nor are they marked by any of the case suffixes. In example 23, the common noun ꯃꯥꯢꯈꯨꯝ-ꯐꯤ (maikʰum-pʰi) acts as a head noun and forms an NP by itself:

Example 23.
ꯃꯥꯢꯈꯨꯝ-ꯐꯤ ꯇꯞꯅ-ꯇꯞꯅ ꯂꯧꯊꯣꯛꯀꯤ
maikʰum-pʰi\Noun təpnə-təpnə\Adv ləutʰoki\Verb
The face mask is slowly removed.

Table 1: Abbreviations used in constituency structure

Abbreviation	Description
Noun	General nouns
Noun-loc	Locative nouns
Noun-case	General nouns with case suffix
Noun-dmn	General nouns with demonstrative suffix
Pron	Pronouns
Pron-case	Pronouns with case markers
Pron-dmn	Pronouns with demonstrative suffix
Verb	General verbs
VNoun	Verbal nouns
Verb-adv	General verbs with adverbial suffix
Verb-cc	General verbs with suffix coordinator
Adj	Adjective
Qtf	Quantifier
Dmn	Demonstrative
Adv	Adverb
Cop	Copula
Conj	Coordinate conjunction
Scompl	Sentential complements
Clmain	Main clause
Clsub	Subordinate clause
Nclause	Nominal clause
Advclause	Adverbial clause
Sclause	Sentential clause
CClause	Coordinate clause
NP	Noun phrase
VP	Verb phrase
SSim	Simple sentence
Scpd	Compound sentence
Scplx	Complex sentence
Scpdcplx	Compound-complex sentence

Table 2: Structure of Manipuri phrases

Phrase	Constituent pattern
NP	(Adj)* (Noun | Pron) (Qtf | Dmn)?
NP	(Noun | Pron) (Adj)* (Qtf | Dmn)?
NP	(Noun | Pron | NP) (Conj) (Noun | Pron | NP)
NP	(Adj)* (Noun-dmn | Pron-dmn)
NP	(Noun | NP) (Noun-loc)
NP	(Noun-case | Pron-case)
NP	(NClause) (Noun | NP)
VP	((Adv)* | Noun)? (Verb)

Table 3: Structure of Manipuri clauses

Clause type	Constituent pattern
Clmain	(SSim)
Clsub	(NClause | AdvClause | SClause | CClause)
Nclause	(NP)* ((Adv)* | Noun)? (Vnoun)
Advclause	(NP)* ((Adv)* | Noun)? (Verb-adv)
Sclause	(SSim) (SCompl)
Cclause	(NP)* ((Adv)* | Noun)? (Verb-cc)

Table 4: Structure of Manipuri sentences

Sentence type	Constituent pattern
Ssim	(NP)* (VP)
Ssim	(NP NP) (Cop)
Scpd	(SSim) (Conj) (SSim)
Scpd	(SCpd) (Conj) (SSim)
Scplx	(ClSub)+ (ClMain)
Scpdcplx	(SSim | SCpd) (Conj) (SCplx)
Scpdcplx	(SCplx) (Conj) (SSim | SCpd)

DOI: 10.3844/jcssp.2021.855.869

We have stated that verbal nouns cannot act as a head by themselves. But there are occasional cases where verbal nouns act as heads with other constituents. Such a case is illustrated in example 24 where the verbal noun ꯑꯄꯅꯕ (əpnbə) forms an NP with the quantifier ꯀꯌꯥ (kəja).

Example 24.
ꯂꯩꯕꯥꯛꯀꯤ ꯆꯥꯎꯈꯠ-ꯊꯧꯔꯥꯡꯗ ꯑꯄꯅꯕ ꯀꯌꯥ ꯄꯤꯔꯤ
[ləibakki caukʰt-tʰəuɹaŋdə]NP [əpnbə\VNoun kəja\Qtf]NP piɹi\Verb
They are giving many hindrances to nation's progress.

In previous sections, we have highlighted the important role played by affixes in language construction. We have seen that some word categories can stand alone as an NP by themselves as a result of the information provided by the suffixes attached. To accommodate such a category of words, the CFG has been appended with the necessary rules. But, for the proposed CFG to be successful, the intended corpus for use should be tagged with the extended version of the BIS tagset we have mentioned. Failing to do so would result in the CFG's inability to recognize the standalone words, that form an NP by themselves, as chunks.

Apart from the issues we have mentioned above, copula and multi-words are also an issue to the CFG as similarly mentioned by Nirmal and Sharma (2018). For these two issues, we follow the solutions suggested by the authors in their work.

Evaluation

We develop a CFG and use Earley's algorithm to effectively parse Manipuri sentences. It is implemented using Python 3.6 and Natural Language Toolkit (NLTK) 3.4.1.

Corpus

In the absence of a large Treebank of Manipuri, we prepare a gold standard corpus consisting of 250 sentences carefully chosen to cover the variety inherent in the language to evaluate the grammar. It is manually annotated using the BIS tagset along with some extensions (Appendix: Table 5). These sentences have been selected from the "Manipuri General Text Corpus". The "Linguistic Resources" has been developed and made available by TDIL, Deity, Government of India.

A total of 220 sentences of the corpus are grammatically correct. These sentences are selected in such a manner that they represent the overall structure of the language. The remaining 30 are manually fabricated negatives, constituted by randomly choosing from the positives. Words and phrases of these chosen sentences are randomly re-arranged and/or deleted to produce grammatically incorrect sentences.

As we are yet to consider punctuation and multi-words, we have preprocessed the sentences. We removed punctuation marks and merged multi-words as single words by hyphenating them.

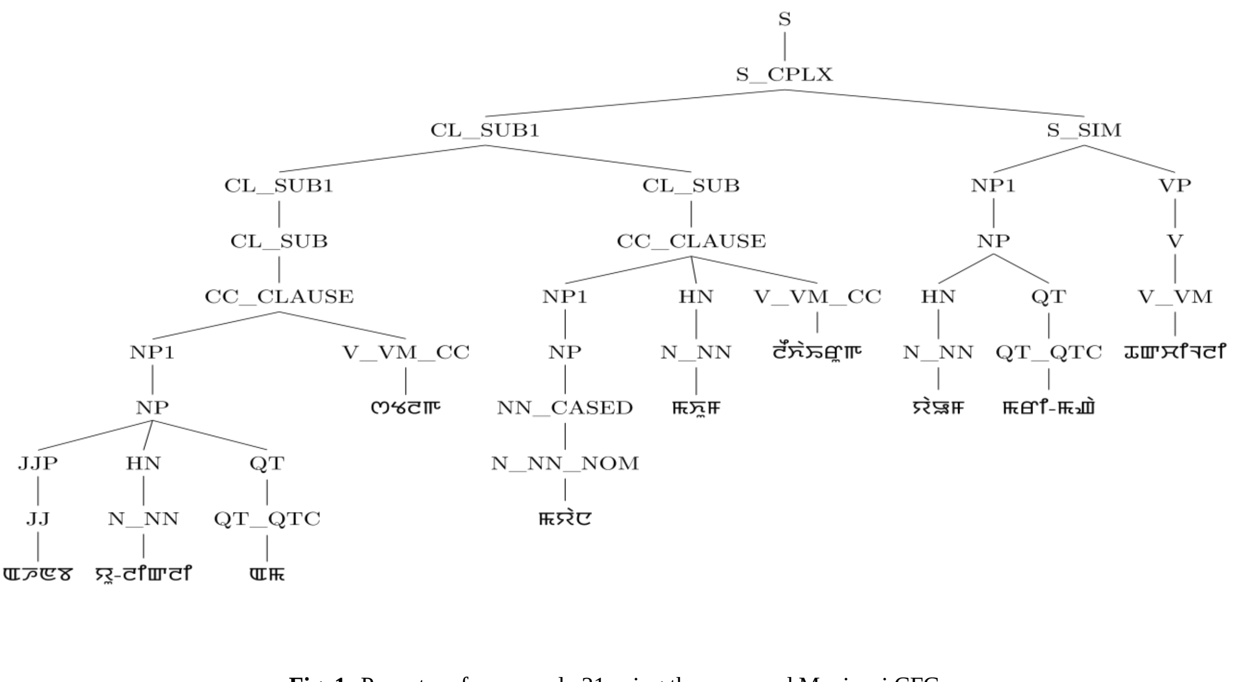

Fig. 1: Parse tree for example 21 using the proposed Manipuri CFG

DOI: 10.3844/jcssp.2021.855.869

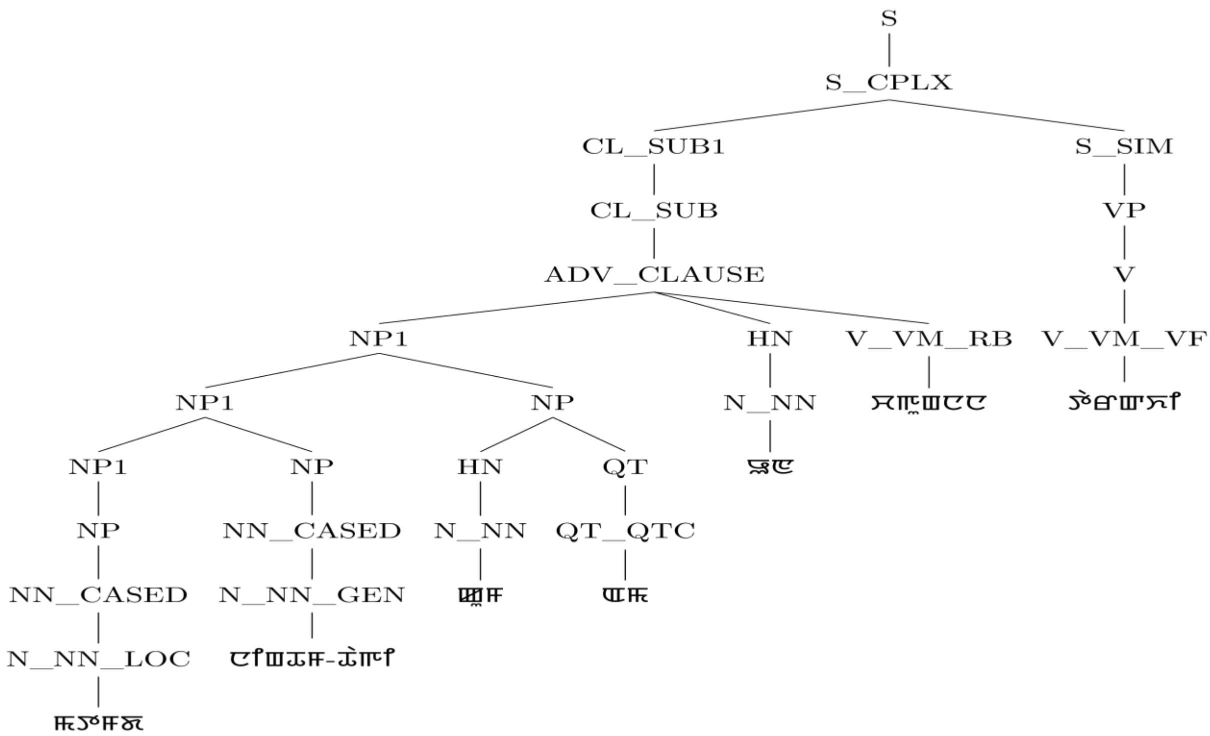

Fig. 2: Parse tree for example 25 (syntactically correct, but semantically wrong)

Results

The parser recognized the sentences into 182 positives and 68 negatives. It is further manually categorized into 178 true positives, 4 false positives, 26 true negatives and 42 false negatives. Since our work is focused on syntax and does not consider semantics, we can consider the 4 false positives as true positives. These sentences are syntactically correct but semantically wrong. One such sentence has been given in example 25 and its respective parse tree in Fig. 2. Thus, we have 182 true positives and 0 false positives.

The accuracy (recognition rate) of the proposed grammar can be obtained as the total sum of true positives and true negatives divided by the total number of sentences. Thus, we obtain an accuracy of 83.20% for our proposed grammar:

Example 25.
* ꯃꯇꯝꯗ ꯅꯤꯡꯊꯝ-ꯊꯥꯒꯤ ꯀꯨꯝ ꯑꯃ ꯎꯟ ꯆꯒꯨꯡꯅꯅ ꯇꯥꯔꯛꯈꯤ
mtmdə niŋtʰm-tʰagi kum əmə un cəguŋnənə taɹəkkʰi
time that winter season one snow thick fall.

Conclusion and Future Work

We have proposed a Manipuri CFG based on the constituency structure of the language. Our CFG attempts to provide an exhaustive framework for representing the syntactic structure of Manipuri. It is quite effective in the syntactic parsing of Manipuri sentences. Due to the unavailability of annotated corpora of the language, purely data-driven syntactic modeling is not feasible for us. Existing work on parsing Manipuri sentences is very few and they are generally preliminary and incomplete. Our work on computational parsing of Manipuri language is pioneering. The grammar covers simple, compound, complex and compound-complex sentences. Using our grammar, we can parse a Manipuri corpus of 14023 sentences that we have collected and thus create a Treebank for Manipuri, though the parse quality is less than perfect.

There are some issues with the grammar, but it can be improved with a more extensive evaluation of the output of parsing. Also, as future work, we can employ a computational method for identifying multi-words and a morphological analyzer that could identify and annotate the type of suffixes attached.

Our model can be a steppingstone for attempting more powerful parsing of this language, as well as other languages, particularly ones from the Tibeto-Burman family.

Acknowledgment

The authors would like to acknowledge TDIL, Deity, Government of India, for providing “Manipuri

General Text Corpus". The authors would also like to acknowledge the Ministry of Social Justice and Empowerment, Govt. of India, New Delhi, for providing financial support through the National Fellowship for Other Backward Classes (NFOBC) to successfully conduct the research.

Funding Information

This research had been partially funded by MHRD sponsored Center of Excellence under FAST Project entitled "Machine Learning Research and Big Data Analysis", Department of Computer Science and Engineering, Tezpur University, India.

Author's Contributions

Yumnam Nirmal: Developing Manipuri Corpus, Text analysis using Manipuri language expertise, development of Manipuri grammar and preparing the manuscript.

Utpal Sharma: Overall supervision and guidance in experiments and presentation.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all the other authors have read and approved the manuscript and no ethical issues are involved.

References

Ababou, N., Mazroui, A., & Belehbib, R. (2017). Parsing Arabic Nominal Sentences Using Context Free Grammar and Fundamental Rules of Classical Grammar. International Journal of Intelligent Systems and Applications, 9(8), 11. https://doi.org/10.5815/ijisa.2017.08.02

Ammar, W., Mulcaire, G., Ballesteros, M., Dyer, C., & Smith, N. A. (2016). Many languages, One Parser. Transactions of the Association for Computational Linguistics, 4, 431-444. https://doi.org/10.1162/tacl_a_00109

Bhat, D. N. S. and Ningomba, M., 1997. Manipuri Grammar. Lincom Europa, München.

Bhat, D. N. S., 2002. Grammatical Relations: The Evidence Against Their Necessity and Universality. Routledge, London.

Blackburn, S., & Opgenort, J. R., (2010). India and the Himalayan chain. In: Atlas of the World's Languages in Danger, Moseley, C. (Eds.), UNESCO, Paris., pp: 59-63.

Chelliah, S. L. (2011). A Grammar of Meithei. De Gruyter Mouton. https://www.degruyter.com/document/doi/10.1515/9783110801118/html

Dorđević, T., & Stojković, S. (2020, September). Syntax Analysis of Serbian Language using Context-free Grammars. In 2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST) (pp. 50-53). IEEE. https://doi.org/10.1109/ICEST49890.2020.9232872

Han, W., Jiang, Y., & Tu, K. (2019, July). Enhancing unsupervised generative dependency parser with contextual information. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 5315-5325). http://dx.doi.org/10.18653/v1/P19-1526

Imoba, S. (2004). Manipuri to English Dictionary. Published by:-S. Ibetombi Devi, Imphal.

Jurafsky, D., & Martin, J. H. (2000). Speech and Language Processing. Pearson Education. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Kapanadze, O., (2019). Parsing the Less-configurational Georgian Language with a Context-Free Grammar. In: Proceedings of the Language Technologies for All (LT4All), European Language Resources Association (ELRA), Paris, UNESCO Headquarters, pp, 342-345. https://lt4all.elra.info/proceedings/lt4all2019/pdf/2019.lt4all-1.86.pdf

Kim, Y., Dyer, C., & Rush, A. M. (2019). Compound probabilistic context-free grammars for grammar induction. arXiv preprint arXiv:1906.10225. http://dx.doi.org/10.18653/v1/P19-1228

Korzeniowski, M., & Mazurkiewicz, J. (2017, June). Rule based dependency parser for polish language. In International Conference on Artificial Intelligence and Soft Computing (pp. 498-508). Springer, Cham. https://doi.org/10.1007/978-3-319-59060-8_45

Le, P., & Zuidema, W. (2015). Unsupervised Dependency Parsing: Let's Use Supervised Parsers. arXiv preprint arXiv:1504.04666. http://dx.doi.org/10.3115/v1/N15-1067

Madhubala, D. P., 1979. Manipuri Grammar. Unpublished dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Poona, Pune, India. http://hdl.handle.net/10603/155110

Makwana, M. T., & Vegda, D. C. (2015). Survey: Natural Language Parsing for Indian Languages. arXiv preprint arXiv:1501.07005. https://arxiv.org/abs/1501.07005

Matisoff, J. A., Baron, S. P., & Lowe, J. B. (1996). Languages and Dialects of Tibeto-Burman. https://stedt.berkeley.edu/pubs_and_prods/STEDT_Monograph2_Lgs-Dialects-TB_with-orig-article.pdf

Mrini, K., Dernoncourt, F., Tran, Q., Bui, T., Chang, W., & Nakashole, N. (2019). Rethinking self-attention: Towards interpretability in neural parsing. arXiv preprint arXiv:1911.03875. http://dx.doi.org/10.18653/v1/2020.findings-emnlp.65

Nirmal, Y., & Sharma, U. (2018). Problems and Issues in Parsing Manipuri Text. In Proceedings of the International Conference on Computing and Communication Systems (pp. 393-401). Springer, Singapore. https://doi.org/10.1007/978-981-10-6890-4_38

Nirmal, Y., & Sharma, U. (2019, December). A Grammar-Driven Approach for Parsing Manipuri Language. In International Conference on Pattern Recognition and Machine Intelligence (pp. 267-274). Springer, Cham. https://doi. org/10.1007/978-3-030-34872-4_30

Pettigrew, W. (1912). Manipuri (Mītei) Grammar with Illustrative Sentences... Pioneer Press.

Primrose, A. J. (1995). A Manipuri Grammar, Vocabulary and Phrase Book. Asian Educational Services.

Sarangthem, B. and Singh, L. L., 2014. Noun phrase in Manipuri (Meiteiron) as a data structure for Computational processes. http://www.academia.edu/download/36391237/ Noun_phrase_structure_of_Meiteilon.pdf

Sharipbay, A., Razakhova, B., Mukanova, A., Yergesh, B., & Yelibayeva, G. (2019, December). Syntax parsing model of Kazakh simple sentences. In Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems (pp. 1-5). https://doi.org/10.1145/3368691.3368745

Sharma, H. (2006). Learners Manipuri Dictionary. Sangam Book Store, Paona Bazar, Imphal.

Sharma, N. L. (1987). Manipuri Grammar. RK Book Agency, Imphal, Manipur, India.

Singh, C. Y. (2000). Manipuri Grammar. Rajesh Publications, New Delhi, India.

Singh, L. N., & Sharma, U. (2012). Modelling the syntax of Manipuri: A Tibeto-Burman language. M. Tech Dissertation form Tezpur University, Tezpur, 784028.

Singh, N. (1987). N: A Meitei Grammar of Roots and Affixes. A Thesis, Unpublish, Manipur University, Imphal.

Singh, S. I. (2013). Manipuri Clause Structure. Unpublished dissertation in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Manipur University, Imphal, India. http://hdl.handle.net/10603/26596

Þorsteinsson, V., Óladóttir, H., & Loftsson, H. (2019, September). A wide-coverage context-free grammar for Icelandic and an accompanying parsing system. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019) (pp. 1397-1404). https://aclanthology.org/R19-1160.pdf

Thoudam, P. C. (1980). A Grammatical Sketch of Meiteiron. Unpublished dissertation in partial fulfilment of the requirements for the degree of Doctor of Philosophy, Jawaharlal Nehru University, New Delhi, India. http://hdl.handle.net/10603/13917

Thoudam, P. C. (1991). Remedial Manipuri. Singjamei, Imphal: ANM Enterprises.

Yang, S., Jiang, Y., Han, W., & Tu, K. (2020). Second-order unsupervised neural dependency parsing. arXiv preprint arXiv:2010.14720. http://dx.doi.org/10.18653/v1/2020.coling-main.347

Yang, S., Zhao, Y., & Tu, K. (2021). PCFGs Can Do Better: Inducing Probabilistic Context-Free Grammars with Many Symbols. arXiv preprint arXiv:2104.13727.https://arxiv.org/abs/2104.13727

Zhou, J., & Zhao, H. (2019). Head-driven phrase structure grammar parsing on Penn treebank. arXiv preprint arXiv:1907.02684. http://dx.doi.org/10.18653/ v1/P19-1230

Appendix

Table 5: Extended Parts-Of-Speech (POS) tags

Category	Suffix Attached	POS Tag
Common Noun	Nominative	N_NN_NOM
	Accusative	N_NN_ACC
	Instrumental	N_NN_INST
	Locative	N_NN_LOC
	Associative	N_NN_ASS
	Genitive	N_NN_GEN

Table 5: Continue

	Genitive	N_NN_GEN
	Accusative	N_NN_ACC
	Demonstrative	N_NN_DM
Proper Noun	Nominative	N_NNP_NOM
	Accusative	N_NNP_ACC
	Instrumental	N_NNP_INST
	Locative	N_NNP_LOC
	Associative	N_NNP_ASS
	Genitive	N_NNP_GEN
	Accusative	N_NNP_ACC
	Demonstrative	N_NNP_DM
Main Verb	Adverbial	V_VM_RB
	Coordinator	V_VM_CC
Personal Pronoun	Nominative	PR_PRP_NOM
	Accusative	PR_PRP_ACC
	Instrumental	PR_PRP_INST
	Locative	PR_PRP_LOC
	Associative	PR_PRP_ASS
	Genitive	PR_PRP_GEN
	Demonstrative	PR_PRP_DM
Reflexive pronoun	Nominative	PR_PRF_NOM
	Accusative	PR_PRF_ACC
	Instrumental	PR_PRF_INST
	Locative	PR_PRF_LOC
	Associative	PR_PRF_ASS
	Genitive	PR_PRF_GEN
	Demonstrative	PR_PRF_DM
Relative pronoun	Nominative	PR_PRL_NOM
	Accusative	PR_PRL_ACC
	Instrumental	PR_PRL_INST
	Locative	PR_PRL_LOC
	Associative	PR_PRL_ASS
	Genitive	PR_PRL_GEN
	Demonstrative	PR_PRL_DM
Reciprocal pronoun	Nominative	PR_PRC_NOM
	Accusative	PR_PRC_ACC
	Instrumental	PR_PRC_INST
	Locative	PR_PRC_LOC
	Associative	PR_PRC_ASS
	Genitive	PR_PRC_GEN
	Demonstrative	PR_PRC_DM
Wh-word pronoun	Nominative	PR_PRQ_NOM
	Accusative	PR_PRQ_ACC
	Instrumental	PR_PRQ_INST
	Locative	PR_PRQ_LOC
	Associative	PR_PRQ_ASS
	Genitive	PR_PRQ_GEN
	Demonstrative	PR_PRQ_DM
Indefinite pronoun	Nominative	PR_PRI_NOM
	Accusative	PR_PRI_ACC
	Instrumental	PR_PRI_INST
	Locative	PR_PRI_LOC
	Associative	PR_PRI_ASS
	Genitive	PR_PRI_GEN
	Demonstrative	PR_PRI_DM

Table 6: Meaning of non-terminals used in the proposed CFG

Syntactic tag	Description
S	Start symbol of the CFG
S_SIM	Simple sentence
S_CPD	Compound sentence
S_CPLX	Complex sentence
S_CPD_CPLX	Compound-complex sentence
CL_SUB	Subordinate clause
S_CLAUSE	Sentential clause
N_CLAUSE	Nominal clause
ADV_CLAUSE	Adverbial clause
CC_CLAUSE	Coordinate clause
NP	Noun phrase
VP	Verb phrase
JJP	Adjective phrase
RBP	Adverb phrase
HN	Head noun
PR	Pronoun
QT	Quantifier
DM	Demonstrative
V	Verb
NNP_CASED	Proper noun with case marker
NN_CASED	Common noun with case marker
PR_CASED	Pronoun with case marker
PR_DM	Pronoun with demonstrative marker

Table 7: Proposed Manipuri context-free grammar

Non terminal symbol	Production rules
S	S_SIM | S_CPD | S_CPLX
S_SIM	NP1 VP | VP | NP1 COP
S_CPD	S_CPD CC_CCD S_SIM | S_SIM CC_CCD S_SIM
S_CPLX	CL_SUB1 S_SIM
S_CPD_CPLX	S_SIM CC_CCD S_CPLX | S_CPD CC_CCD S_CPLX
S_CPD_CPLX	S_CPLX CC_CCD S_SIM | S_CPLX CC_CCD S_CPD
CL_SUB1	CL_SUB1 CL_SUB | CL_SUB
CL_SUB	N_CLAUSE | ADV_CLAUSE | S_CLAUSE | CC_CLAUSE
N_CLAUSE	NP1 RBP N_NNV | NP1 N_NNV | NP1 HN N_NNV
N_CLAUSE	RBP N_NNV | HN N_NNV | N_NNV
ADV_CLAUSE	NP1 RBP V_VM_RB | NP1 HN V_VM_RB | NP1 V_VM_RB
ADV_CLAUSE	RBP V_VM_RB | HN V_VM_RB | V_VM_RB
S_CLAUSE	S_SIM CC_CCS_UT
CC_CLAUSE	NP1 RBP V_VM_CC | NP1 HN V_VM_CC | NP1 V_VM_CC
CC_CLAUSE	RBP V_VM_CC | HN V_VM_CC | V_VM_CC
NP1	NP1 NP | NP
NP	JJP HN | HN JJP | HN DM | HN QT | JJP HN QT
NP	HN JJP QT | JJP HN DM | HN JJP DM | N_NNP | PR
NP	HN CC_CCD HN | NP CC_CCD NP
NP	NP CC_CCD NP | NP CC_CCD NP
NP	NN_CASED | NNP_CASED | PR_CASED
NP	PR_DM | JJP PR_DM | N_NN_DM |
NP	N_NNP_DM | JJP N_NN_DM | JJP N_NNP_DM
NP	N_NN N_NST | N_NNP N_NST | NP N_NST
NP	N_CLAUSE HN | N_CLAUSE NP
VP	RBP V | N_NN V | V
V	V_VM | V_VAUX | V_VM_VNG

Table 7: Continue

V	V_VM_VF | V_VM_VNF | V_VM_VINF
JJP	JJP JJ | JJ
RBP	RBP RB | RB
HN	N_NN | N_NNP | PR
PR	PR_PRP | PR_PRF | PR_PRL | PR_PRC | PR_PRQ | PR_PRI
QT	QT_QTF | QT_QTC | QT_QTO
DM	DM_DMD | DM_DMR | DM_DMQ | DM_DMI
NNP_CASED	N_NNP_NOM | N_NNP_ACC | N_NNP_INST
NNP_CASED	N_NNP_LOC | N_NNP_ASS | N_NNP_GEN
NN_CASED	N_NN_NOM | N_NN_ACC | N_NN_INST
NN_CASED	N_NN_LOC | N_NN_ASS | N_NN_GEN
PR_CASED	PR_PRP_NOM | PR_PRP_ACC | PR_PRP_INST | PR_PRP_LOC
PR_CASED	PR_PRP_ASS | PR_PRP_GEN | PR_PRF_NOM | PR_PRF_ACC
PR_CASED	PR_PRF_INST | PR_PRF_LOC | PR_PRF_ASS | PR_PRF_GEN
PR_CASED	PR_PRL_NOM | PR_PRL_ACC | PR_PRL_INST | PR_PRL_LOC
PR_CASED	PR_PRL_ASS | PR_PRL_GEN | PR_PRC_NOM | PR_PRC_ACC
PR_CASED	PR_PRC_INST | PR_PRC_LOC | PR_PRC_ASS | PR_PRC_GEN
PR_CASED	PR_PRQ_NOM | PR_PRQ_ACC | PR_PRQ_INST | PR_PRQ_LOC
PR_CASED	PR_PRQ_ASS | PR_PRQ_GEN | PR_PRI_NOM | PR_PRI_ACC
PR_CASED	PR_PRI_INST | PR_PRI_LOC | PR_PRI_ASS | PR_PRI_GEN
PR_DM	PR_PRP_DM | PR_PRF_DM | PR_PRL_DM
PR_DM	PR_PRC_DM | PR_PRQ_DM | PR_PRI_DM