A Finite-State Morphological Analyser for Sindhi

Raveesh Motlani[1], **Francis M. Tyers**[2] **and Dipti M. Sharma**[1]
[1]FC Kohli Center on Intelligent Systems (KCIS), International Institute of Information Technology
Hyderabad, Telangana, India
[2] HSL-fakultehta, UiT Norgga árktalaš universitehta
N-9019 Romsa
raveesh.motlani@research.iiit.ac.in, francis.tyers@uit.no, dipti@iiit.ac.in

Abstract

Morphological analysis is a fundamental task in natural-language processing, which is used in other NLP applications such as part-of-speech tagging, syntactic parsing, information retrieval, machine translation, etc. In this paper, we present our work on the development of free/open-source finite-state morphological analyser for Sindhi. We have used Apertium's *lttoolbox* as our finite-state toolkit to implement the transducer. The system is developed using a paradigm-based approach, wherein a paradigm defines all the word forms and their morphological features for a given stem (lemma). We have evaluated our system on the Sindhi Wikipedia, which is a freely-available large corpus of Sindhi and achieved a reasonable coverage of about 81% and a precision of over 97%.

Keywords: Sindhi, Morphological Analysis, Finite-State Machines

1. Introduction

Morphology describes the internal structure of words in a language. A morphological analysis of a word involves describing one or more of its properties such as: gender, number, person, case, lexical category, etc. Morphological analysis of a word thus becomes a fundamental and crucial task in natural-language processing for any language. Morphological analysis is vital in various NLP applications such as spell-checking, part-of-speech tagging, dependency parsing, information retrieval, machine translation, etc.

Sindhi is an Indo-Aryan language with an inflectional morphology. The language can be written in one of two writing systems, either in the Devanagari writing system or the Perso-Arabic writing system. In this paper we will discuss our approach in building a morphological analyser for Sindhi in Perso-Arabic script.

The paper is organised in the following manner. We give an overview of the language in Section 2. We have discussed the work done previously in Section 3. We have described our developmental process in Section 4. The evaluation of our system is described in Section 5. The future work is discussed in Section 7. The concluding remarks are given in Section 7.

2. The linguistic properties of Sindhi

Sindhi belongs to the Indo-Aryan language family. It has a long history and its origins can be traced back to 1500 BCE. The name Sindhi is derived from *Sindhu* which was the local name for the Indus river. The Sindh region has gone through various invasions and as a result the vocabulary contains many loan words from Persian, Arabic, Hindi and Urdu.

Sindhi is an official language in both Pakistan and India. Estimated number of native speakers are approximately 59 million in these two countries (Khubchandani, 2003). However, it is also spoken by people in various other countries. Some of the linguistic properties of Sindhi are discussed below.

ڳ	[ŋ]	ڃ	[ɲ]	ٻ	[ɓ]
گ	[g]	ڄ	[ʄ]	ڀ	[$b^ɦ$]
ڪ	[k]	ڇ	[c^h]	ڏ	[$ɗ^ɦ$]
ڻ	[ɳ]	ٺ	[$ʈ^h$]	ڊ	[ɗ]
ق	[p^h]	ٽ	[ʈ]	ڍ	[ɖ]
ڙ	[ɽ]	ٿ	[t^h]	ڌ	[$ɖ^ɦ$]

Table 1: The characters in the Sindhi alphabet which are not found in the Persian alphabet and their phonetic value.

2.1. Orthography

Historically, Sindhi has been written using many writing systems: Landa, Khojki, Waranki, Khudawadi, Gurmukhi, Perso-Arabic and Devanagari. During the colonial rule, the British chose Perso-Arabic as standard script and most of the literature has been written in Perso-Arabic since then. Currently, both Devanagari and Perso-Arabic are official forms of writing Sindhi in India, while in Pakistan the only official form is Perso-Arabic. In our work we used the Perso-Arabic script, which is used by most Sindhi speakers on the Internet and has large amount of content freely available on the web.

The Sindhi alphabet in Perso-Arabic is a variant of the Persian alphabet. It shares a lot of characters with Arabic and Persian alphabets. It is composed of 52 letters, which includes Persian letters, digraphs and eighteen other letters (see Table 1) to capture the sounds particular to Sindhi and other Indo-Aryan languages.

2.2. Morphology

Sindhi, like many Indo-Aryan languages, has a very rich morphology. It uses suffixes for constructing derivational and inflectional morphemes. Below, we have described certain aspects of Sindhi morphology in various lexical categories.

2.2.1. Nominals

Sindhi words generally end in a vowel, which also help them classify into their appropriate gender (in case of nouns). There are two genders in Sindhi : masculine and feminine. Feminine nouns generally end with the following vowels: ا [ə], ا [a] and اِ [i] and the masculine nouns usually end with و [o] or اُ [u]. Nouns inflect according to number (singular and plural) and case (nominative and oblique). Cases are marked to identify the function of the noun. Nouns in oblique case are generally followed by a postposition.

Pronouns, like nouns, also inflect with gender and number. Pronouns are a closed category but may be categorized into several subcategories: personal, demonstrative, indefinite, interrogative, reflexive, relative and co-relative.[1]

Adjectives can be classified into two main classes, declinable (سنو *suṭho* 'good') and indeclinable (آسان *āsan* 'easy'). All adjectives ending in و [o] are declinable and agree in gender, number and case with the following noun. Adjectives have three degrees for comparison: analytical positive, comparative (هُن کان وڌو *hun khā vado* 'older than him') and superlative (سڀ کان وڌو *sabh khā vado* 'the oldest').

2.2.2. Verbs

Verbs are morphologically the richest and largest category of all. Here are some properties of the Sindhi verbs.

The auxiliary verbs modify the action expressed by the main verb. They may indicate mood, tense and aspect. They constitute a small class of words (سگ *səgh-* 'be able').

Sindhi verbs also exhibit transitivity. An example with a transitive verb would be : مان کت لکان ٿو *mã khat lIkhã tho* 'I write a letter'. Similarly, an example with an intransitive verb : مان سمهان ٿو *mã sumhã tho* 'I am sleeping'. The distinction between transitive and intransitive verbs is important for morphological disambiguation and parsing as the transitivity of the verb determines the subject case in certain tenses.

As mentioned previously, verbs are the largest category and work on adding a lot of verbs is still in progress.

According to the book of grammar[2] we have have covered all inflectional forms for auxiliary verbs and copula (هو *ho-* 'be'). In case of transitive (ڪرڻ *karan* 'to do') and intransitive (سمهڻ *sumhan* 'to sleep') verbs, we have covered inflectional forms for simple tenses with all numbers, person and genders.

2.2.3. Other categories

The remaining parts of speech are uninflecting closed categories, these include : adverbs, particles, postposition, conjunction and interjections.

Postpositions are functional words which are used to show grammatical relations. They are indeclinable with the exception of جو *jo* 'of', which declines like an adjective for gender, number and case.

Conjunctions are indeclinable in Sindhi. These are further classified into two categories: coordinating conjunctions and subordinating conjunctions. Coordinate conjunctions are either *cumulative*, which add one statement to another (۽ *ain* 'and'), or *alternative*, which express a choice between two alternatives (يا *ya* 'or'). Subordinate conjunctions join subordinate clauses to construct a complex sentence. They may also at times express time, location, direction, manner, reason, condition, result, concession etc. (جڏهين - تڏهن *jadhein - tadhein* 'when - then').

Interjections are words or phrase which expresses some sudden feeling or emotion. Some examples are: واقعي *wakaI* 'Really!' , واهه *wah* 'Wow!'.

3. Previous work

Apertium's *lttoolbox* (Forcada et al., 2011) has been used to develop finite state morphological analysers for as many as 46 languages. Those developed for some of the other Indo-Aryan languages are: Tamil (Parameswari, 2010), Assamese (Rahman and Sarma, 2015), Oriya (Jena et al., 2011) and Malayalam (Rinju et al., 2013; Vinod et al., 2012).

Some research on Sindhi has been done in the past few years. A rule-based POS tagger (Mahar and Memon, 2010) had been developed for Sindhi (Perso-Arabic). The authors developed a lexicon of 26,355 entries and a tagset containing 67 tags. Using both these resources along with about 186 disambiguation rules, their Sindhi POS tagger reported an accuracy of 96.28% . They have also contributed towards other aspects of natural language processing such as text segmentation, language modelling, etc.

Rahman et al. (2010) have worked on Sindhi noun morphology. They have tried to capture Sindhi noun inflections through finite-state machines. Unfortunately both the above resources are not publicly available so we could not evaluate them or use them in our work.

A statistical POS tagger using Conditional Random Fields was developed by Motlani et al. (2015) for Sindhi Devanagari. They reported 92% average accuracy, which was evaluated using 10-fold cross-validation. They used a POS annotated corpus of 37162 words, developed using an adaptation of BIS (Bureau of Indian Standards) tagset (Jha, 2010) for Sindhi. This work is one of the first works published on NLP for Sindhi-Devangari.

(Oad, 2012) has implemented computational resource grammar for Sindhi in Grammatical Framework[3]. Grammatical Framework (Ranta, 2009) (abbreviated as, GF) is a functional and natural language processing programming language, which is designed for writing grammars. The Sindhi GF library has around 360 entries in its lexicon. These number of entries belonging to each part-of-speech category is tabulated in Table 2. Sindhi grammar library has used different categories and functions to manage the morphology and syntax implementation. The library has 44 categories and 190 functions. We had referred this library during initial stages of development of our work and we had also found and reported some mistakes in it. For instance, some feminine nouns were assigned to a paradigm

[1] A co-relative pronoun is a feature of some Indo-Aryan languages where a relative pronoun in the relative clause has a counterpart (the co-relative) in the main clause.

[2] http://www.ciil-lisindia.net/Sindhi/sindhi_struct.html

[3] http://www.grammaticalframework.org/lib/src/sindhi

Part of speech	Number of stems	
	Apertium	*GF*
Noun	1191	179
Verb	88	54
Adjective	766	49
Proper noun	958	1
Adverb	267	21
Numeral	52	4
Conjunction	17	7
Interjection	7	1
Abbreviation	6	0
Postposition	66	21
Pronoun	23	14
Determiner	13	10
Total:	3454	361

Table 2: Number of stems in each of the categories in the Apertium and Grammatical Framework lexica.

for masculine nouns.

4. Developing the morphological analyser

We initiated our work on developing Sindhi morphological analyser with the help of three resources. The first one was an article by (Rahman and Bhatti, 2010), which described how Sindhi nouns inflect. This aided us in creating our first few paradigms for nouns. Along with noun paradigms we also created paradigms for some closed categories, such as, prepositions, conjunctions, and open categories of adjectives and adverbs.

The second resource was Sindhi GF library. It helped us in verifying some of the paradigms that we had already defined. It was also helpful in adding verb and pronoun paradigms and improve the paradigms for nouns. The third resource was a corpus, a collection of articles from Sindhi Wikipedia. Then, we used our knowledge of Sindhi, the Wikipedia corpus and lexicon from GF to add words in the Apertium lexicon. A small dictionary of Sindhi, Hindi and Urdu words was also developed alongside adding lexicon to Apertium.

The process of adding words was completely manual. We had parsed the corpus to create a list of words with their frequency, sorted in descending order. We went through this list word by word and added each word we knew to the lexicon along with the corresponding paradigm that it belonged to. An example paradigm and entry can be found in Figure 2. Also, an example sentence and its morphological analysis produced by Apertium is shown in Figure 3.

We referred to dictionaries[4] and grammar books[5] as well, that were available online or in printed hard copy form, to add more words and paradigms. We also tried crowdsourcing for understanding words that we could not find anywhere, by asking learned people through social media. The current statistics of words in lexicon are tabulated in Table 2, also drawing a comparison with GF lexicon. There are total 72 paradigms in our analyser right now.

[4] http://www.sindhila.org/dics.php?dic=sindhidevnagarienglishdic

[5] http://www.ciil-lisindia.net/Sindhi/sindhi_struct.html and http://www.sindhila.org/sindhilearning/

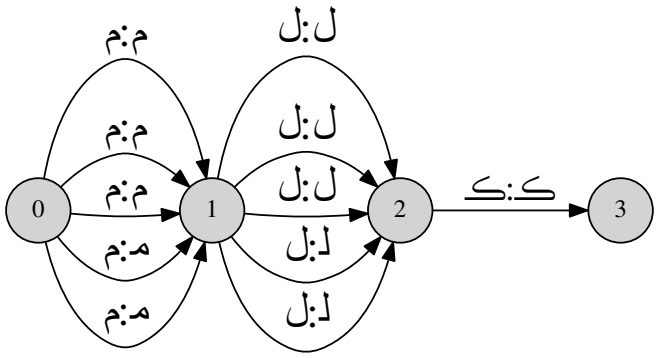

Figure 1: Example of a fragment of a transducer for ملڪ *mulk* 'country' demonstrating how badly encoded text is dealt with. The left side is the output and the right side is the input. Note that the ڪ [k] character also has initial, medial and final forms, but these are produced with a separate code point, ـ (U+0640).

4.1. Orthographic issues

We use Unicode as a character set for our lexicon as this is a global standard. However, when working with Sindhi and other similar writing systems it presents a number of issues:

1. One letter may have many forms, all of which have separate Unicode code points: isolate, initial, medial, final. For example, the letter م [m] (U+0645 in its canonical form) may appear as مـ (U+FEE3, initial), ـمـ (U+FEE4, medial), ـم (U+FEE2, final) or م (U+FEE1, isolate). In most text these specific presentation forms do not appear, as the choice between them is determined by the layout software.

2. Since Sindhi shares its script with Persian, Arabic and Urdu , there are a lot of character homophones that get introduced into Sindhi. One example is the letter 'h' : ه U+0647 (in Sindhi) and ھ U+06BE (in Urdu). Both these letters are used interchangeably in the text.

3. A lot of Perso-Arabic script based languages do not use diacritics marks in their texts. This creates several issues:

   (a) Semantic ambiguities: Words may have multiple interpretations when used without diacritics. For example: ملڪ *mlk* can be either *mulk* 'country' or *milk* 'milk'.

   (b) Syntactic ambiguities: Sometimes presence of diacritics changes not only the meaning but lexical category as well. For example: هو *ho* can be verb 'was' or pronoun 'that'.

   The practice of using or not using diacritics in text is not standard among writers. A lot of texts contain both kinds of words. This created problem for us as many times our analyser could not analyse a word (with diacritics) despite the correct analysis of that word (without diacritics) was present in the lexicon.

```
Paradigm:
    <pardef n="چوڪر/و__n_m" c="I">
      <e><p><l>و</l><r>و<s n="n"/><s n="m"/><s n="sg"/><s n="nom"/></r></p></e>
      <e><p><l>ي</l><r>و<s n="n"/><s n="m"/><s n="sg"/><s n="obl"/></r></p></e>
      <e><p><l>ا</l><r>و<s n="n"/><s n="m"/><s n="pl"/><s n="nom"/></r></p></e>
      <e><p><l>ن</l><r>و<s n="n"/><s n="m"/><s n="pl"/><s n="obl"/></r></p></e>
    </pardef>
Entry:
    <e lm="چوڪرو"><i>چوڪر</i><par n="چوڪر/و__n_m"/></e>
```

Figure 2: An example paradigm `pardef` and entry `e` for the noun چوڪرو *chhokro* 'boy' in XML format.

```
^ڏياري/ڏياري<n><f><sg><obl>$
^جي/جي<post><m><sg><obl>$
^موقعي/موقع<n><m><sg><obl>$
^تي/تي<post>$
^ماڻهو/ماڻهو<n><m><pl><nom>$
^دڪانن/دڪان<n><m><pl><obl>$
^۽/۽<cnjcoo>$
^گهرن/گهر<n><m><pl><obl>$
^جي/جي<post><f><sg><obl>$
^صفائي/صفائي<adj><f><sg><obl>$
^ڪرڻ/ڪندا<vblex><tv><pres><hab><p3><m><pl>$
^آهن/هو<vbser><pres><p3><pl>$
```

Figure 3: Example output for the sentence
ڏياري جي موقعي تي ماڻهو دڪانن ۽ گهرن جي صفائي ڪندا آهن
dyarī jey maukey tey manhoo dukānan ain gharan jī safaI kandā āhin "On the occasion of Diwali, people clean shops and houses."

In order to get around the problem of analysing badly or incorrectly encoded Unicode text, for each canonical letter, we allow the other code points as variants on the input side of the transducer (shown in Figure 1). We also do the same for character homophones.

5. Evaluation

We have evaluated the morphological analyser in two ways. The first was by calculating the naïve coverage and mean ambiguity on freely available corpora (shown in Table 4) and the second was by calculating precision and recall. Naïve coverage refers to the percentage of surface forms in a given corpora that receive at least one morphological analysis. Although, forms counted by this measure may have other analyses which are not delivered by the transducer. The mean ambiguity measure was calculated as the average number of analyses returned per token in the corpus.

5.1. Corpora

There are a lot of websites, books, blogs, etc. on the internet which can serve as sources of raw text in Sindhi. Our primary source of text data was Wikipedia, which fortunately exists for Sindhi language too. We will describe how we sourced the data below.

5.1.1. Sindhi Wikipedia

The collection of articles published on Wikipedia for all the languages available in the forms of compressed dumps on a website[6]. These dumps are updated on a regular basis. We downloaded a Sindhi Wikipedia dump[7] and created our corpus in the following manner.

1. We decompressed the `sdwiki-20150826-pages-articles.xml.bz2` file, which gives us all the articles in XML format.
2. We extracted the raw text from compressed file itself by using a script.[8] The extracted raw text was of size 2.6MB.
3. We noticed there were still a lot of problems in the data. For instance, there was XML metadata and portions of non-Sindhi (English, Urdu, Arabic, Persian) texts. So, we cleaned it manually and eventually got 2.5MB of Sindhi text.
4. The final text has 303,401 words while the filtered out non-Sindhi text had 7,570 words.

5.1.2. Web corpus

We also gathered more textual data by scraping from various domains[9] on the web, as done in (Rahman, 2010). These include texts from news articles (politics, current affairs, sports, editorials), blog posts, forum posts, etc. The size of this collection is about 6.4 MB, with about 805,000 words.

5.2. Precision and Recall

Precision and recall are measures of the average accuracy of analyses provided by a morphological transducer. Precision represents the number of the analyses given for a form that are correct. Recall is the percentage of analyses that are deemed correct for a form (by comparing against a gold standard) that are provided by the transducer.
To calculate precision and recall, it was necessary to create a hand-verified list of surface forms and their analyses.

[6] `https://dumps.wikimedia.org/`
[7] `https://dumps.wikimedia.org/sdwiki/sdwiki-20150826-pages-articles.xml.bz2`
[8] `http://wiki.apertium.org/wiki/Wikipedia_Extractor`
[9] `http://svn.code.sf.net/p/apertium/svn/incubator/apertium-snd/URLs.txt`

	Precision (%)	Recall (%)
Known tokens	97.68	97.52
All tokens	97.68	72.61

Table 3: Precision and recall presented as percentages

We extracted 1000 unique surface forms at random from the Wikipedia corpus, and checked that they were valid words in the languages and correctly spelled. Where a word was incorrectly spelled or deemed not to be a form used in the language, it was discarded.

This list of surface forms was then analysed with the analyser, and each analysis was checked. Where an analysis was erroneous, it was removed; where an analysis was missing, it was added. This process gave us a 'gold standard' morphologically analysed word list of 384 forms. The list is publicly available for each language in Apertium's SVN repository.

We then took the same list of surface forms and ran them through the morphological analyser once more. Precision was calculated as the number of analyses which were found in both the output from the morphological analyser and the gold standard, divided by the total number of analyses output by the morphological analyser.

Recall was calculated as the total number of analyses found in both the output from the morphological analyser and the gold standard, divided by the number of analyses found in the morphological analyser plus the number of analyses found in the gold standard but not in the morphological analyser. The results for precision and recall are presented in Table 3.

5.3. Qualitative evaluation

We manually analysed the output for error analysis and found the following problems :

1. **Diacritics**: When the diacritised input is given, it is difficult to lookup in the lexicon and disambiguate.
2. **Miscategorization**: Some stems were assigned to wrong categories or paradigms. For example, لکڻ *likhan* 'to write' was initially marked as an intransitive verb and later corrected to transitive. Such mistakes also existed in GF lexicon.
3. **Incomplete Paradigms**: Some paradigms were insufficient or incorrect. For example, the suffix ءَ [ə] that is attached to some nouns and proper nouns in oblique cases was missing in the paradigms earlier.
4. **Size of Lexicon**: Although the coverage is greater than 80%, the lexicon is still small. Therefore, some of the random surface forms selected for evaluation had unanalysed output.

6. Future work

We plan to improve the analyser further by taking up the following steps. We would add more stems to, especially verbs and nouns. We would also like to explore automatic lexicon extraction techniques (Hulden et al., 2014) since we have some paradigms defined already.

Corpus	Tokens	Coverage (%)	Mean ambig.
Wiki.	341.5k	81.12	3.2
Blogs	805k	76.68	3.4
Average	–	78.90	3.3

Table 4: Results of naïve coverage tests.

The verb morphology of Sindhi is quite extensive and we do not currently cover all productive processes. An important addition would be to improve the verbal paradigms by adding missing productive suffixes.

Given the problems caused by diacritics, we would also like to look at how diacritics can be integrated into the transducer. The ideal way would be to include all stems in the lexicon with diacritic forms, and then allow forms without diacritics to be analysed. However, the main obstacle to this is that most lexical resources for Sindhi do not use forms with diacritics, so this would need to be done manually by a proficient native speaker.

We can also use this work to improve the Sindhi resource grammar in GF. It would then be able to produce parse trees of more sentences than it currently can.

Another plan that we have is to develop rule-based machine translation systems in Apertium for Sindhi. The morphological analyser presented in this work also has a monolingual dictionary which is an essential part of a language pair in rule-based machine translation systems in Apertium. We already have an Urdu-Hindi MT system in Apertium and since these languages are closely related and share common linguistic properties, we can use them to develop MT systems for Urdu-Sindhi and Sindhi-Hindi language pairs.

7. Concluding remarks

We have presented the first freely available morphological analyser for Sindhi. The analyser is based on a word-and-paradigm model of Sindhi morphology and is implemented as a finite-state transducer which can also be used for generation. The lexicon is entirely encoded in Unicode and has reasonable coverage for few lexical entries. The precision of the paradigms is good, as would be expected from a manually constructed resource, but the recall is limited by the small size of the lexicon.

Acknowledgements

We would like to thank Mehtab Ahmed Solangi for helping us with annotation and translation of several words in the Sindhi language. We would also like to thank, Mr. Bhagwan Babani and Mr Chunnilaal Wadhwani for helping us in understanding Sindhi (Perso-Arabic) better. We are also thankful to the anonymous reviewers for their invaluable feedback. This publication reflects the authors views only.

8. References

Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., Sánchez-Martínez, F., Ramírez-Sánchez, G., and Tyers, F. M. (2011). Apertium: a free/open-source platform for rule-based machine translation. *Machine Translation*, 25(2):127–144.

Hulden, M., Forsberg, M., and Ahlberg, M. (2014). Semi-supervised learning of morphological paradigms and lexicons. In *Proceedings of EACL 2014*, pages 569–578.

Jena, I., Chaudhury, S., Chaudhry, H., and Sharma, D. M. (2011). Developing oriya morphological analyzer using lt-toolbox. In *Information Systems for Indian Languages*, pages 124–129. Springer.

Jha, G. N. (2010). The tdil program and the indian language corpora initiative (ilci). In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010). European Language Resources Association (ELRA)*.

Khubchandani, L. M. (2003). Sindhi. In George Cardona et al., editors, *The Indo-Aryan Languages*, pages 622–658. Routledge, Abingdon.

Mahar, J. A. and Memon, G. Q. (2010). Rule based part of speech tagging of Sindhi language. *International Conference on Signal Acquisition and Processing*, 134:101–106.

Motlani, R., Lalwani, H., Shrivastava, M., and Sharma, D. M. (2015). Developing part-of-speech tagger for a resource poor language: Sindhi. In *Proceedings of the 7th Language and Technology Conference (LTC 2015), Poznan, Poland*.

Oad, J. D. (2012). *Implementing GF Resource Grammar for Sindhi language*. Msc. thesis, Chalmers University of Technology, Gothenburg, Sweden.

Parameswari, K. (2010). An improvized morphological analyzer cum generator for tamil: A case of implementing the open source platform apertium. *Knowledge Sharing Event*.

Rahman, M. U. and Bhatti, M. I. (2010). Finite state morphology and Sindhi noun inflections. *In Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation, PACLIC 24, Tohoku University, Japan*, 134:669–676.

Rahman, M. and Sarma, S. K. (2015). An implementation of apertium based assamese morphological analyzer. *arXiv preprint arXiv:1503.03989*.

Rahman, M. U. (2010). Towards sindhi corpus construction. *Conference on Language and Technology, Lahore, Pakistan*.

Ranta, A. (2009). Gf: A multilingual grammar formalism. *Language and Linguistics Compass*, 3.

Rinju, O., Rajeev, R., and Sherly, E. (2013). Morphological analyzer for malayalam: Probabilistic method vs rule based method.

Vinod, P., Jayan, V., Bhadran, V., and Thiruvananthapuram, C. (2012). Implementation of malayalam morphological analyzer based on hybrid approach. *ROCLING XXIV (2012)*, page 307.