Research Article

Shihadeh Alqrainy* and Muhammed Alawairdhi

Towards Developing a Comprehensive Tag Set for the Arabic Language

Abstract: This paper presents a comprehensive Tag set as a fundamental component for developing an automated Word Class/Part-of-Speech (**PoS**) tagging system for the Arabic language. The aim is to develop a standard and comprehensive PoS tag set that based upon PoS classes and Arabic inflectional morphology useful for Linguistics and Natural Language Processing (NLP) developers to extract more linguistic information from it. The tag names in the developed tag set uses terminology from Arabic tradition grammar rather than English grammar. The usability of the presented Tag set has been tested in manual tagging and built up a set of tagged text to serve as a goal corpus used to compare it with the results obtained from the tagger. The tagger has achieved an average accuracy of 90% using the developed detailed tag set.

Keywords: Arabic Language, Part-of-Speech (PoS), Tag set, Natural Language Processing (NLP)

2010 Mathematics Subject Classification: 68-XX

1 Introduction

Whatever the natural language, its words are classified into grammatical categories called word class/Part-of-Speech (PoS), such as Name, Verb and Particle. The set of all classes is called a PoS tag set, where these sets are used in the PoS tagging process which is a crucial part of any tagging system that gives a good explanation to any tagged corpus [1]. A tag process entails assigning a symbol attached to each word that indicates what part of speech a word is [2]. The PoS tag set is a set of labels or symbols that can be used to describe the words in any giving text [3]. This paper focuses on how to developed a standard and comprehensive PoS Tag set to be valid for any PoS tagging system for Arabic regardless of the technique the PoS tagging system was built.

The main task of any PoS tagging system is corpus linguistic. It is also an extremely necessary step and an important practical problem with potential NLP applications in many areas such as: Information Retrieval, Parsing System, Word Processing, Speech Synthesis System, Machine Translation and Building Dictionaries. However, the PoS Tag set that the PoS tagging system will use must be valid for any purpose for which the PoS tagging system is built. This paper does not delve into describing the techniques of PoS tagging process. Instead, the focus is simply on presenting a comprehensive PoS Tag set as a fundamental component for developing an automated Word Class/Part-of-Speech (PoS) tagging system for the Arabic language.

The paper is organized as follows: in Section 2, a background information regarding developing PoS tag set for Arabic are presented. We illustrate Part-of-Speech Tag Set Criteria in Section 3. In Section 4, the proposed method of designing the developed PoS Tag set is described. we present the usability test of the developed the tag set via experiment in Section 5. We conclude this paper in Section 6.

*Corresponding Author: Shihadeh Alqrainy:** Faculty of Artificial Intelligence, Al-Balqa Applied University, Jordan; Email: alqrainy@bau.edu.jo; alqrainy62@gmail.com
Muhammed Alawairdhi: College of Computing and Informatics, Saudi Electronic University; Email: malawairdhi@seu.edu.sa

2 Related Work

The current literature shows many attempts of presenting and developing a PoS tag set for Arabic to use *in PoS tagging systems the authors presented.*

El-Kareh and Al-Ansary [4] defined the tag set for the Arabic language, which contains three verbs, forty-six nouns, and twenty-three particles.

Khoja *et al.* [5] defines tag set for the Arabic language, which contains a hundred and seventy-seven tags, as follows: fifty-seven verbs, one hundred and three nouns, nine Particles, seven residuals and one punctuation.

Alshamsi and Guessom [6] defined a tag set for the Arabic language and they used it for their Hidden Marcove Model POS tagger system, which contains fifty-five tags.

Gharaibeh and N Gharaibeh [12] used a tag set presented in [5] to extract Arabic Noun Phrase to build their system using information retrieval techniques

Ababou and Mazroui [14] presented a tag set contains eighty-two tag proposed by the Alkhalil_Morpho_Sys analyzer [15]. Most of the tags belong to Particle class and not based upon inflectional feature of the word. The tags were built to show the grammatical arrangement of words (Syntax System).

Albared *et al.* [16] presented a tag set contains twenty-four tags. However, their tag set not cover all the sub-categories of the three major grammatical category/PoS class of the original Arabic word, Verb, Noun and Particle.

Zerouala *et al.* [17] presented a tag set has hundred-ten basic tags classified into Hierarchical levels of: Noun categories and their tags, Verb categories and their tags and Particle categories and their tags.

Hadni *et al.* [18] used thirty-two tag set extracted from Quranic tag set web site [19] to tagged Quranic and Kalimat Corpus.

Alian,and Awajan [20] presented a significant part of the work has been undertaken in the area of Arabic PoS tag set.

All these tag sets have been developed for different purposes based upon the PoS tagger own objective. However, the adaption of the above tag sets is problematic for Semitic languages as [17] claimed. "The majority of currently used tag sets are derived from English, which is a drawback for a morphologically complex language such as Arabic, namely those recommended by the Expert Advisory Group on Language Engineering Standards(EAGLES), are designed for Indo-European languages". The developed tag set in this work follows Arabic tradition grammar.

3 Part-of-Speech Tag Set Criteria

The tag set developer should take into considerations while designing the tag set a certain number of criteria; these criteria have been presented in more detail in [8]. The following subsections describe these criteria in a little bit of details.

3.1 Mnemonic Tag names

The developer must make the name of the tag easy for the user to remember. For example, [**Ve**] for a verb, [**Nu**] for noun and [**Pr**] for particle.

3.2 Fundamental of linguistic theory

The characteristics of the language and the theory of that language should be covered. For Example, the inflectional features of the Arabic language such as gender, number, person, etc.

3.3 Classification by form or function

The developer should have mentioned the paradigmatic forms (representative set of the inflections of a verb, noun, etc.), and syntagmatic functions (a syntactic function of the words) of the language. For example, vowels and other diacritical marks in the Arabic Language.

3.4 Idiosyncratic words

Most languages have special idiosyncratic behavior. Arabic like any other language has words does not fit into the POS tagger. For example, words belong to Particle class.

3.5 Categorization problem

The tag should be clearly defined and unambiguously. For example, Brown and Lancaster-Oslo/Bergen (LOB) tag sets (for English) analyzed "a" as article tag AT, but UPenn tag set analyzed it as determiner DT.

3.6 Tokenization issues: what counts as a word?

The tokenization process is also needed in the Arabic language to split not each word in the text but also the punctuation marks.

3.7 Multi-word lexical items

Unlike other languages, the Arabic language does not have many multi-words lexical items; most of these words appear in proper nouns and treated as one word.

3.8 Target users and/or application

Developing a tag set to create noun groups for building search engine differ than a tag set for POS tagger produced tagged corpora for education purpose.

3.9 Availability and/or adaptability of the tagger system

The tag set should develop to be compatible with the target language grammar. On another word, it covers all the features of the language. The proposed tag set compatible with Arabic grammar.

3.10 Adherence to standards

Languages are different in structure, features, linguistic attributes and order of the constituents within the sentence. The developers aimed to develop a standardize tag set.

3.11 Genre, register or type of languages

The text may either written text or spoken. Based on the type of text, the developer must take this criterion into account while developing the tag set.

3.12 Degree of the delicacy of the tag set

The tag set must have developed at a very high level of granularity. All the subclasses in the main grammar categories and the inflectional features should have covered.

4 Proposed Method - ARBTAGS Design and analysis

This section describes the proposed method of developing our Arabic Language PoS Tags (Abb: ARBTAGS). The initial version of this tag set design has been published and can be seen in [7]. In this paper, we extend and complete that design. The PoS tag set is based on tradition of Arabic grammar. The proposed PoS Tag set hierarchy is shown in Figure 1. Throughout this section, abbreviation symbols used to represent the name of each PoS class and sub-class in addition to a value of the inflectional features which have been used to represent the PoS tag in the proposed tag set is shown between square brackets.

ARBTAGS contains a hundred and sixty-one *detailed* PoS tags, a hundred and one nouns, fifty verbs, nine particles, one punctuation; the tags are augmented with inflectional features information such as gender, number, person, case, and mood. In addition, it contains twenty-seven PoS *general* tags covering the sub-classes of the three main PoS classes (**Verb**, **Noun**, and **Particle**). Verb class was categorized into three sub-classes, Noun class was categorized into sixteen sub-classes and particle class was categorized into seven sub-classes. One for Punctuation [**Pun**]. Another tag that used to represent the Arabized words [**Fw**] which consider a foreign word, were added to the above tags, to get a total of twenty-eight tags. These overall tags used to represent main classes and sub-classes in Arabic language with no inflectional features.

Despite the work done by Khoja *et al.* [5] as the author(s) described that their PoS Tag set has a hundred and seventy-seven tags which consider the largest PoS Tag set in the literature, but still subclasses belong to Noun such as verbal, diminutive, instrument, noun of place, noun of time and subclasses belong to Particle
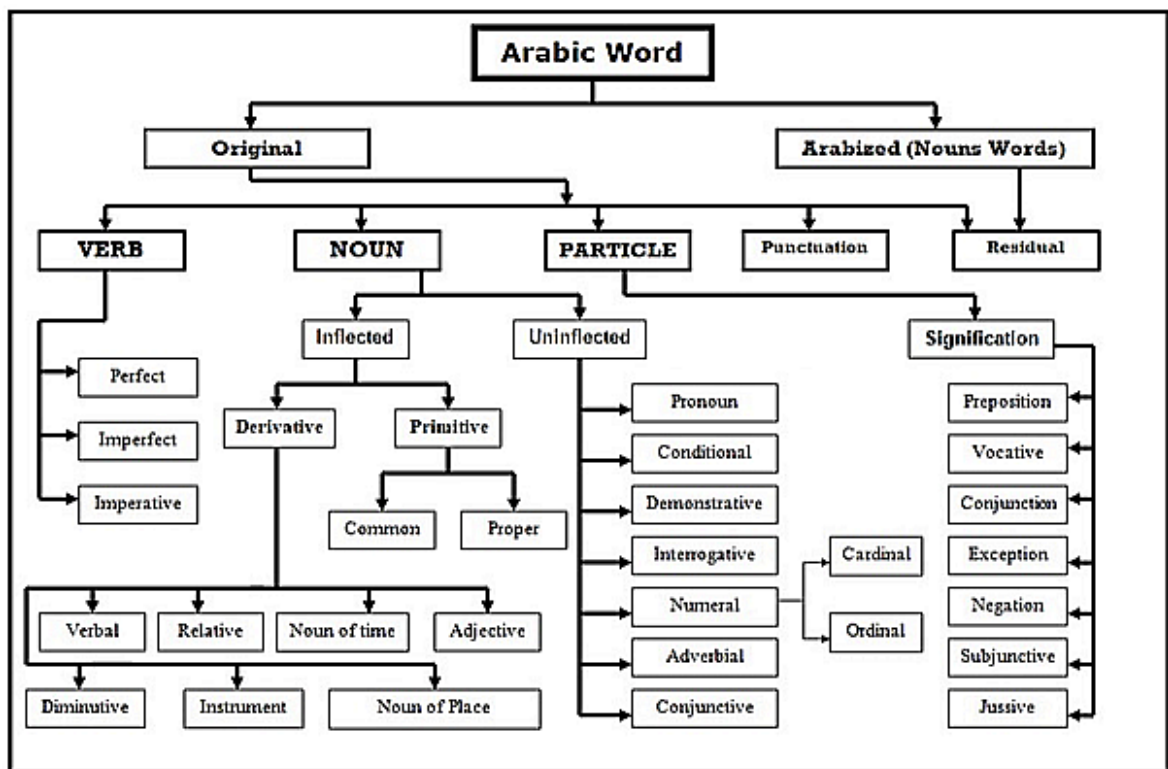

Figure 1: ARABTAGS Tag Set Hierarchy

such as vocative, subjunctive and jussive have not been mentioned. The reason is, they used terminology from English grammar rather than Arabic tradition grammar. As [8] point out, their PoS tag set came from the Lancaster University Centre for Computer Corpus Research on Language (URCEL) tradition of corpus linguistic; the authors were influenced by the English tag sets such as Constituent Likelihood Automatic Word-tagging System (CLAWS) heritage of tag set for Lancaster-Oslo/Bergen (LOB) corpora [8, 13].

The tag names in the developed tag set used terminology from Arabic tradition rather than English grammar. The tags were developed with a good level of granularity, where each tag is developed to include the inflectional features that meet the need of linguists and NLP developers.

An important feature of the Arabic language is that it is characterized by having an extensive morphological system as well as an inflectional system, so the tag set should be richly articulated, providing distinct coding for all classes of Arabic words. Therefore, using an earlier tag set will not capture all the sub-classes shown in the ARBTAGS tag set shown in Figure 1. The ARBTAG hierarchy based on important references in Arabic grammar, such as the Lexicon of Arabic Language grammar in tables and tablets [21] and Tatbiq Al-Nahwi by Rajhi [22]. We also collaborated with experts in this field.

The three major grammatical category/PoS class in Arabic language, Verb [**Ve**], Noun [**Nu**] and Particle [**Pr**]. Verb defined as a word used to denote an action that may be combined with particle. Noun defined as a word used to denote an essence that may be combined with the article "اَل" 'the' as a prefix [9].

As with other Semitic languages, Arabic is deficient in tenses [10]. Unlike Indo-European languages, the tenses in Arabic do not have accurate time significances. Table 1 describes the tenses of the class verb.

Table 1: The tense of class Verb

Tense	Tag
Perfect	[Pe]
Imperfect	[Pi]
Imperative	[Pm]

Table 2: The sub-categories of class Noun

Sub-category	Tag
Proper	[Po]
Infinitive	[If]
Instrument	[Is]
Pronoun	[Ps]
Demonstrative	[De]
Numeral	[Nn]
Common	[Cn]
Relative	[Re]
Noun of place	[Pn]
Conjunctive	[Cv]
Interrogative	[In]
Adjective	[Aj]
Diminutive	[Dm]
Noun of time	[Tn]
Conditional	[Cd]
Adverb	[Ad]

Table 3: The sub-categories of class Particle

Sub-category	Tag
Preposition	[Pp]
Conjunction	[Co]
Vocative	[Vo]
Annulment	[An]
Exception	[Ex]
Subjunctive	[Sb]
Elision	[Jv]

Table 4: The inflectional features of Arabic words

Inflectional Feature	Description	Tag	Used for
Gender	Masculine	[Ma]	
	Feminine	[Fe]	
	Neuter	[Ne]	
Number	Singular	[Sn]	both Verb and Noun
	Dual	[Du]	
	Plural	[Pl]	
Person	First	[Fs]	
	Second	[Sc]	
	Third	[Th]	
Case	Nominative	[Nm]	
	Accusative	[Ac]	
	Genitive	[Ge]	Class Noun
State	Definite	[Df]	
	Indefinite	[Id]	
Mood	Indicative	[Dc]	
	Subjunctive	[Sj]	Class Verb
	Jussive	[Js]	

In Arabic, verb tenses generally denote the time of action. The action completed in the perfect tense.

The prefixes and suffixes play a great role in determining the time of the verb. Therefore, the present/Imperfect verb is differing from past/perfect based on that affixes. The present or imperfect is used with actions that are still in progress or with repeated actions. The Imperative verb considered a modification of the Imperfect verb [10].

In the Arabic system, from every Arabic verb, many forms of words belong to noun categories, such as instrument noun, noun of place, adjective-noun, and noun of time, diminutive noun and verbal noun may be derived. The subcategories of the class Noun included also adjectives, pronouns; adverb, etc. are shown in Table 2.

The subcategories of the class Particle are shown in Table 3.

Inflection is a grammatical morpheme; it is the marking of the written word to add grammatical information such as gender, number, person, case and mood. Inflections do not change the meaning of the word. Arabic as the biggest member of Semitic languages exhibits not only a complex morphological system but also a richly inflected one. Derivational and Inflectional constitute the main categories of the Arabic morphological system.

In Arabic, inflectional features such as gender, number, and person are used for both main classes Verb and Noun, while case is used for class Noun. Also, the inflectional feature mood is used for class Verb. The tag of each inflectional feature and the description as well are shown in Table 4.

5 Experimental Results and Analysis

Arabic Morph Syntactic Tagger (AMT) used Rule-based and Pattern techniques was implemented to highlight the use of the developed tag set via various experiments. Figure 2 shows a snapshot of the AMT main screen.

Figure 2: AMT system main screen

In this regard, the usability of ARBTAGS has been tested in manual tagging and built up a set of tagged text to serve as a goal corpus used to compare it with the results obtained from the AMT tagger.

The performance of AMT system was tested on data extracted and collected from the official site of ministry of Education-Jordan, with a permission and authorization from the department of curricula and text-books management. The small corpus used in the current work consisting of 11,528 words; not limited to a particular domain; it covers wide range of topics such as scientific topics and literary topics. This testing data spread across three sets; set-1: consists of 3518 words represent an article extracted from the book of computer science for class 12 of secondary level. Set-2: consists of 3988 words represent an article extracted from the book of Arabic language for class 10 of elementary level and set-3: consists of 4022 words represent two articles extracted from the book of literary for class 11 of secondary level. A pattern, namely, correctness used to evaluate the performance of the developed system. Correctness is the ratio of number of words that tagged correctly to the number of words present in the text.

Table 5: Results obtained from AMT system

Data Sets	# of words	Tagged Correctly	Percentage
Set 1	3518	2885	82%
Set 2	3988	3886	97%
Set 3	4022	3700	92%

The domain of each data set mentioned in Table 5 affect the percentage of correctly tagged words. For example, the text in data set 1 is related to computer science topic; some words belong to Arabized words (e.g كَمِبيُتَر meaning "Computer"); these words not original Arabic words, and do not have a pattern/root from one hand and no Arabic grammatical rule may apply to it from other hand. Unlike data set 1, data set 2 comes from Arabic language topic and specified for elementary school level, it is an expected result. In addition, the percentage of correctly tagged words that were verbs or punctuation is higher. The different subject of data set 3 which contains many of proper nouns and irregular verbs leads to low accuracy. These are not recognized and tagged correctly.

PoS tag may be very coarse (e.g: **Ve** "Verb") or very fine (e.g: **VePiMaPlFsSj** "Verb, Imperfect, Masculine, Plural, First Person, Subjunctive"), depending on the task or application [11]. Since the main aim of AMT system is to produce a tagged corpus, tags were developed with a good level of granularity, inflectional features were added to each tag, that satisfied what linguists and NLP developers need. The tags are valid either general tags (**Nu**, **Ve** and **Pr**) or detailed tags (e.g: **VePiMaPlFsSj)** . Using the developed tag set, a word "يَكتُبن", yktboon meaning +"they writing" will be tagged as **VePiMaPlFsSj**, which means [Imperative verb, masculine gender, plural number, third person, subjunctive mood], as a fine or detailed tag while the general or coarse tag is **Ve**, depend on the aim of the POS tagger system.

Our experiments confirm that the implementation phase of the developed tag set in the PoS tagging process using AMT system is satisfactory. The developed tag set is valid for any PoS tagger purpose. For example, a tagger, that it aims is to extract noun groups useful for information retrieval system needs very coarse tag/general tag (e.g: **Ve**, **Nu**, and **Pr**). Furthermore, a tagger which has been used to produce a tagged corpus useful for developing an educational system needs very fine/detailed tags (e.g: **VePiMaPlFsSj**).

The significant point of the developed PoS tag set is that, it is easier for the linguistic and NLP developer to extract not only the three major grammatical category/POS class in Arabic language, Verb [Ve], Noun [Nu] and Particle [Pr] as general tags, but also the inflectional feature of the tagged text produced by the PoS tagger system depend on their PoS tagging system purpose. For example, they may extract all the words from the tagged corpus related to Verb class and at the same time the inflectional feature is Masculine gender by using **VeMa** parameter in their search system and so on.

Furthermore, the developed PoS tag set is valid regardless the type of the Arabic standard modern text or classical text may use either vocalized or un-vocalized since the researcher may pick the inflectional feature of the PoS tag he/she need depend on his/her PoS tagger system purpose. However, ATM presented in this paper aims to produce a tagged corpus. A sample of general PoS tags shown in Appendix A, while Appendix B shows a sample of detailed PoS tags.

6 Conclusions and Future Work

A comprehensive tag set was developed in this paper depends on ancient Arabic grammar. The developed tag set is based on the Arabic system of inflectional morphology. The tag set does not follow the traditional Indo-European tag set that is based on Latin, but instead, it's based on the Semitic tradition of analyzing language. These tags contain a large amount of information and add more linguistic attributes to the word. The presented tag set contains a hundred and sixty-one detailed tags and twenty-eight general tags that cover an Arabic language major POS classes and sub-classes that have been compiled and introduced in this paper.

We aimed to conduct further tests on more interesting dataset to evaluate the real performance of the developed PoS tag set. In addition, the lack of the Arabic resources such as tagged corpora constitute the main challenge involved in constructing any NLP system for Arabic. We aimed to offer a huge tagged corpus. Therefore, building a lexicon contains all Arabic roots to enhance the performance of the tagger system and to employ the developed tag set is extremely necessary.

References

[1]   S. Khoja, Apt: Arabic part-of-speech tagger, in: Proceedings of the Student Workshop at the Second Meeting of (NAACL2001), Carnegie Mellon University, Pittsburgh, Pennsylvania.

[2]   V. Halteren, Syntactic Word class Tagging, Kluwer Academic Publishers, Vol 9,(1999), The Netherlands.

[3]   A Hardie, Developing a tag set for automated part-of-speech tagging in Urdu, In Proceedings of the Corpus Linguistics (2003) conference, Lancaster University, UK.

[4]   El-Kareh and Al-Ansary, An Arabic interactive multi-feature PoS tagger, In Proceeding of the international conference on Artificial and Computational intelligence for Decision Control and Automation in engineering and Industrial Application (ACIDCA 2000) conference, Tunisia.

[5]   Khojah, Graside, and Knowels, A tagset for the morphosyntactic tagging of Arabic, In presented at Corpus Linguistics (2001), Lancaster University, UK.

[6]   F Shamsi and A Guessoum, A hidden Markov model-based PoS tagger for Arabic, In JADT (2006) - 8th International Conference on the Statistical Analysis of Textual Data, France.

[7]   S Alqrainy and A Ayesh, Developing a tag set for automated PoS tagging in Arabic, WSEAS Transactions on Computers 5, 11 (2006), pp 2787–2792.

[8]   E Atwell, "Development of tag sets for part-of-speech tagging", In AnkeLudeling & Merja Kyto, editor, Corpus Linguistics: An International Handbook. Mouton de Gruyte, (2007).

[9]   K Versteegh, The Arabic Language", (2001), Edinburgh University Press.

[10] Haywood and Nahmad, A new Arabic grammar: of the written language, (2005), Lund Humphries, USA.

[11] Prez-Ortiz, and Forcada., Part-of-speech tagging with recurrent neural networks, In Proceedings of the International Joint Conference on Neural Networks, IJCNN (2001).

[12] I Gharaibeh and N Gharaibeh, Towards Arabic Noun Phrase Extractor (ANPE) using information retrieval techniques. Int. J. Softw Eng. (2012) 2, 36–42.

[13] S Alqrainy, A morphological-syntactical analysis approach for Arabic textual tagging. (2008), De Montfort University, UK.

[14] N. Ababou and A. Mazroui, A hybrid Arabic POS tagging for simple and compound morph syntactic tags, Int. J. Speech Technol. (2015), pp. 1-14.

[15] M Bebah et al., Alkhalil morpho sys. In 7th International computing conference in Arabic (2011).

[16] Albared et al., Developing a competitive HMM arabic POS tagger using small training corpora. Intelligent Information and Database Systems, Springer (2011), pp. 288-296.

[17] Zerouala et al., Towards a standard Part of Speech tagset for the Arabic language. Journal of King Saud University - Computer and Information Sciences. Volume 29, Issue 2, April (2017), Pages 171-178.

[18] Hadni et al., Hybrid Part-Of-Speech Tagger for Non-Vocalized Arabic Text. Int. J. Nat. Lang. Comput., 2 (2013), pp. 1-15.

[19] Quran Tagset: http://corpus.quran.com/documentation/tagset.jsp

[20] M Alian and A Awajan. Arabic Tag Sets: Review. Intelligent Systems and Applications, (2019), pp 592-606.

[21] A Al-Dahdah, The grammar of the Arabic language in tables and lists, (1989), Beirut Maktabat Lebnan Arab.

[22] A Rajhi, Tatbiq Al-Nahwi, (2000), Dar Annahda Al Arabia for printing, publishing and distribution.

Appendix A – Sample of general PoS tags

Tag	Dsecription
VePe	Perfect verb
VePi	Imperfect verb
VePm	Imperative verb
NuPo	Proper noun
NuCn	Common noun
NuAj	Adjective noun
NuIf	Infinitive noun
NuRe	Relative noun
NuDm	Diminutive noun
NuIs	Instrument noun
NuPn	noun of Place
NuTn	noun of Time
NuPs	Pronoun
NuCv	Conjunctive noun

Tag	Description
NuCd	Conditional noun
NuDe	Demonstrative noun
NuIn	Interrogative noun
NuAd	Adverb
NuNn	Numeral noun
Fw	Foreign noun
Pun	Punctuation mark
PrPp	Preposition
PrVo	Vocative Particle
PrCo	Conjunction Particle
PrEx	Exception Particle
PrAn	Annulment Particle
PrSb	Subjunctive Particle
PrJs	Jussive Particle

Appendix B – Sample of detailed PoS tags

Tag	Dsecription
VePeMaSnThSj	Verb, Perfect, Masculine, Singular, Third Person, Subjunctive
VePeMaSnFsDc	Verb, Perfect, Masculine, Singular, First Person, Indicative
VePeMaSnSeSj	Verb, Perfect, Masculine, Singular, First Person, Subjunctive
VePeFeSnSeJs	Verb, Perfect, Feminine, Singular, Second Person, Jussive
VePeFeSnThJs	Verb, Perfect, Feminine, Singular, Third Person, Jussive
VePeNeDuSeSj	Verb, Perfect, Neuter, Dual, Second Person, Subjunctive
VePeMaDuThSj	Verb, Perfect, Masculine, Dual, Third Person, Subjunctive
VePeFeDuThSj	Verb, Perfect, Feminine, Dual, Third Person, Subjunctive
VePeMaPlFsSj	Verb, Perfect, Masculine, Plural, First Person, Subjunctive
VePeMaPlSeJs	Verb, Perfect, Masculine, Plural, Second Person, Jussive
VePeFePlSeJs	Verb, Perfect, Feminine, Plural, Second Person, Subjunctive
VePeFePlThJs	Verb, Perfect, Feminine, Plural, Third Person, Subjunctive
VePeMaPlThDc	Verb, Perfect, Masculine, Plural, Third Person, Indicative

Tag	Description
NuPsMaPlThNmId	Personal Noun, Masculine, Plural, Third Person, Nominative, Indefinite
NuPsFeSnThAcId	Personal Noun, Feminine, Singular, Third Person, Accusative,Indefinite
NuPsFePlThAcId	Personal Noun, Feminine, Plural, Third Person, Accusative, Indefinite
NuPsMaSnSeAcId	Personal Noun, Masculine, Singular, Third Person, Accusative, Indefinite
NuPsNeDuSeAcId	Personal Noun, Neuter, Dual, Third Person, Accusative,Indefinite
NuPsMaPlSeNmId	Personal Noun, Masculine, Plural, Third Person, Nominative,Indefinite
NuPsFeSnSeGeId	Personal Noun, Feminine, Singular, Third Person, Genitive,Indefinite
NuPsFePlSeAcId	Personal Noun, Feminine, Plural, Third Person, Accusative,Indefinite
NuPsNeSnFsAcId	Personal Noun, Neuter, Singular, First Person, Accusative,Indefinite
NuPsNePlFsNmId	Personal Noun, Neuter, Plural, First Person, Nominative,Indefinite
NuDeSnAcId	Demonstrative Noun, Singular, Accusative,Indefinite
NuDeDuGeId	Demonstrative Noun, Dual, Genitive, Indefinite