Staged Approach for Grammatical Gender Identification of Nouns using Association Rule Mining and Classification

Shilpa Desai[1], Jyoti Pawar[1], and Pushpak Bhattacharyya[2]

[1] Department of Computer Science and Technology
Goa University, Goa - India
`sndesai@gmail.com`, `jyotidpawar@gmail.com`
[2] Department of Computer Science and Engineering
IIT-Bombay, Mumbai - India
`pb@iitb.ac.in`

Abstract. In some languages, gender is a grammatical property of the noun. Grammatical gender identification enhances machine translation of such languages. This paper reports a three staged approach for grammatical gender identification that makes use of word and morphological features only. A Morphological Analyzer is used to extract the morphological features. In stage one, association rule mining is used to obtain grammatical gender identification rules. Classification is used at the second stage to identify grammatical gender for nouns that are not covered by grammatical gender identification rules obtained in stage one. The third stage combines the results of the two stages to identify the gender. The staged approach has a better precision, recall and F-score compared to machine learning classifiers used on complete data set. The approach was tested on Konkani nouns extracted from the Konkani WordNet and an F-Score 0.84 was obtained.

1 Introduction

Gender is a grammatical property of nouns in many languages[3] including Indian language such as Sanskrit, Hindi, Gujarati, Marathi and Konkani. In such languages adjectives and verbs in a sentence agree with the gender of the noun. For example translation of "*He is a good boy*" and "*She is a good girl*" in Hindi is "*vaha eka achchhaa laDakaa haai*[4]" and "*vaha eka achchhii laDakii haai*", respectively. We see that adjective good is translated into "*achchhaa*" or "*achchhii*" based on the gender of the noun. Using a freely available and highly used Google Translate machine translation system, translation of "*It is a nice river*" to Hindi is "*yaha eka achchhaa nadai haai*". The grammatical gender

---

[3] List of languages with type of grammatical gender followed is listed at (http://en.wikipedia.org/wiki/List_of_languages_by_type_of_grammatical_genders)

[4] In the paper we have used transliteration for Hindi and Konkani examples that resembles ITRANS (http://www.aczoom.com/itrans/html/tblall/tblall.html).

of river i.e. "*nadai*" in Hindi is feminine hence "*achchhii*" should be used in place of "*achchhaa*" for the translation to be more appropriate. Thus, machine translation into any language with a grammatical gender will need to predict the grammatical gender of the noun to ensure agreement. Hence, determining grammatical gender of nouns in these languages is essential.

Features such as grammatical gender of nouns required for machine translation can be obtained from a lexical resource like WordNet. However all WordNets do not maintain gender of nouns. Many WordNets in Indian Languages namely Marathi, Konkani, Gujarati, Bengali, Sanskrit etc. have been created using the Expansion Approach[1] with Hindi WordNet as source. Hindi nouns have either *masculine* or *feminine* grammatical gender whereas Bengali nouns do not possess grammatical gender. Nouns in Marathi, Konkani, Gujarati and Sanskrit can have either *masculine* or *feminine* or *neuter* grammatical gender. Grammatical gender of nouns was not directly borrowed from the source WordNet as all language do not share the same grammatical gender. There is a need to augment the WordNet nouns with grammatical gender so that WordNet can be a valuable resource for machine translation, specially for languages which have an agreement system with another aspect of the language, such as adjectives, articles or verbs.

In this paper we present an automatic three stage gender identification approach using word and morphological features, which can be used to assign grammatical gender to nouns in a WordNet. Our method was tested for nouns in Konkani WordNet[5] . The main contributions of this work can be summarized as follows:

- Grammatical Gender Identification rules obtained using association rule mining.
- Classification model for grammatical gender identification where association rule found insufficient.
- New grammatical gender category called non-neuter[6]

The rest of the paper is organized as follows: Section 2 is devoted to related work. A description of grammatical gender identification is discussed in section 3. The staged approach used for grammatical gender identification is described in section 4. In Section 5, experimental results and evaluation are discussed. Section 6 concludes the paper.

2   Related Work

Related work for lexicon extension include semi-supervised approach used to predict lexical properties like grammatical gender for Wambaya[2], a low density language. Here co-occurrence frequencies with demonstratives and machine

---

[5] The Konkani WordNet can be accessed at
  http://konkaniwordnet.unigoa.ac.in/public/wordnet/wordnet.php?langid=19&id=2
[6] Discussed in section 3

learning using context windows are used to predict the lexical properties. Gender classification, for Dutch nouns, has been attempted for rule-based MT generation tasks[3]. In which information derived from large annotated monolingual corpora, a set of context-checking rules based on co-occurrence of nouns and determiners are used. An SVM based tool has been developed for Romanian[4] to distinguish neuter gender which uses n-gram features.

General sound of the language has been exploited to determine the grammatical gender for German and Romanian[5]. Noun's ending have been used as a strong clue for gender in French[6]. Cognitive aspects of grammatical gender for Italian nouns referring to animals have been tested[7]. Bootstrapping process which make use of word forms, determiners, quantifiers and adjectives to predict gender for nouns in context[8] have been carried out.

Other reported work on gender identification has been for anaphora resolution wherein gender information has been extracted from enhanced part-of-speech tagger[9]. Methods to classify all occurrences of nouns in a document using a wide variety of contextual, morphological and categorical gender features have been done[10]. These works are targeted towards biological gender identification and not grammatical gender.

Most of the work reported is for non-Indian languages and make use of determiners or gender-marked pronouns like his, her etc. Some of these methods are used for gender identification within a context. Other methods require manual preparation of annotated data like morphological suffix list with corresponding gender which is time consuming.

3   Grammatical Gender Identification

Noun-class[11] in linguistics refers to a particular category of nouns. Some natural languages categorize nouns into noun-classes based on either semantic criterion, morphology or some arbitrary convention. Noun-classes form a system of grammatical agreement. Some examples of common criteria that define noun classes include:

- animate vs. inanimate
- rational vs. non-rational
- human vs. non-human
- masculine vs. feminine
- masculine vs. feminine vs. neuter
- common vs. neuter

Grammatical gender[12] is a specific form of noun-class system. In languages with grammatical gender, every noun inherently carries one value (*masculine* or *feminine* or *neuter*). The values present in a given language are called the genders of that language[7]

---

[7] The grammatical gender for a noun mostly coincides with natural or biological gender but not always. For example "*chalii* (girl) in Konkani, "*Madchen* (girl) in German have grammatical gender as *neuter* and natural or biological gender as *feminine*

In Indian languages such as Sanskrit, Marathi, Konkani and Gujarati, the grammatical gender property of nouns is one of the gender values namely *Masculine*, *Feminine* or *Neuter*. Earlier profession names (job titles) like doctor, lawyer, professor, "*aadaogada* (advocate) etc were assigned *Masculine* gender because such professions were only practiced by men. However with changing times we see many women taking up such profession. Profession names now be used in either feminine or masculine sense and gender value can only be determined in the context of usage. For example "*Professor baraii shaikayata*" can be translated as "*The professor teaches well*" or "*She is a good professor*". Here the noun "*Professor*" refers to a female professor which we can induce from the adjective agreement "*baraii*" in the sentence. However, the example "*Professor barao shaikayata*", translated as "*The professor teaches well*" or "*He is a good professor*" where the noun "*Professor*" refers to a male professor which we can induce from "*barao*". As a result such profession names cannot be assigned *Masculine* gender in languages like Konkani. To reflect this changing behavior in language a new gender value namely *Non-Neuter* was proposed. WordNet for languages like Konkani do not maintain two different entries for male Professor and female Professor, in such cases it is more relevant to assign *Non-Neuter* as grammatical gender value.

Definition *Non-Neuter* : *Non-Neuter* is a grammatical gender value assigned to a noun when the noun gender value can be either *masculine* or *feminine* based on the context and cannot be determined out of context.

3.1 Problem Statement

Given a set of nouns $N = \{n_1, n_2, \ldots, n_k: n_i$ is Konkani noun$\}$ and a set $W = \{(w_1, f_1), (w_2, f_2), \ldots, (w_m, f_m): w_i \in$ Corpus C, f_i is the frequency count of w_i in $C\}$ the objective is to automatically identify appropriate grammatical gender for nouns in N and create set $N_G = \{(n_1, g_1), (n_2, g_2), \ldots, (n_k, g_k): n_i \in N, g_i \in G\}$ where set $G = \{ Masculine, Feminine, Neuter, Non\_Neuter\}$

```
Input    : Nouns
Output   : Gender label (Masculine, Feminine, Neuter or
           Non_Neuter).
Resources/tools used: Raw corpus, Morphological Analyzer and
                      WordNet.
Hypothesis: Morphological features along with word features can be
            used to identify grammatical gender of nouns.
```

To identify grammatical gender, we first identified all possible morphological and word features for nouns. We than prepared a data set based on these features. We describe the details of data set generation in the following sub-sections.

3.2 Data Set Generation

The features in the data set are grouped into two categories namely word features and morphological features. The morphological features are generated using

the Konkani Morphological Analyzer[13]. The desired output feature, gender, is manually assigned for each tuple in the data set. The features used are listed in Table 1.

Let set α and set β hold all possible values for features BeginVowel and EndVowel respectively. To generate our data set we followed the following data preprocessing steps:-

Step 1: Extract nouns from the WordNet.

Step 2: Compute word features listed in Table 1 for each noun. Convert Unicode based string features to ASCII based string features using a map table.[8]

Step 3: Compute set α and set β

Step 4: For nouns extracted with each feature in set α and set β, randomly pick ten nouns of each word length wherever possible.

Step 5: Prune noun entries corresponding to morphological variants.[9]

Step 6: For each noun, compute values for morphological features in Table 1 using Morphological Analyzer tool for the language in this case Konkani.

Step 7: For each noun, manually assign gender feature.

We have a total of 1264 nouns in the data set used and the same can be found on http://www.cicling.org/2015/data/119/. The data set was used to identify a model for gender identification using supervised machine learning techniques. We ran various classification algorithms on our data sets[10] using 66% as training set and the rest as test set to determine the best training model. The performance of machine learning classifiers on our data set is presented in the next subsection.

3.3 Performance of Machine Learning Classifiers

The performance of various types of machine learning classifiers were tabulated based on word features only, morphology features only and both word and morphology features. The results[11] obtained for each feature type are tabulated in Table 2.

Analysis of Results: The performance of the various classifiers indicate that word + morphological features give better classification results compared to word features or morphological features used alone. The *ZeroR* Rule-Based classifier, which simply outputs the most common class irrespective of the attributes, is used as baseline or lower bound to evaluate classifier performance. *Bayesian*

---

[8] A map table assign a roman transliteration code to every Unicode characters. This mapping was done to execute machine learning classifiers which expect ASCII strings

[9] We found that some entries in WordNet were plural forms of the noun which were pruned

[10] Such data sets used to select a learning model are some times referred to as development (dev) set.

[11] Here Precision (P), Recall (R) and F-score (F) are the weighted average values generated

Shilpa Desai, Jyoti Pawar, Pushpak Bhattacharyya

Table 1. Data Set Features.

Feature Name	Feature Description
Word Features	
Begin Vowel	The vowel attached to the letter with which the word begins. If the word does not have a vowel attached to first letter, then this feature is set to null.
End Vowel	The vowel if any with which the word ends. If the word does not end with a vowel, then this feature is set to null.
Length	The number of characters in the word.
Morphological Features	
PID	The matching paradigm identifier.
SOFS	The suffix attached to the noun base to get the singular oblique stem for the noun.
POFS	The suffix attached to the noun base to get the plural oblique stem for the noun.
PDCS	The suffix attached to the noun base to get the plural direct form of the noun.
FreqDSF	Number of times the direct singular form of the noun occurs in the corpus
FreqDPF	Number of times the direct plural form of the noun occurs in the corpus.
FreqSOF	Number of times the oblique singular form of the noun occurs in the corpus.
FreqPOF	Number of times the oblique plural form of the noun occurs in the corpus.
TotSOV	Total number of oblique singular forms of the noun occurring in the corpus.
TotalPOV	Total number of oblique plural forms of the noun occurring in the corpus.
SOVWin	Distinct number of types amongst oblique singular forms of the noun occurring in the corpus.
POVWin	Distinct number of types amongst oblique plural forms of the noun occurring in the corpus.
SR	Rule applied to obtain the noun stem.

Table 2. Performance of Machine Learning Classifiers.

	Features								
Algorithm	**Word**			**Morphological**			**Word + Morphological**		
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
Bayesian									
Naive Bayes	0.676	0.679	0.658	0.770	0.781	0.766	0.775	0.784	0.769
Bayes Net	0.672	0.677	0.656	0.769	0.786	0.774	0.781	0.802	0.788
Function									
Logistic	0.696	0.712	0.689	0.788	0.809	0.792	0.786	0.805	0.794
MultilayerPerceptron	0.694	0.705	0.693	0.804	0.812	0.798	0.756	0.763	0.752
RBFNetwork	0.653	0.670	0.646	0.776	0.784	0.765	0.765	0.781	0.759
SimpleLogistic	0.709	0.691	0.661	0.774	0.805	0.786	0.792	0.821	0.803
SMO	0.721	0.681	0.644	0.775	0.807	0.787	0.783	0.816	0.798
Instance-Based									
IB1	0.713	0.714	0.697	0.757	0.767	0.759	0.774	0.781	0.766
KStar	0.728	0.714	0.697	0.747	0.753	0.744	0.761	0.763	0.754
Ensemble									
AdaBoost	0.416	0.563	0.437	0.605	0.576	0.467	0.496	0.626	0.542
Bagging	0.702	0.723	0.702	0.802	0.805	0.791	0.773	0.802	0.787
Random Sub Space	0.720	0.698	0.668	0.761	0.795	0.775	0.796	0.828	**0.811**
Decorate	0.688	0.714	0.695	0.801	0.814	**0.800**	0.802	0.819	0.804
Logit Boost	0.712	0.695	0.667	0.760	0.788	0.769	0.811	0.819	0.803
Rule-Based									
PART Decision List	0.701	0.723	0.700	0.770	0.786	0.776	0.790	0.809	0.798
Ridor	0.678	0.572	0.583	0.720	0.740	0.727	0.779	0.802	0.788
ZeroR	0.248	0.498	0.331	0.248	0.498	0.331	0.239	0.488	0.320
Decision Tree									
Random forest	0.726	0.726	**0.708**	0.777	0.791	0.779	0.784	0.809	0.794
Logistic Model Tree	0.709	0.691	0.661	0.774	0.865	0.786	0.799	0.826	**0.811**
REPTree	0.707	0.723	0.702	0.807	0.809	0.792	0.781	0.814	0.795

classifiers use a probabilistic framework, *Instance based classifier*s use nearest neighbor, distance based method for solving classification problem. These perform better than *ZeroR* the baseline used but don't provide the best learning model for the data. The Function based classifiers like *Logistic Regression, Multilayer Perceptron* fit a function on the attribute vectors in the training data to determine the class. They have been used as a good training model for many applications. However for grammatical gender identification, Logistic Model Tree classifier that combines logistic regression and decision tree learning performs better and is more appropriate learning model. Ensemble learning method namely Random Subspace which is a decision tree based classifier has a performance similar to Logistic Model Tree. These observations suggest that a combination of classifier provide a better model for grammatical gender identification. One amongst *Random subspace* and *Logistic Model Tree* can be used as a supervised machine learning model to identify grammatical gender with a F-Score of 0.811.

We were curious to know if we could extract some rules from our data which could be directly used to determine grammatical gender, and if so what is the precision or confidence level of such rules generated. Grammar books[14] and linguistic resources[15][16] available for the Konkani language do not explicitly specify any grammatical gender identification rules but suggest possibility of such rules. We wanted to check if the rules suggested and any new rule were generated by our data set and if so the precision or confidence level associated with such a rule. This motivated us to use a three stage grammatical gender identification approach described in the next section.

4   Staged Grammatical Gender Identification

4.1   Approach

A three stage approach was used wherein stage one used association rule mining to obtain a list of grammatical gender identification rules. Stage two used machine learning classifiers to identify grammatical gender of nouns that could not be identified in stage one using grammatical gender identification rules. In stage three the rules generated and best classification model obtained were combined to calculate the performance of the staged approach on a separate test data set with 200 entries. We describe the three stages used in the following sub-sections.

4.2   Stage One: Association Rule Mining

Predictive apriori association mining algorithm available in knowledge analysis tool WEKA 3.6.11 was used to generate gender identification rules. Three different experiments were conducted and rules were generated based on

1. Word features only
2. Morphological features only
3. Both word and morphological features.

For each type of feature multiple grammatical gender identification rules were obtained. Only those rules which had more than 0.95 confidence were accepted. Table 3 tabulates the number of rules extracted and some sample rules for each feature type. The rules obtained when both word and morphological features

Table 3. Grammatical Gender Identification Rules obtained.

Features Used	# Grammatical Gender Rules	Sample Grammatical Gender Identification Rules Obtained
Word	13	$EndVowel = o \rightarrow Gender = masculine$ acc:(0.99498) $EndVowel = ehm \rightarrow Gender = neuter$ acc:(0.99497) $BeginVowel = se\ EndVowel = a \rightarrow Gender = feminine$ acc:(0.99454) ......
Morphological	10	$PID = 20 \rightarrow Gender = masculine$ acc:(0.99496) $PID = 19 \rightarrow Gender = neuter$ acc:(0.99493) $PID = 13 \rightarrow Gender = feminine$ acc:(0.99476) ......
Word + Morphological	35	$PID = 13 \rightarrow EndVowel = be\ Gender = feminine$ acc:(0.99469) $PID = 9 \rightarrow EndVowel = a\ Gender = feminine$ acc:(0.99433) ......

were used had some rules which were same as rules obtained using only word features and only morphological features. In addition some rules got combined to give more refined rules. Repetitive rules and refined rules were pruned and 16 rules were retained and implemented as Gender Identification Rule (GR).

The main contribution of using association rule mining is the 16 gender identification rules obtained with confidence above 0.95 for each rule. These rules alone can be used to determine gender of 49.76% of entries in the data set with a precision of 0.97. The grammatical gender rules were not sufficient to identify grammatical gender for the entire data set. A substantial number of nouns remained unassigned. Hence original data set was reduced to create a new data set which held only those entries which cannot be assigned a grammatical gender using a rule. This new data set created is referred to as *Reduced Data Set* (*RDS*) The best model for *Reduced Data Set* (*RDS*) was picked by training various machine learning classifiers on reduced data set.

4.3 Stage Two: Classification

The reduced data set had 635 entries for which gender could not be determined using association rule mining. To decide which classification model will best represent our data, various classification algorithms available in WEKA 3.6.11 were executed. Table 4 gives the best performing classifications algorithm. Table 4

Table 4. Classification Algorithm Results.

Features	Algorithm	Precision	Recal	F-Score
	IB1	0.768	0.758	0.692
Word	LogitBoost	0.528	0.726	0.611
	MultilayerPerceptron	0.657	0.705	0.615
	Random Forest	0.612	0.747	0.63
	IB1	0.598	0.579	0.588
Morphological	LogitBoost	0.698	0.695	0.697
	MultilayerPerceptron	0.589	0.579	0.584
	Random Forest	0.669	0.684	0.676
	IB1	0.705	0.705	**0.705**
Word+Morphological	LogitBoost	0.528	0.726	0.611
	MultilayerPerceptron	0.734	0.716	**0.723**
	Random Forest	0.72	0.747	**0.712**

indicates that word and morphological features used together provide a better classification model with best F-score of 0.723[12] obtained by Multi-layer Perceptron. This neural network function based classifier works well with both categorical and continuous attributes, performs well on the test set and hence is chosen as a learning model for training.

4.4 Stage Three: Staged Approach

The rules obtained in stage one and training model namely multilayer perceptron obtained in stage two were combined into a staged approach. The pseudo-code of the algorithm used by the staged approach to assign grammatical gender to noun entries in test data set is illustrated in Figure 1.

The experimental results obtained and evaluation are presented in the next section.

5 Experimental Results and Evaluation

The goal of the experiment was to build a grammatical gender identification system. Two approaches namely machine learning classifiers approach and staged

---

[12] Here Precision, Recall and F-score are the weighted average values generated

---

Algorithm 1: Assign Grammatical Gender to Nouns in WordNet

```
for each noun in WordNet
   if noun compatible with grammatical gender rule GR
      gender = GR(noun);
   else
      gender = MultilayerPerceptronClassifier(noun);
   end if
end for
```

Fig. 1. Algorithm for assigning grammatical gender to nouns in WordNet.

approach were used for the same. A separate test data set with 200 entries was created to test the performance of the approaches. The staged approach algorithm presented in Figure 1 was tested on the test data set to determine its performance. Table 5 tabulates the results obtained using staged approach. Table 6 compares the staged grammatical gender identification approach with

Table 5. Grammatical Gender Identification using Staged Approach.

Grammatical Gender	True Positives	False Positives	False Negatives	Precision	Recall	F-Score
Feminine	91	5	4	0.95	0.96	0.95
Masculine	50	11	15	0.82	0.77	0.79
Neuter	31	12	7	0.72	0.82	0.77
Non-neuter	0	0	2	0	0	0

best performing machine learning classifier. A minor improvement is seen in the weighted average values of precision, recall and F-score in the staged approach. The results obtained were analyzed and are presented in the next subsection.

Table 6. Staged Approach Vs Machine Learning Classifiers.

Algorithm	Precision	Recall	F-Score
Staged Approach	0.854	0.86	**0.84**
Random Sub Space	0.78	0.807	0.79
Logistic Model Tree	0.807	0.837	0.82

5.1 Analysis of Results

To find out where the algorithm fails, the results obtained were analyzed from two perspectives namely association rule mining and classification model.

Association Rule Mining Analysis: Analysis of grammatical gender identification rules generated by association rule mining indicated that the feminine gender is identified using these rules . As expected the precision and recall for the feminine gender is high as can be seen in Table 5. An example where assignment of feminine gender fails is for the Konkani word "*shaeta*" (goldsmith) which has grammatical gender masculine but is assigned as feminine. This is an exceptional case wherein its morphological features resemble those of a feminine noun. Another example of failure is word like *baenagalaii* (Bengali language/Bengali person) which has multiple senses wherein the different senses have different gender namely in the Bengali language sense the noun is feminine and in the Bengali person sense the noun is masculine. The algorithm fails for such cases. A few instances of masculine and neuter gender can be identified using rules, but most of masculine and few neuter gender nouns cannot be identified based on rules. There are no rules generated corresponding to non-neuter gender.

Classification Model: The classification model has been used mainly to identify masculine, neuter and non-neuter genders. This word and morphological features used can classify nouns in the above stated gender values with a best F-Score of 0.72. Additional features like context based features may be used to improve the performance of such classifiers.

6 Conclusion

Grammatical gender of a noun in Konkani can be determined based on morphological and word features using a staged approach, with a weighted average F-Score of 0.84 thus supporting our hypothesis. On examination of grammatical gender identification rules we found that some morphological paradigms suggest grammatical gender. For some words like lawyer two paradigms get assigned suggesting that the noun has non-neuter grammatical gender. Some association rules example $BeginVowel = NULL\ EndVowel = o \rightarrow PID = twozero\ Gender = masculine$ acc:(0.99315) generated can also be used to automatically assign morphological paradigms to new noun entries based on word features. Future work will be focused on finding word, morphological or context features which give more clues to determine masculine and neuter gender.

Acknowledgments

provided to clarify some of the doubts with respect to Konkani Language. We also acknowledge that we were able to carry out the work using the Konkani WordNet which was developed as part of the Indradhanush WordNet Project funded by Department of electronics and information technologY (DeitY), New Delhi 11(13)/2010-HCC(TDIL) dated 03/08/2010

References

1. Bhattacharyya, P.: Indowordnet. In: Proceedings of Seventh International Conference on Language Resources and Evaluation (LREC), Malta (2010) 3785–3792
2. Nicholson, J., Nordlinger, R., Baldwin, T.: Deep lexical acquisition of type properties in low-resource languages: A case study in wambaya. In: Proceedings of the 26th Pacific Asia Conference on Language,Information and Computation. (2012) 75–81
3. Babych, B., Geiger, J., Rosell, M., Eberle, K.: Deriving de/het gender classification for dutch nouns for rule-based mt generation tasks. In: Proceedings of the 3rd Workshop on Hybrid Approaches to Translation (HyTra) @ EACL 2014, Gothenburg, Sweden (2014) 75–81
4. Dinu, L., Niculae, V., Sulea, M.: Dealing with the grey sheep of the romanian gender system, the neuter. In: Proceedings of COLING 2012: Demonstration Papers, Mumbai, India (2012) 119–124
5. Nastase, V., Popescu, M.: Whats in a name? in some languages, grammatical gender. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore (2009) 1368–1377
6. Spalek, K., Franck, J., Schriefers, H., Frauenfelder, U.: Phonological regularities and grammatical gender retrieval in spoken word recognition and word production. Journal of Psycholinguistic Research **37** (2008) 419–442
7. Vigliocco, G., Vinson, D., Paganelli, F.: Grammatical gender and meaning. In: Proceedings of the 26thMeeting of the Cognitive Science Society. (2004)
8. Cucerzan, S., Yarowsky, D.: Minimally supervised induction of grammatical gender. In: Proceedings of HLT-NAACL, Edmonton (2003) 40–47
9. Kennedy, C., Boguraev, B.: Anaphora for everyone: Pronominal anaphora resolution without a parser. In: Proceedings of COLING, Boulder,Colorado (1996) 113–118
10. Bergsma, S., Lin, D., Goebel, R.: Glen, glenda, or glendale: Unsupervised and semi-supervised learning of english noun gender. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL), Boulder,Colorado (2009) 120–128
11. : Noun class (2015) "http://en.wikipedia.org/wiki/Noun_class".
12. : Grammatical gender (2015) "http://en.wikipedia.org/wiki/Grammatical_gender".
13. Desai, S., Desai, N., Pawar, J., Bhattacharya, P.: Autoparse: An automatic paradigm selector for nouns in konkani. In: Proceedings of the 11th International Conference on Natural Language Processing (ICON), Goa,India (2014) 154–162
14. Borkar, S. In: Konkani Vyakran. Konkani Bhasha Mandal (1992)
15. Almeida, M. In: A Description of Konkani. Thomas Stephens Konknni Kendr (1989)
16. Sardessai, M. In: Some Aspects of Konkani Grammar. Unpublished M.Phil. Thesis (1986)