Learning rules for morphological analysis and synthesis of Macedonian nouns, adjectives and verbs

Aneta Ivanovska*, Katerina Zdravkova†, Tomaž Erjavec*, Sašo Džeroski*

* Department of Knowledge Technologies, Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
{aneta.ivanovska, tomaz.erjavec, saso.dzeroski}@ijs.si

†Institute of Informatics, Faculty of Natural Sciences and Mathematics
Arhimedova 5, 1000 Skopje, Macedonia
keti@ii.edu.mk

Abstract

This paper presents a machine learning approach to morphological analysis and synthesis of Macedonian nouns, adjectives and verbs. An inductive logic programming (ILP) system, Clog, was used to learn the inflectional paradigms of Macedonian words. Clog learns first-order decision lists, i.e., ordered sets of rules. Training and testing of the rules was performed on the words originating from Orwell's "1984". High accuracies (over 90%) were achieved, which are encouraging both for further research in annotation of Macedonian language resources as well as practical use (eg. in web search engines).

Učenje pravil za oblikoslovno analizo in sintezo makedonskih samostalnikov, pridevnikov in glagolov

V prispevku predstavimo metodo za strojno učenje oblikoslovne analize in sinteze makedonskih samostalnikov, pridevnikov in glagolov. Za učenje pregibnih paradigem makedonskih besed smo uporabili sistem Clog, ki temelji na induktivnem logičnem programiranju. Clog se nauči odločitvenih seznamov prvega reda, to je urejenih seznamov pravil. Za učenje in testiranje pravil smo uporabili besede, ki jih vsebuje roman "1984" G. Orwella. Dosegli smo visoko točnost (preko 90%), kar je spodbudno tako za nadaljnje raziskave pri označevanju makedonskih jezikovnih virov, kot za praktično uporabi, npr. pri mrežnih iskalnikih.

1. Introduction

The Macedonian language belongs to the South-Slavic family of languages and with the other Slavic languages shares a rich system of inflections. Although morphological rules for word formation in Macedonian have been exhaustively studied by linguists for decades (Koneski, 1952, 2004), they have not been systematized until recently (Petrovski, 2005). This paper is concerned with the problem of machine learning of morphological rules for producing the inflectional forms of nouns, adjectives and verbs given the base form (lemma), and of deducing the base form from those inflectional forms, i.e., of learning rules for morphological synthesis and analysis.

This task has been previously addressed for other Slavic languages, such as Slovenian (Erjavec and Džeroski, 2004), Czech, Bulgarian, etc. The first attempt to do this for Macedonian was by learning rules for morphological analysis and synthesis of nouns (Ivanovska et al., 2005). In this study the same approach was applied to the Macedonian adjectives and verbs.

The examples (word-forms) used in the process of learning rules for morphological analysis and synthesis were taken from the Macedonian translation of George Orwell's "1984", which itself is meant to become a part of the Multext-East (Multilingual Text Tools and Corpora for Eastern and Central European Languages) language resources (Erjavec, 2004).

In this paper, we first describe the preprocessing of the data, which consists of the annotation of words made in line with Multext-East notation and transforming the data in a format for running the ILP system Clog. Particular attention in the paper is paid to the process of learning rules for morphological analysis and synthesis of Macedonian words using the ILP system Clog (Manandhar et al., 1998). Section 2 explains in details the preprocessing of the data. The experiments and the experimental results are presented in Section 3. Conclusions discuss the results and directions for future work.

2. Preparing the lexicon

The morphosyntactic annotation of words was made according to the Multext-East specification (Erjavec, 2004), where each word-form is associated with a morphosyntactic description (MSD) presented as a packed string. Its first character, always uppercase, represents the part-of-speech (grammatical category). It is followed by a list of character values corresponding to the part-of-speech dependent attributes. For instance, the MSD Ncmsnn expands to PoS (part-of-speech): Noun; Type: *common*; Gender: *masculine*; Number: *singular*; Case: *nominative*; and Definiteness: *no*.

Nouns in Macedonian are marked for type (2 values), gender (3 values), number (3 values), case (3 values) and definiteness (4 values); adjectives are marked for type (3), degree (4), gender (3), number (2) and definiteness (4); verbs are marked for type (4), form (3), tense (3), person (3), number (2), gender (3), negative (2) and aspect (2) (Figures 1, 2, 3).

The annotation of Macedonian words was facilitated with a specially designed program (Fig. 5). PoS tagging was automatically performed according to word prefixes and suffixes. Correction of misclassified words was done during MSD tagging. This process was also performed automatically, and polished manually (Ivanovska et al

2005). However, even with manual correction, annotation of the words was not error free, as noticed during the process of learning morphological rules, and was afterwards corrected. In parallel with the MSD annotation, lemmas were added in a separate column. The structure of the resulting lexicon is illustrated in Figure 4.

```
= ============== ============== =
P ATT            VAL
= ============== ============== =
1 Type           common         c
                 proper         p
- -------------- -------------- -
2 Gender         masculine      m
                 feminine       f
                 neuter         n
- -------------- -------------- -
3 Number         singular       s
                 plural         p
                 count          t
- -------------- -------------- -
4 Case           nominative     n
                 vocative       v
                 oblique        o
- -------------- -------------- -
5 Definiteness   no             n
                 yes            y
                 distal         d
                 proximal       p
- -------------- ---------------
```

Figure 1. Attributes and their possible values of the grammatical category Nouns

```
= ============== ============== =
P ATT            VAL
= ============== ============== =
1 Type           qualificative  f
                 ordinal        o
- -------------- -------------- -
2 Degree         positive       p
                 comparative    c
                 superlative    s
                 elative        e
- -------------- -------------- -
3 Gender         masculine      m
                 feminine       f
                 neuter         n
- -------------- -------------- -
4 Number         singular       s
                 plural         p
- -------------- -------------- -
5 Definiteness   no             n
                 yes            y
                 distal         d
                 proximal       p

- -------------- -------------- -
```

Figure 2. Attributes and their possible values of the adjectives

```
================ ============== =
P ATT            VAL
= ============== ============== =
1 Type           main           m
                 auxiliary      a
                 modal          o
                 copula         c
- -------------- -------------- -
2 VForm          indicative     i
                 imperative     m
                 participle     p
- -------------- -------------- -
3 Tense          present        p
                 imperfect      i
                 aorist         a
- -------------- -------------- -
4 Person         first          1
                 second         2
                 third          3
- -------------- -------------- -
5 Number         singular       s
                 plural         p
- -------------- -------------- -
6 Gender         masculine      m
                 feminine       f
                 neuter         n
- -------------- -------------- -
7 Negative       no             n
                 yes            y
- -------------- -------------- -
8 Aspect         progressive    p
                 perfective     e
- -------------- -------------- -
```

Figure 3. Attributes and their possible values of the verbs

Next, the lexicon had to be transformed into the format suitable for running Clog. The tables were split into three documents (one for every analyzed grammatical category), and the text was (for easier inspection) transliterated into the Latin alphabet.

авион   авион   Ncmsnn
авиони   авион   Ncmpnn
автентични   автентичен   Afpmp-n
автоматска   автоматски   Aopfs-n
асимилира   асимилира   Vmip3s--n-----p
асоцираа   асоцира   Vmii3p--n-----p

Figure 4. An example of the structure of the lexicon

Look_up_podvid_na_zbor : Table

ID	vid_na_zbor	nad_vid
1	именки општи	1

ID	zborovi	type	gender	number	case	definiteness
1	'рбет	c	m	s	n	n
4	'рбетот	c	m	s	n	y
5	'рѓата	c	f	s	n	y
7	'ртот	c	m	s	n	y
10	август	c	m	s	n	n
11	авенијата	c	f	s	n	y
12	авион	c	m	s	n	y
13	авиони	c	m	p	n	n
14	авионите	c	m	p	n	y
15	авионот	c	m	s	n	y
16	Австралазија	p	f	s	n	n
20	авто	c	n	s	n	y
21	автобус	c	m	s	n	n
22	автобуската	c	f	s	n	[illegible]
23	автомат	c	m	s	n	y
24	автомати	c	m	p	n	n
25	автоматот	c	m	s	n	c
29	автомобил	c	m	s	n	d
30	автомобилине	c	m	p	n	d
31	автор	c	m	s	n	n

Record: 19 of 4867

Figure 5. An example of the structure of the lexicon

3. Experiments and results

This section describes how the morphological analysis and synthesis of nouns, adjectives and verbs was carried out using the inductive logic programming system Clog.

We will fist explain the notion of first-order decision lists on the problem of synthesis of the past tense of English verbs, one of the first examples of learning morphology with ILP, using the ILP system FOIDL (Mooney and Califf, 1995).

The ILP formulation of the problem is as follows. A logic program has to be learned defining the relation past(PresentVerb, PastVerb), where PresentVerb is an orthographic representation of the present tense of a verb and PastVerb is an orthographic representation of its past tense. PresentVerb is the input and PastVerb the output argument. Given are examples of input/output pairs, such as past([b,a,r,k], [b,a,r,k,e,d]) and past([g,o], [w,e,n,t]). The program for the relation past uses the predicate split(A, B, C) as background knowledge: this predicate splits a list of letters A into two lists B and C, e.g., split([b,a,r,k,e,d], [b,a,r,k], [e,d]).

Given examples and background knowledge, FOIDL (Mooney and Califf, 1995) learns a first-order decision list defining the predicate past. And example of such list is given in Figure 6.

```
past([g,o], [w,e,n,t]) :- !.
past(A, B) :-  split(A, C, [e,p]), split(B, C, [p,t]), !.
past(A, B) :-  split(B, A, [d]), split(A, _, [e]), !.
past(A, B) :-  split(B, A, [e,d]), !.
```

Figure 6. A first-order decision list

Clog is similar to FOIDL in the sense that it also learns first-order decision lists, i.e., ordered sets of rules, from positive examples only (Manandhar et al., 1998). For the process of learning rules, triplets from the training data, presented earlier, were used. Each triplet was an example of analysis of the form *msd(orth, lemma)*, where *orth* and *lemma* are the orthographic representations of the word-form and the lemma, respectively. Within the learning setting of inductive logic programming, *msd(Orth, Lemma)* is a relation or predicate that consists of all pairs (word-form, lemma) that have the same morphosyntactic description. *Orth* is the input and *Lemma* is the output argument. In Clog the predicate *mate* is used as background knowledge instead of the predicate *split*. *mate* generalizes *split* to deal also with prefixes (useful for analyzing superlative forms of Macedonian adjectives).

A set of rules was learned for each of the *msd* predicates. The rules were encoded as PROLOG facts. An example of rules for the analysis of Macedonian qualificative indefinite feminine adjectives is given in Figure 7.

```
afpfs_n(A,B):-mate(A,B,[],[],[r,n,a],[r,e,n]),!.
afpfs_n(A,B):-mate(A,B,[],[],[d,n,a],[d,e,n]),!.
afpfs_n(A,B):-mate(A,B,[],[],[t,n,a],[t,e,n]),!.
afpfs_n(A,B):-mate(A,B,[],[],[v,n,a],[v,e,n]),!.
afpfs_n(A,B):-mate(A,B,[],[],[b,n,a],[b,e,n]),!.
```

Figure 7. An example of PROLOG rules for analysis the Macedonian qualificative indefinite feminine adjectives

3.1. Morphological analysis and synthesis

As said previously, morphological analysis is the process of deducing the base form (lemma) from the inflectional forms of the words (word-forms), while morphological synthesis is the process of producing the inflectional forms given the base form. In this section, these learnt rules are described for the three grammatical categories –adjectives, verbs and nouns.

3.1.1. Adjectives

The morphological analysis and synthesis were carried out over 5,078 word-forms of adjectives. A set of rules was learned for every MSD that has more than 100 examples. MSDs with less than 100 examples do not provide enough data to induce good rules. The rule sets vary in size and complexity over different MSDs and refer to how the suffix or/and the prefix of the word changes to obtain the base form. An example of induced exceptions and rules for analyzing the singular of Macedonian qualificative definite feminine adjectives is given in Figure 8.

```
afpfs_y([z,e,m,j,i,n,a,t,a],[z,e,m,j,i,n]):- !.
afpfs_y ([t,o,p,l,a,t,a],[t,o,p,o,l]):- !.
afpfs_y ([t,e,n,k,a,t,a],[t,e,n,o,k]):- !.
afpfs_y ([s,t,a,r,a,t,a],[s,t,a,r]):- !.

afpfs_y (A,B):-mate(A,B,,[],[],[t,r,a,t,a],[t,a,r]),!.
afpfs_y (A,B):-mate(A,B,[],[],[s,a,t,a],[s]),!.
afpfs_y (A,B):-mate(A,B,[],[],[v,a,t,a],[v]),!.
afpfs_y (A,B):-mate(A,B,[],[],[e,t,a],[e,t]),!.
afpfs_y (A,B):-mate(A,B,[],[],[m,a,t,a],[m]),!.
afpfs_y (A,B):-mate(A,B,[],[],[g,a,t,a],[g]),!.
afpfs_y (A,B):-mate(A,B,[],[],[o,k,a,t,a],[o,k]),!.
```

Figure 8. Exceptions and rules in analysis of Macedonian qualificative definite feminine adjectives in the singular

Some of the words were not correctly analyzed and this happens because of three reasons: a) the rule applied to the word-form generates a word that is not equal to the lemma, b) there is no rule that corresponds to that combination of word-form – msd, and c) there is an error due to the manual annotation of the words. An example of incorrectly lemmatized adjectives is given in Figure 9.

```
slatka   Afpfs-n   sladok   |slatk|   ERR
dobra   Afpfs-n   dobar   |dobr|   ERR
mrtvi   Afpmp-n   mrtov   |mrtv|   ERR
polna   Afpfs-n   poln   |polen|   ERR
mekoto   Afpns-y   mek   |???|   ERR
kratko   Afpns-n   kratok   |???|   ERR
slabite   Afpns-y   slab   |???|   ERR
negibnata   Afpfs-y   negibnat   |???|   ERR
blagoto   Afpns-n   blag   |???|   ERR
spokojni   Afpns-n   spokoen   |spokojen|   ERR
```

Figure 9. Incorrectly lemmatized adjectives

In Figure 9 the adjectives *slatka (sweet), dobra (good), mrtvi (dead)* and *polna (full)* were incorrectly lemmatized because the rules applied to them generated words that are not equal with their lemmas. For the adjectives *mekoto (soft), kratko (short)* and *slabite (thin)* there were no rules that could be applied so it did not generate any word. And the errors in the last three adjectives – *negibnata (untouched), blagoto (sweet)* and *spokojni (calm)*, were due to the incorrect annotation (errors in the MSDs).

To test the accuracy of the obtained rules, 10-fold cross validation was performed. After correcting the errors, which were due to the manual annotation of the adjectives, we have achieved average accuracy of 93.12%. The obtained average accuracy of the rules for the analysis is given in Table 1.

Morphological synthesis was carried out over the same set of adjectives, only the structure of the set was changed, i.e., the columns of lemmas and word-forms were swapped. The data consisted of triplets lemma-wordform-msd. Again, rules were learned only for those MSDs that have more than 100 examples. Examples of exceptions and rules for synthesizing the neuter singular form of the qualificative adjectives are given in Figure 10.

```
afpns_n([o,d,g,o,v,o,r,e,n],[o,d,g,o,v,o,r,e,n,o]):- !.
afpns_n ([u,t,v,r,d,e,n],[u,t,v,r,d,e,n,o]):- !.
afpns_n ([v,i,s,o,k],[v,i,s,o,k,o]):- !.
afpns_n ([s,t,u,d,e,n],[s,t,u,d,e,n,o]):- !.

afpns_n(A,B):-mate(A,B,[i],[i],[l,e,n],[l,n,o]),!.
afpns_n(A,B):-mate(A,B,[v,o],[v,o],[e,n],[e,n,o]),!.
afpns_n (A,B):-mate(A,B,[],[],[p],[p,o]),!.
afpns_n(A,B):-mate(A,B,[n,a],[n,a],[l,e,n],[l,e,n,o]),!.
afpns_n (A,B):-mate(A,B,[],[],[a,r],[r,o]),!.
afpns_n (A,B):-mate(A,B,[],[],[t],[t,o]),!.
afpns_n (A,B):-mate(A,B,[],[],[a,n],[a,n,o]),!.
```

Figure 10. Exceptions and rules in synthesizing the neuter singular form of the Macedonian qualificative adjectives

The accuracy of the synthesis rules was tested using 10-fold cross validation and the average accuracy is 82.77%, which is lower than the accuracy of the rules obtained in the morphological analysis, mostly because of the large number of inflectional forms in which one adjective can be found. The average accuracy and standard deviation for the morphological synthesis of the adjectives are given in Table 1.

	Analysis	Synthesis
Accuracy (%)	93.12	82.77
Standard deviation	1.30610915	1.066208235

Table 1. Average accuracy and standard deviation of the rules for analysis and synthesis of Macedonian adjectives

3.1.2. Verbs

Verbs are the most complex grammatical category in Macedonian language (they have the largest number of attributes and MSDs). The morphological analysis and synthesis of verbs were carried out over 5483 word-forms of verbs. The process of learning rules is the same as described earlier. Only the MSDs that have more than 100 examples were included in the process of learning rules. An example of exceptions and rules for analyzing the participle form of the Macedonian verbs is given in Figure 11.

```
vmpa3sm_n__e([d,o,b,i,l],[d,o,b,i,e]):- !.
vmpa3sm_n__e ([z,e,l],[z,e,m,e]):- !.
vmpa3sm_n__e ([r,a,z,b,r,a,l],[r,a,z,b,e,r,e]):- !.
vmpa3sm_n__e ([i,s,p,i,l],[i,s,p,i,e]):- !.

vmpa3sm_n__e(A,B):-mate(A,B,[p],[p],[z,n,a,l],[z,n,a,e]),!.
vmpa3sm_n__e (A,B):-mate(A,B,[],[],[k,a,l],[k,a]),!.
vmpa3sm_n__e(A,B):-mate(A,B,[],[],[l,e,g,o,l],[l,e,z,e]),!.
vmpa3sm_n__e(A,B):-mate(A,B,[],[],[d,a,l],[d,a,d,e]),!.
vmpa3sm_n__e (A,B):-mate(A,B,[],[],[n,a,l],[n,e]),!.
vmpa3sm_n__e (A,B):-mate(A,B,[],[],[i,l],[i]),!.
```

Figure 11. Exceptions and rules in analysis of participle form of Macedonian verbs

Again, there were incorrectly lemmatized verbs and the reasons for that are the same as for the adjectives. Figure 12 shows some incorrectly lemmatized verbs.

```
dobija  Vmia3p--n-----e  dobie  |dobi|  ERR
razberat  Vmip3p--n-----e  razbere  |razberi|  ERR
dozna  Vmia2s--n-----e  doznae  |dozne|  ERR
umrat  Vmip3p--n-----e  umre  |umri|  ERR
zaprea  Vmia3p--n-----e  zapre  |???|  ERR
odigraa  Vmia3p--n-----e  odigra  |???|  ERR
zagriza  Vmia3s--n-----e  zagrize  |???|  ERR
sporedat  Vmia3p--n-----e  sporedi  |???|  ERR
nadvladea  Vmii2s--n-----p  nadvladee  |???|  ERR
nagovorila  Vmii3s--n-----p  nagovara  |???|  ERR
```

Figure 12. Incorrectly lemmatized verbs

The verbs *dobija (to get), razberat (to understand), dozna (to find out)* and *umrat (to die)* were incorrectly lemmatized because the rules applied to them did not generate the right base-forms. For the verbs *zaprea (to stop), odigraa (to play)* and *zagriza (to bite)* there were no rules that could be applied and the errors in the last three verbs – *sporedat (to compare), nadvladea (to dominate), nagovorila (to persuade)* – were due to errors in the annotation (wrong MSDs).

The average accuracy of the obtained rules for morphological analysis of verbs, tested with 10-fold cross validation, is 91.65% (Table 2).

For the process of morphological synthesis the data were transformed into triplets lemma-wordform-msd and rules were learned over it. Some of the rules and exceptions for synthesizing the participle form of the Macedonian verbs are presented in Figure 13.

```
vmpa2sm_n__e([s,v,r,t,i],[s,v,r,t,e,l]):- !.
vmpa2sm_n__e ([r,a,z,b,e,r,e],[r,a,z,b,r,a,l]):- !.
vmpa2sm_n__e ([p,r,i,v,r,z,e],[p,r,i,v,r,z,a,l]):- !.
vmpa2sm_n__e ([p,o,s,t,o,i],[p,o,s,t,o,e,l]):- !.

vmpa2sm_n__e(A,B):-mate(A,B,[],[],[z,n,a,e],[z,n,a,l]),!.
vmpa2sm_n__e(A,B):-mate(A,B,[],[],[v,i,d,i],[v,i,d,e,l]),!.
vmpa2sm_n__e (A,B):-mate(A,B,[],[],[e,e],[e,a,l]),!.
vmpa2sm_n__e(A,B):-mate(A,B,[],[],[l,e,z,e],[l,e,g,o,l]),!.
vmpa2sm_n__e(A,B):-mate(A,B,[],[],[d,a,d,e],[d,a,l]),!.
vmpa2sm_n__e (A,B):-mate(A,B,[],[],[a],[a,l]),!.
vmpa2sm_n__e (A,B):-mate(A,B,[],[],[n,e],[n,a,l]),!.
vmpa2sm_n__e (A,B):-mate(A,B,[],[],[i],[i,l]),!.
vmpa2sm_n__e (A,B):-mate(A,B,[],[],[i,e],[i,l]),!.
```

Figure 13. Exceptions and rules in synthesizing the participle form of Macedonian verbs

The average accuracy of the obtained rules for the synthesis of verbs, tested with 10-fold cross validation, is 95.71%. (Table 2).

	Analysis	Synthesis
Accuracy (%)	91.65	95.71
Standard deviation	1.383468	1.424956

Table 2. Average accuracy and standard deviation of the rules for analysis and synthesis of Macedonian verbs

3.1.3. Nouns

The process of morphological analysis and synthesis of nouns was the same as for adjectives and verbs. Some exceptions and rules for analyzing the common nouns of feminine gender are presented in Figure 14.

10-fold cross validation was used to test the accuracy of the rules for morphological analysis and synthesis of nouns. The average accuracy obtained was 97.01% (Ivanovska et al., 2005). Morphological synthesis of the nouns resulted with slightly smaller average accuracy of 94.81%.

```
ncfpnn([r,a,s,p,r,a,v,i,i],[r,a,s,p,r,a,v,a]):- !.
ncfpnn([s,t,r,u,i],[s,t,r,u,j,a]):- !.
ncfpnn([r,a,c,e],[r,a,k,a]):- !.
ncfpnn([n,o,z,e],[n,o,g,a]):- !.

ncfpnn(A,B):-mate(A,B,[],[],[s,t,i],[s,t]),!.
ncfpnn(A,B):-mate(A,B,[],[],[i,i],[i,j,a]),!.
ncfpnn(A,B):-mate(A,B,[i,d],[i,d],[i],[j,a]),!.
ncfpnn(A,B):-mate(A,B,[],[],[i],[a]),!.
```

Figure 14. Exceptions and rules in analysis of common Macedonian nouns of feminine gender

Some exceptions and rules for synthesis of common neuter nouns in plural are presented in Figure 15.

```
ncnpny([d,e,t,e],[d,e,c,a,t,a]):- !.
ncnpny([z,i,v,o,t,n,o],[z,i,v,o,t,n,i,t,e]):- !.
ncnpny([b,e,b,e],[b,e,b,i,n,j,a,t,a]):- !.

ncnpny(A,B):-mate(A,B,[p,o],[p,o],[e],[i,n,j,a,t,a]),!.
ncnpny(A,B):-mate(A,B,[],[],[c,e],[c,a,t,a]),!.
```

Figure 15. Exceptions and rules in synthesizing the common neuter nouns in plural

4. Conclusions

In this paper we have addressed the problem of morphological analysis and synthesis of Macedonian adjectives, verbs and nouns. To learn rules for morphological analysis and synthesis we used word-forms from the Macedonian translation of Orwell's "1984".

We successfully applied the ILP system Clog for learning rules for analysis and synthesis of Macedonian words. The obtained average accuracies of the learned rules for analysis and synthesis of adjectives are 93.12% and 82.77%, respectively; for verbs – 91.65% and 95.71%; and for nouns - 97.01% and 94.81%.

The high accuracies achieved are encouraging for further research as well as practical use. For example, morphological analysis/synthesis could be used in web search engines.

To this end the rules for morphological analysis and synthesis need to be connected with PoS tagging to perform lemmatization. Although preliminary research has been done on learning PoS tagging for Macedonian (Vojnovski, 2005), further work with a larger and better annotated corpus is needed. Once we have a PoS tagger we can use the rules learned in this work for lemmatization, which is relevant both for practical use as well as for example another syntactic annotation of Macedonian language resources.

5. References

Dimitrova, L., T. Erjavec, N. Ide, H.-J. Kaalep, V. Petkevič, and D. Tufis. 1998. Multext-east: Parallel and comparable corpora and lexicons for six central and eastern european languages. In *Proceedings of the COLING-ACL '98*, pages 315–319, Montreal, Quebec, Canada. http://nl.ijs.si/ME/.

Erjavec, T. 2004. MULTEXT-East Version 3: Multilingual Morphosyntactic Specifications, Lexicons and Corpora. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'2004)*, ELRA, Paris., pp. 1535-1538.

Erjavec, T., and S.Džeroski, 2004. Machine Learning of Morphosyntactic Structure: Lemmatizing unknown Slovene words. *Applied Artificial Intelligence* 18(1), 2004, pp. 17-41.

Ivanovska, A., K.Zdravkova, S.Džeroski, and T.Erjavec, 2005. Learning rules for morphological analysis and synthesis of Macedonian nouns. In *Proceedings of SiKDD 2005 (Conference on Data Mining and Data Warehouses)*, Ljubljana, Slovenia, pp. 195-198.

Koneski, B., 1952. Grammar of Standard Macedonian. *Prosvetno Delo*, first edition, in Macedonian

Koneski, B., 2004. Grammar of Macedonian Language. *Prosvetno Delo*, in Macedonian

Manandhar, S., S.Džeroski, and T.Erjavec, 1998. Learning multilingual morphology with CLOG. In *Proceedings of Inductive Logic Programming: 8th International Workshop (ILP-98),* Number 1446 in Lecture Notes in Artificial Intelligence, ed. D. Page, pages135-144, Berlin, Springer-Verlag

Mooney, R. J., and M. E. Califf. 1995. Induction of first-order decision lists: Results on learning the past tense of English verbs. *Journal of Artificial Intelligence Research* 3(1):1–24.

Petrovski, A., 2005. About Macedonian Computational Dictionary. In *Proceedings of BCI2005,* Ohrid, Macedonia, pp. 76-81

Vojnovski, V., S. Džeroski, and T. Erjavec, 2005. Learning PoS tagging from a tagged Macedonian text corpus. In *Proceedings of SiKDD 2005 (Conference on Data Mining and Data Warehouses)*, Ljubljana, Slovenia, pp. 199-202.