Recognizing Verb-based Croatian Idiomatic MWUs

Kristina Kocijan and Sara Librenjak

Department of Information and Communication Sciences
Faculty of Humanities and Social Sciences, Zagreb
krkocijan@ffzg.hr, sara.librenjak@gmail.com

Abstract.
This paper tackles the computational problems of Croatian verbal idioms. Croatian language has very rich phraseme structure, as described in Matešić (1982), Menac et.al. (2003; 2007) and Menac-Mihalić (2004), as well as many others. This work is one of the few attempts of computational analyis of idioms in Croatian language as multi-word units. We used rule-based approach and NooJ syntactic grammars in order to recognize any verb based idiom (of the ~1500 analyzed) in any syntactic position. The Croatian Dictionary of Idioms (Menac et al., 2003) was used for the initial list, which was implemented with new additions during training phase. Grammars were tested within the corpora constructed specifically for this work, and used to calculate statistical measures of recall, precision and f-measure for our grammars. With the final results of recall <98%, precision < 96% and f-measure <97%, we consider this a successful attempt in the recognition of verb based idioms in Croatian language.

Keywords: croatian, idioms, verbal phrases, NooJ, MWU, frozen expressions, semi-frozen expressions

1 Introduction

Idioms, or the non-literal expressions, are considered to be an important and quite large field of any language, but are believed to be especially rich in the Croatian language. They are often a topic of discussion for many linguists and Croatian language researchers. Various types of idioms are present in most styles of spoken and written language, although pertaining more to spoken and journalistic style. They can be found in many internet texts written by users, newspaper (especially sport and gossip), literature, and of course natural speech of native speakers when one tends to be more poetic or hyperbolic in their expression. On the other hand, they are rarely found in the scientific and specialized texts, since they constitute a type of speech which is not always

specific and stylistically marked. The literature (Fink Arsovski et al., 2010) cites that the first papers on Croatian idioms were written not earlier than 1970-ies covering different aspects but also different variants of Croatian (contemporary, old, dialectal).

Although the existence of idioms in most (or all) languages is a known fact, they are not always so present in the computational analysis of natural language. The idioms pose as a difficult field for computational approach. This is due to few unavoidable factors: they are mostly multi-word in nature, their meaning is hard to grasp for a computer, and polysemy is frequent. Also, their cultural and historical nuances render them very difficult to process or translate without special preparation.

We believe that the approach for dealing with the idioms in Croatian language should be rule-based, and the NooJ was the ideal tool for the task. Provided that we assure the detailed syntactic description and categorization of idioms, NooJ can be used to construct grammars to detect any known idioms in a Croatian text. Using somewhat different approaches than we propose in this paper, NooJ has already been tried in the area of multi word expressions, including, among others, Bulgarian verbal idioms (Todorova, 2008), Italian support verb constructions (Chatzitheodorou, 2014), English phrasal verbs (Machonis, 2010, 2012) and Greek idioms (Gavrilidou et al., 2012).

Thus, this work describes the process of categorizing and describing the idioms, writing the grammars and testing them on corpora, using NooJ as the environment. We specifically concentrate on the most complex and most numerous type of idioms, those based on a verbal phrase. Idioms based on a verbal head take up a bigger portion of idioms, not only in Croatian but in other languages as well (Wehrli, 1998; Radikovna Sakaeva et al, 2013). It is thus of a great importance that they are analyzed and classified with care so that their identification in text may serve as a good basis for any future work (e.g. translation, information retrieval, language learning (Granger et al., 2006) etc.).

2 Methodology and corpora

As a starting point for this work, we collected idioms from the printed Croatian Dictionary of Idioms - CDI (Menac et al.; 2003). We found there many syntactically different types of idioms, such as:

a) fixed phrase which does not change in any syntactic environment
b) noun phrase with an attribute or apposition
c) verbal phrase with a direct object
d) verbal phrase with the optional direct object which can disrupt the syntactic structure
e) comparative structure (A/V as N).

In this article, we will provide the detailed analysis of the third and fourth category, i.e. verb based idioms. This work is a continuation of the work presented in Kocijan & Librenjak (to appear 2015) where general description of all the types is given and Kocijan & Librenjak (to appear 2016) where we described the comparative structures in

more detail. Verbal phrases are most complex of all the types and require special attention, thus we chose to concentrate our interest on this type in more detail in this article.

After consulting the CDI, we analyzed all the verb-based idioms syntactically, taking special care when it comes to structures that can be disrupted with objects, those that can be inverted or negated. Closer research yielded more sub-types which will be discussed in latter chapters. In addition to CDI, additional idioms were found while working on this project and added to the corpus of idioms. This gives the cumulative number of approximately 1500 verb-based idioms which were analyzed for the purpose of this work. All the idioms were listed in the NooJ dictionary, along with their type, sub-type and possible objects.

In the following step, we constructed syntactic NooJ grammars in order to recognize verbal idioms in all possible contexts and syntactic variations that can come to being in the Croatian language. As this is the most important part of this work, it is discussed at length in the sections that follow, along with the grammars, corpora for both training corpus and testing corpus needed to be constructed.

The training corpus was used in order to improve the grammars during their construction period, and it was constructed specifically for the purposes of our idioms research. This is a smaller corpus of approximately 60 Kw (60 000 words). It consists of sentences from Croatian newspaper articles and contemporary Croatian authors and each sentence has at least one version of the processed idioms[1].

Subsequently, finished grammars were tested on the web based Croatian corpus sample (Agić and Ljubešić, 2014), and compared with manually marked results. Sentences were extracted randomly, and a corpus of approximately 100 Kw (100 000 word) was built. Each sentence was manually checked and marked if an idiom was found. This enabled us to get statistical measures of our grammars, such as recall, precision and f-measure. The results are described in the Evaluation section of the article.

3 Verb-based Idioms

Verbs that are part of a comparative MWU (Verb as NP, Verb as PP or Verb as NP+PP) are placed into the category 5c (Kocijan, Librenjak, to appear). Such examples are *buljiti kao tele u šarena vrata, čekati kao ozeblo sunce, govoriti kao iz knjige*.

Also, those MWUs that have a verb in a fixed position and no variation i.e. the verb never changes gender, number, person or tense, are placed in the category 1 (*nije šala, ni pas s maslom ne bi pojeo, povuci-potegni, pričam ti priču*). They are listed in the dictionary as simple entries (Kocijan, Librenjak, to appear)[2] and belong to fixed expressions, as termed by Sag et al. (2002).

---

[1] An on-line version of Croatian corpus of idioms is prepared and maintained by Rittgasser and Fink-Arsovski at http://www.lingua-hr.de/phraseologie/stichwort.html.

[2] In the two papers published (Kocijan and Librenjak, to appear in 2015, 2016) we have used a special category NW to describe the MWUs. Since that feature is no longer supported in NooJ 5, we have decided to change the NW notation with the FXC. The remaining of our dictionaries and grammars remain the same.

Verbs appearing as a head of an idiom that undergo any type of morphological variation as well as word order variations and short insertions belong to Type 3 and Type 4 classes of idioms. Each of these classes belongs to syntactically flexible expressions (Sag et al., 2002), thus their subtypes (7 and 8 respectively) describe different syntactic patterns.

We use the following model to describe these two classes inside the NooJ dictionary: the verb (used in the idiom) is considered to be the main entry in the dictionary while the remaining of the phrase is entered as the verb's semantic description using the notation PX, SUFX, SUFXA, SUFXB, SUFXC and SUFXV. The main entry is marked as a category verb (V) with additional feature PHR (short for phrase) and special feature FXC. This special feature in NooJ is associated with lexical entries "that must not result in real annotations" but remain visible to the syntactic grammar (Silberztein, 2003). This way we are able to avoid the unnecessary annotations of the text if only the verb is used in a sentence. Thus, only if the entire expression is recognized (verb and its suffix part/s) the string will be annotated as an idiom. In all other cases, each word will be annotated as a single lexical unit with the lexical information that inherits from the dictionary and related grammars. Since verbs change their form (gender, number, tense, person), each dictionary entry is also provided with the FLX value so that any variation is recognized. Also, each entry has its type (Type=3|4) and subtype (SType=a|b|m|n|p|v|w and SType=a|b|c|d|e|f|g|p) defined. This classification of types and subtypes is used inside the corresponding syntactic grammars for the purposes of constraining and annotating the recognized strings. The distribution of types, subtypes and PX, SUFX, SUFXA, SUFXB, SUFXC and SUFXV used in the dictionary is given in Fig. 1.

Although there are 605 type 3 and 592 type 4 main entries in the dictionary, there are more than 1197 verbal idioms that we recognize. This is possible since one dictionary entry may hold more than one SUFX, SUFXA and/or SUFXB values. For example, the verb *dobiti* (en. to get) forms an idiom with 17 possible SUFX endings, like: *dobiti šipak, dobiti brus, dobiti figu, dobiti nogu, dobiti krila* etc. and verb *imati* (en. to have) has 7 SUFXA values that combine with one and the same SUFXB (*ruke* – en. hands) to form 7 different idioms like: *imati krvave* | *slobodne* | *okrvavljene* | *dvije lijeve* | *odriješene* | *prljave* | *zlatne ruke* (en. to have blody|free|gory|two left|loose|dirty|golden hands) while the verb *biti* (en. to be) has one SUFXA (*na* – en. on) and 20 SUFXB values and forms idioms like *biti na aparatima* | *konju* | *izdisaju,* | *cijeni* (en^{lit}. to be on life support | horse | exhale | price). Majority of type 3 verbs are subtype **a** (573 main entries) and they use 730 SUFX, 328 SUFXA and 479 SUFXB values. Other subtypes in this category are quite small with 10 or less main entries but were necessary in order to raise the precision of the grammars.

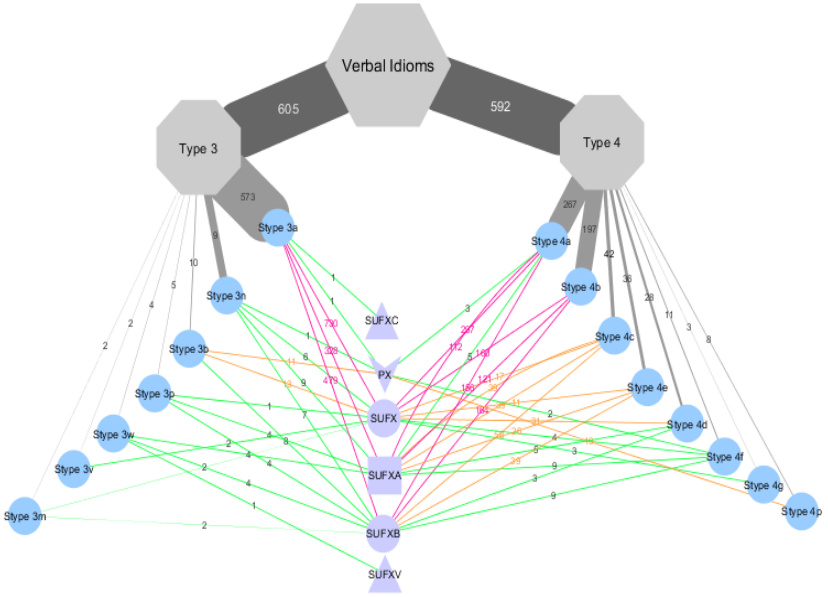

Fig. 1. - Distribution of verbal idioms and their types

The largest subtype group in type 4 verbal idioms are subtypes a and b with 267 and 197 entries respectively. These two subcategories use 457 SUFX, 233 SUFXA and 324 SUFXB values in total.

Many of our dictionary entries have more than one possible variants (ex. *staviti **bubu** u uho = staviti **buhu** u uho*). Regardless the fact that some are more widely used than others (Fink and Menac, 2008), we have decided on including all available variants. Thus the idioms *baciti prašinu u oči* and *baciti pijesak u oči* (en[lit]: 'throw dust into the eyes' and 'throw sand into the eyes', meaning 'to deceive someone') have only one dictionary entry with two SUFXA attributes and one SUFXB attribute (since it is same for both variants):

- baciti, NW+FLX=BIRATI+Type=3+SType=a+**SUFXA=prašinu** +**SUFXA=pijesak**+SUFXB=u oči

In the following two sections we will discuss type's 3 and 4 forms, grammars used for their detection and will provide examples of both dictionary entries and concordances for each subtype.

3.1 Type 3 and subtypes

The main difference between type 3 and type 4 idioms is that type 3 has continuous suffix section while type 4 allows discontinuity i.e. insertion of noun or prepositional phrases. In order to facilitate our grammars, we have decided on the following subcategorization of type 3 verbal idioms:

SType=**a** category allows for the verb and the suffix to be interrupted with some other word categories, but the entire suffix (if it is built from two sections like SUFXA and SUFXB) must remain uninterrupted (see Fig2. for the segment of 3a grammar).

- ostajati,V+PHR+FXC+Type=3+SType=a+FLX=SMIJATI+SUFX=kratkih rukava
  - *On ostaje (uvijek) kratkih rukava.* – en[lit] He stays (always) with short sleeves.
- ostati,V+PHR+FXC+Type=3+SType=a+FLX=ZASTATI+SUFXA=bez+ SUFXB=riječi+SUFXB=teksta
  - *Ostao je bez riječi.* – en[lit] He stayed withouth words. (*He was speechless.)*

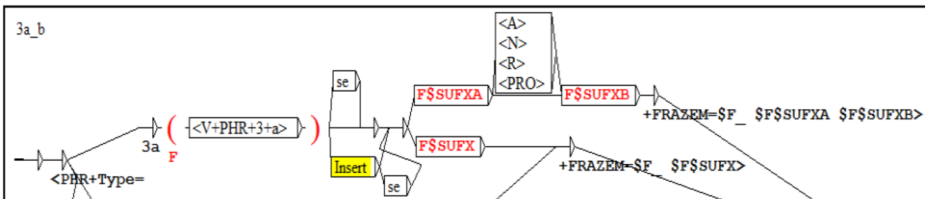

Fig. 2. Grammar for recognizing verbal idioms Type 3a

SType=**b** is similar to the SType **a** but it uses SUFX and SUFXB attributes. The first one holds the obligatory part of expression that must always appear next to the verb, while the SUFXB holds an elective part of the expression that may or may not appear in order for the idiom to be valid.

- nestati,V+PHR+FXC+Type=3+SType=b+FLX=ZASTATI+SUFX=bez traga+ SUFXB=i glasa
  - *Nestao je bez traga (i glasa).* – en[lit] He disappeared without a trace (and voice).

SType=**m** has two verbs, one on the each side of the idiom, both of which may change in tense, gender, person and number all of which must match between the both verbs in order for the expression to be recognized.

- rezati,V+PHR+FXC+Type=3+SType=m+FLX=PODIZATI+SUFX=na kojoj+ SUFXV= sjediti
  - *rezati granu na kojoj sjediš* – en[lit] to cut the branch you are sitting on

SType=**n** uses verbs that exist in an idiom only if negated, thus the expression '*nije vidio ni prst pred nosom*' is marked an idiom, while the same construction without the negation 'nije' is not ('*vidio je prst pred nosom*'). Grammar recognizing this subtype is given in Fig. 3.

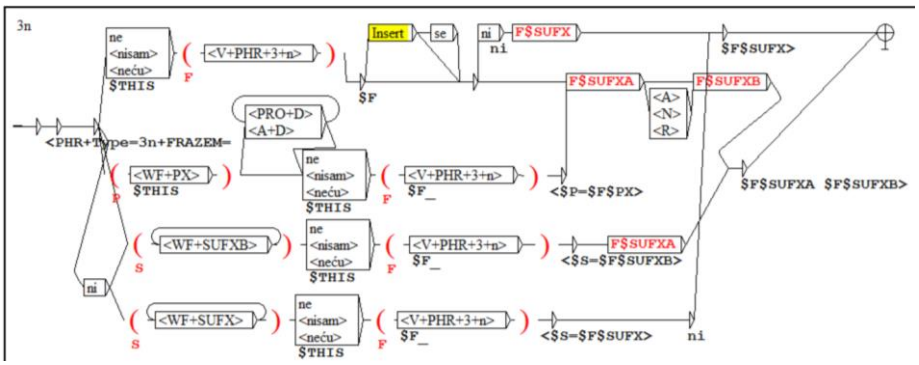

Fig. 3. Grammar for recognizing verbal idioms type 3n

– vidjeti,V+PHR+FXC+Type=3+SType=n+FLX=VIDJETI+SUFXA=prst pred+ SUFXB=nosom
  - *nije vidio ni prst pred nosom* – en[lit] didn't see not even a finger in front of his nose

SType=**p** has the verb that may only exist as an active (PDR) or passive (PDT) participle.

– premazati,V+PHR+FXC+Type=3+SType=p+FLX=KAZATI+SUFXA=svim+ SUFXB= farbama

The grammar (Fig. 4) recognizes both regular word order and inversion:

  - *premazan si svim mastima* – en[lit] you are painted with all different greases
  - *svim si mastima premazan*

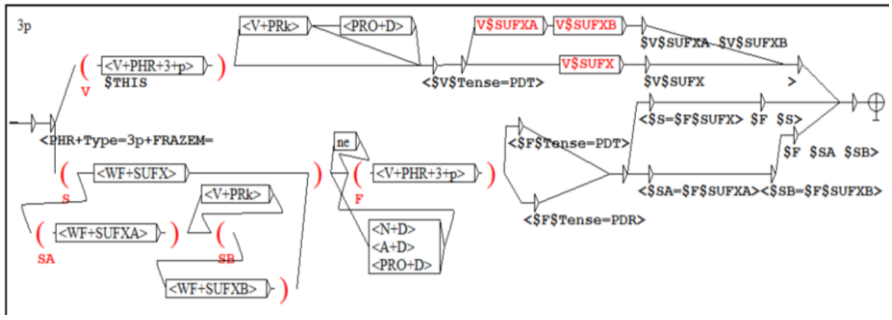

Fig. 4. Grammar for verbal idioms Type 3p

SType=**v** has two verbs (both verbs have to match in person, gender, number and tense) and have an expression 'nit | niti' before each verb. The grammar for subtypes v and w is given in Fig. 5 (lower path).

– smrdjeti,V+PHR+FXC+Type=3+SType=v+FLX=STIDJETI+SUFX=mirisati
  - *niti smrdi nit miriši* – en[lit] it neither stinks nor smells

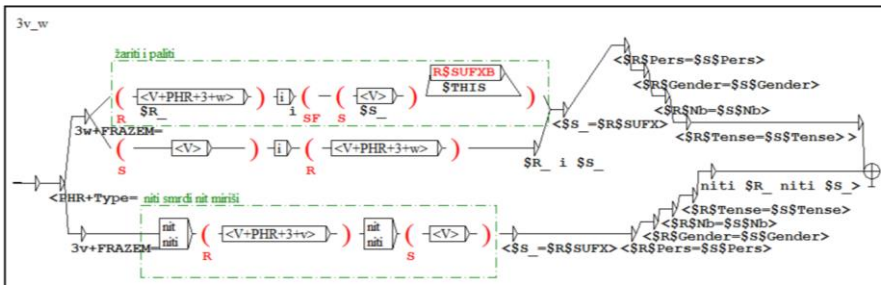

Fig. 5. - Grammar for types 3v and 3w

SType=**w** is similar to the subcategory v as it also has two verbs that must match in person, gender, number and tense, but the two verbs are connected with a conjunction '*i*' (en. and) (see upper path in Fig. 5).

- žariti, NW+Type=3+SType=w+FLX=BILJEŽITI+SUFX=paliti
  - *on žari i pali* – en[lit] he anneals and burns

3.2 Type 4 and subtypes

As already noted, the main characteristic of type 4 verbal idioms is that the main verb requires noun phrase or prepositional phrase in particular case (Nominative, Genitive, Dative, Accusative, Locative or Instrumental) to be present in the sentence. Thus, this type allows for the insertion of an NP and/or a PP after the verb but also, although rarely, between the suffix sections. It is also possible that an NP or PP required by the verb is outside of the idiom borders, i.e. it occurs in the text before or after the verbal idiom. In such cases, the verbal idiom is recognized without the NP/PP section.

Another quite frequent insertion is that of an additional attribute(s) or pronoun either between a verb and a SUFX section or between SUFXA and SUFXB sections. In the later case, SUFXA is always a preposition. Examples for this would be *osvojio (njeno) srce* where an idiom 'to win a heart' has additional pronoun 'her' ('he won her hurt') while *ubacio je buhu u (njegovo malo) uho* has a pronoun (*njegovo*) and an adjective (*malo*) inserted between SUFXA=*u* and SUFXB=*uho* (en[lit] he inserted a bug into (his little) ear).

Verbal idioms of type 4 are further subcategorized into the following 8 subtypes:

SType=**a** has the verb that requires dative construction which may be inserted between a verb and the remaining part of the expression or it can appear between SUFXA and SUFXB.

- baciti,V+PHR+FXC+Type=4+SType=a+FLX=BACITI+SUFXA=rukavicu+ SUFXB= u lice

The grammar (Fig. X) recognizes both regular word order and inversion:

- *<u>bacio</u> mu je <u>rukavicu</u> <u>u lice</u>* – en[lit] he throw him the glove into face
- *<u>u lice</u> mu je <u>bacio</u> <u>rukavicu</u>*

SType=**b** has the verb that requires accusative construction.

- poslati,V+PHR+FXC+Type=4+SType=b+FLX=POSLATI+SUFXA=k+ SUFXB=vragu
  - *poslao svu birokraciju k vragu* – en[lit] send all the bureaucracy to the devil

SType=**c** has the verb that requires genitive construction.

- izbiti,V+PHR+FXC+Type=4+SType=c+FLX=ZABITI+SUFXA=iz+ SUFXB=glave
  - *izbio (ga) je iz glave* – en[lit] he beat (him) out of the head

SType=**d** has the verb that requires instrumental construction.

- imati,V+PHR+FXC+Type=4+SType=d+FLX=SJATI+SUFXA=neraščišćene+ SUFXB= račune
  - *imaju s nama neraščišćene račune* – en[lit] have with us unclear bills

SType=**e** has the verb that requires possessive pronoun or possessive adjective.

- zakucati,V+PHR+FXC+Type=4+SType=e+FLX=SJATI+SUFXA=na+ SUFXB=vrata
  - *zakucati na (njegova) vrata* – en[lit] to knock on (his) door

SType=**f** has the verb that requires nominative construction.

- visjeti,V+PHR+FXC+Type=4+SType=f+FLX=SJATI+SUFXA=u+ SUFXB=zraku
  - *visjelo je (to) u zraku* – en[lit]

SType=**p** has the verb that requires obligatory prefixal word (*čudom se čuditi, curkom curiti, daleko dogurati, bijelo gledati, lose|crno|slabo mu se piše*).

- dogurati,V+PHR+FXC+Type=4+SType=p+FLX=SJATI+SUFX=daleko
  - *oni će daleko dogurati* – en[lit] they will get far and away.

4 Evaluation

In this section, we will give the results (P=precision, R=recall and F=f measure) for each category (3 and 4) separately and in total to get a better understanding of graphs used for recognizing each category.

The problems we encountered may be divided into three different categories that we will exemplify in the following paragraphs.

	words	Total in %			**Type=3**			**Type=4**		
		P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**

Training corpus	58 223 w	98,9	96,3	97,6	100	100	100	99,2	96,1	97,6
Testing corpus	2 247 Kw	100	96,2	98	100	96,9	98,4	100	95	97,4

Table 1. Results from the corpora

The examples that were not recognized by our grammars are mostly due to the long distances between MWU sections or inserted comma sign. This is also, out of three, the most numerous category of unrecognized idioms.

- *Podršku jurišnicima počele su, i to obilno,* ***šakom i kapom****,* ***davati*** *i banke.* (inversion with inserted comma sign)
- *Karijeristi su brzo napredovali, a on, sposobniji od njih,* ***OSTAO JE*** *zbog svoje skromnosti cijeli život* ***U SJENI****.* (long distance between MWU sections)

Another category of unrecognized idioms belongs to the false positives like it was the case with the sentence:

- ***Stajao*** *je pred učiteljem oborene* ***glave****.*

where the idiom *stajao glave* (meaning 'to cost someone his/her life') was falsely recognized. The similar situation is with the sentence:

- *Samo se htio poigrati s tvojim psićem i slučajno mu je* ***stao na rep****.*

where the expression *stao na rep* exists as an idiom meaning 'to stop someone', but in this case it is used in its literal meaning (he stepped on his tail).

The last group of unrecognized idioms belongs to colloquial usage of language, i.e. to colloquialized idioms. Since our dictionary has idioms written only in standard Croatian language, such colloquial examples remain unrecognized by our grammars.

5 Conclusion

In this paper, we have argued for a grammar driven approach to two types of verb-based idioms in Croatian. We collected the idioms from Croatian Dictionary of Idioms (Menac et al.; 2003) and described them syntactically. For verbs, we found two distinct categories – one that cannot be interrupted with the insertion of an inflected word, and one that can. Since this work is a continuation of our comprehensive approach to idioms as MWE (Kocijan and Librenjak, to appear 2015), we refer to the verb-based idioms as types 3 and 4, as other types are described in other articles. After describing them, we constructed the NooJ grammars for both types and all of their subtypes. Although we have covered all of the examples found in Menac et al. (2003), the list of verb-based idioms is still not complete and can be extended with additional ones, for example (Fink

Arsovski et al., 2010). During the testing phase, we have also added some additional idioms found in web corpus.

Since this is the first attempt at grammar driven approach (compare with Ljubešić et al. 2014 for statistical approach) to identify such occurrences in the text, we expect some additional changes and adjustments to take place in any future projects that would include verbal idioms. The results for this work, being greater than 98%, 96% and 97% for precision, recall and f-measure respectfully, show that grammars suggested in this paper prove quite capable in recognizing the Croatian verb-based idioms.

6 References

1. Agić, Ž., Ljubešić, N. 2014. The SETimes.HR Linguistically Annotated Corpus of Croatian. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, P. 1724–27. Reykjavik.
2. Chatzitheodorou, K. 2014. Paraphrasing of Italian Support Verb Constructions based on Lexical and Grammatical Resources. In Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing, Coling 2014, Dublin, Ireland, P. 1-7.
3. Fink Arsovski, Ž., Kovačević, B., Hrnjak, A. 2010. Bibliografija hrvatske frazeologije i popis frazema analiziranih u znanstvenim i stručnim radovima.Knjigra. Zagreb.
4. Fink, Ž. and Menac A. 2008. Hrvatska frazeologija – staro i novo. In Mokienko, W. and Walter, H. (eds). Frazeologia. Komparacja spółczesnych języków słowiańskich, 3. Opole: Universität Greifswald – Institut für Slawistik, Uniwersytet Opolski – Instytut Filologii Polskiej, P. 88-100.
5. Granger, S., Paquot, M. and Rayson, P. 2006. Extraction of multiword units from EFL and native English corpora. The phraseology of the verb 'make'. In Buhofer, A.H. & H. Burger (eds.) Phraseology in Motion I, Baltmannsweiler: Schneider. P. 57-68.
6. Kocijan, K., Librenjak, S. 2015. The Quest for Croatian Idioms as Multi Word Units. To appear in Monti, J., Mitkov, R., Corpas Pastor, G. and Seretan V. (eds.) In *Multiword Units in Machine Translation and Translation Technology*, John Benjamins Publishing. [in print]
7. Kocijan, K., Librenjak, S. 2016. Comparative Idioms in Croatian: MWU Approach. To appear in Proceedings of EUROPHRAS 2015, Malaga, Spain.
8. Menac, A; Fink-Arsovski, Ž; Venturin, R. 2003. Hrvatski frazeološki rječnik. Zagreb: Naklada Ljevak.
9. Ljubešić N., Dobrovoljc K., Krek S.,, Peršurić Antonić M. and Fišer D.// hrMWELex – A MWE lexicon of Croatian extracted from a parsed gigacorpus. Language technologies: Proceedings of the 17th International Multiconference Information Society IS2014. Ljubljana, Slovenia.
10. Radikovna Sakaeva, L. and Gumerovna Nurullina, A. 2013. Comparative Analysis of Verbal, Adjectival, Adverbial and Modal Phraseological Units with a Lexeme "Devil" in English and Russian Languages. In Middle-East Journal of Scientific Research 18 (1): P. 50-54, DOI: 10.5829./idosi.mejsr.2013.18.1.12354.
11. Silberztein, M., 2003. NooJ Manual, www.nooj4nlp.net.
12. Gavriilidou Z., Papadopoulou E. and Chadjipapa E. 2012. Processing Greek Frozen Expressions with NooJ. Formalising Natural Languages with NooJ: Selected Papers from the NooJ 2011 International Conference (Dubrovnik, Croatia). Cambridge Scholars Publishing, Newcastle., UK: P. 63-74

13. Machonis, P.A. 2010. English Phrasal Verbs: from Lexicon-Grammar to Natural Language Processing. In Southern Journal of Linguistics 34,1. P.21-48.
14. Machonis P.A. 2012. Sorting NooJ out to take Multiword Expressions into account. Formalising Natural Languages with NooJ: Selected Papers from the NooJ 2011 International Conference (Dubrovnik, Croatia). Edited by Kristina Vučković, Božo Bekavac and Max Silberztein. Cambridge Scholars Publishing, Newcastle., UK: P. 152-165
15. Sag, I.A., Baldwin, T., Bond, F., Copestake, A. and Flickinger D. 2002. Multiword Expressions: A Pain in the Neck for NLP. In Proceedings of the Third International Conference on Computational Linguistics and Intelligen Text Processing, Springer-Verlag, London. P.1-15.
16. Todorova M. 2008. Morpho-Syntactic Properties of Bulgarian Verbal Idiomatic Expressions. In Proceedings of the 2007 International NooJ Conference.Edited by Xavier Blanco and Max Silberztein.Cambridge Scholars Publishing, Newcastle, UK: P. 273-279.
17. Wehrli, E. 1998. Translating Idioms, in Proceedings of the 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics: COLING/ACL-98, Montreal, Canada, P. 1388-1392.