



**HAL**  
open science

# Structural derivation and meaning extraction: a comparative study of French/Serbo-Croatian parallel texts

Cvetana Krstev, Dusko Vitas

► **To cite this version:**

Cvetana Krstev, Dusko Vitas. Structural derivation and meaning extraction: a comparative study of French/Serbo-Croatian parallel texts. Geoff Barnbrook; Pernilla Danielsson; Michaela Mahlberg. Meaningful Texts. The Extraction of Semantic Information from Monolingual and Multilingual Corpora, Continuum, pp.165-176, 2005, 0-8264-9181-2. hal-01866673

**HAL Id: hal-01866673**

**<https://hal.science/hal-01866673>**

Submitted on 3 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## 16

# Structural derivation and meaning extraction: a comparative study of French/Serbo-Croatian parallel texts

*Cvetana Krstev*  
*University of Belgrade and*

*Duško Vitas*  
*University of Belgrade*

### 1. SERBIAN VS. SERBO-CROATIAN

Serbian literary language is the national language of Serbs and Montenegrins: it originated from the language reform of Vuk Stefanović Karadžić in the first half of the nineteenth century. At the base of this reform is the reform of the alphabet and the adoption, in principle, of a phonetic orthography. The Cyrillic alphabet was the main alphabet of the reformed language, but today Serbian language equally uses the Latin alphabet<sup>ix</sup>. Its counterpart is Croatian literary language as a national language of Croats that originated from the reform conducted by the Illyrian Movement also in the first half of the nineteenth century. This reform was partly modified at the middle of the nineteenth century by acceptance of Vuk's principles. Croatian literary language uses the Latin alphabet exclusively. There are three main dialects in the vernacular language: shtokavian, kajkavian, and chakavian, named according to the different pronunciation of the pronoun *šta* (Engl. what). Both literary languages were, however, developed on the basis of the new-shtokavian dialect. The name Serbo-Croatian appears with the establishment of the Kingdom of Yugoslavia in 1918 and existed as a diasystem of Serbian and Croatian literary languages. Up to the 1960s, Serbo-Croatian was seen as a single language with two alphabets. From that time on, it is rather seen as a language with two variants – Serbian and Croatian. Since the breakup of Yugoslavia, the Croatian variant of Serbo-Croatian has been called Croatian literary language, and the Serbian variant Serbian literary language. The overlapping in the framework of this diasystem is, however, still present. Ljubomir Popović (1996, 2000) gave the more detailed description of the development of the Serbian language.

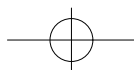
## 2. STRUCTURAL DERIVATION

The phenomenon of structural derivation in Serbo-Croatian consists of those derivational processes that induce change to the lexeme meaning in a predictable way. Traditional Serbo-Croatian lexicography determines the form of an entry according to its inflective features while the systematic character of structural derivation is not taken into consideration. For those entries that were derived through the process of structural derivation, the lexicographic description usually consists of the indication of the derivational process and the reference to the basic entry. However, this is not done in a systematic way. It could be said that the border between inflection and structural derivation is especially vague in the cases in which the structural derivation does not change the part of speech, such as derivation of diminutive and augmentative. But even in the cases in which the change of the part of speech occurs, such as derivation of possessive adjectives from nouns or gerunds from verbs, the semantic links between the basic entry and the derived entries remain stable.

As an example, we present in Table 1 the derivational forms of the nouns *glumac* (Engl. actor) and *baron* (Engl. baron), according to (RMS-MH 1967). This phenomenon enables the derivation from the basic entry of the forms for feminine gender, and respective diminutives, augmentatives and possessive adjectives. Traditional lexicography records only some of them—the shaded boxes in Table 1 are not described in (RMS-MH 1967). The lexicographic description for these missing forms, if produced by analogy, is given in the column 4 of the Table 1.

**Table 1. Structural derivation for nouns *glumac*, *baron***

glum + ac ( <i>actor</i> , [RMS-MH]: <i>the artist that acts in a drama</i> )			
baron + ε ([RMS-MH]: <i>noble title of lower degree</i> )			
type of process	base	suffix	definition
Gender motion	glum; baron	ica	<i>woman actor – (actress)</i>
Augmentative	glum	cyina	augmentative of <i>actor</i>
	baron	cyina	?augmentative of <i>baron</i>
Diminutive	glum;baron	cyicx	diminutive of <i>actor, baron</i>
	glum; baron	icyica	? diminutive of <i>actress, baroness</i>
Adjective Poss.	glum; baron	cyev; ov	belonging to <i>actor, baron</i>
	glum; baron	icyin	belonging to <i>actress, baroness</i>
	glum; baron	cyicxev	? belonging to dim. of <i>actor, baron</i>
	glum; baron	icyicin	? belonging to dim. of <i>actress, baroness</i>
Adjective Rel.	glum; baron	acyki; ski	relating to <i>actor, baron</i>
Adverb	glum; baron	acyki; ski	in the way of <i>actor, baron</i>



An analysis that included several dictionaries of Serbo-Croatian showed that at least 20–25 per cent of all dictionary entries is produced by this sort of derivation. As a consequence, the inconsistent treatment of structural derivation phenomenon implies at least one of the following problems: (a) redundant entries, (b) missing entries.

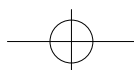
### 3. ELECTRONIC DICTIONARY AND STRUCTURAL DERIVATION

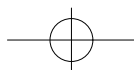
The system of electronic morphological dictionaries was originally constructed for French by Blandine Courtois. Its production was explained in full detail in (Courtois & Silberztein 1990; Courtois 1992; Gross & Perrin 1989). The basic part of this system consists of morphological dictionaries of simple words. The notion of simple word is here formally introduced as a character string between two consecutive separators, such as blanks, punctuation marks, etc. There are two different morphological dictionaries of simple words, known as DELAS and DELAF. The DELAS dictionary consists of simple lexical words represented in the form of the entries in traditional dictionaries: their infinitive form represents verbs while nouns are represented by their nominative singular form. To each lexical word a code is attached that uniquely describes its inflectional paradigm. Entries in DELAF are all inflectional forms of the lexical words presented in the corresponding DELAS. To each entry its corresponding uniquely coded lemma from DELAS is attached together with the sequence of codes representing its possible grammatical realisation. The entries for *baron* (Engl. *baron*) in DELAS and DELAF are represented in Table 2.

**Table 2. The entries for the noun *baron* in Serbian morphological dictionaries DELAS and DELAF**

DELAS	baron.N1
DELAF	baron,,N1:msn barona,baron.N1:msg:msa:mpg barone,baron.N1:msv:mpa baroni,baron.N1:mpn:mpv baronima,baron.N1:mpd:mpi:mpl baronom,baron.N1:msi baronu,baron.N1:msd:msl

The interpretation of the code N1 attached to lemma *baron* is that it is a noun that belongs to the morphological class 1 (Vitas 1993). The interpretation of the information *baron.N1:msv:mpa* attached to the textual word *barone* is that it can be either the vocative singular or accusative plural of the masculine noun *baron* from the morphological class 1.





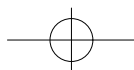
In the context of the construction of an electronic dictionary of Serbo-Croatian according to this model, the phenomenon of structural derivation raises several problems. The first problem arises from the way traditional lexicography determines the entries. If the results of structural derivation are considered as separate entries, then relations with derived entries having the predictable meaning are lost on the dictionary level. Also, automatic text indexing does not enable the identification of synonymous compounds, such as Adj. N and N N<sub>gen</sub>—*baronov sin* (Engl. baron's son) and *sin baron* (Engl. son of the baron). Second, if the processes of structural derivation are not described as exhaustively as the inflective processes, then automatic indexing by e-dictionaries identifies as unknown all those words that are not separate entries in a dictionary of DELAS type, but have origins in structural derivation. Contributions to the solving of problems that occur due to derivational morphology in the application of the system of electronic dictionaries were given by (Clemanceau 1993) and (Mota 1998), while the possible solutions of problems limited to the structural derivation were discussed in (Vitas and Krstev 1999). Third, the results of tagging by e-dictionaries of texts in different languages are difficult to compare if the phenomenon of structural derivation is not treated in the e-dictionaries of respective languages in the same or comparable way. Finally, the problem of maintenance of the system of the e-dictionary becomes a very complex task, if the phenomenon of structural derivation is not solved in a systematic way.

#### 4. AN EXPERIMENT

The investigation of the frequency of the phenomenon of structural derivation commenced with the following experiment. We chose texts from French classical literature for which there are multiple translations into Serbo-Croatian. From these texts a bilingual parallel corpus was produced. Using the language of the original (French in this case) as a meta-language, we observed how the entries of the original text were mapped to the target language in respect to the phenomenon of structural derivation.

As a first text for this experiment Voltaire's *Candide* was chosen, for which two independent translations in Serbo-Croatian exist\*. Both translations belong to the Serbian variant. The first one was published before the Second World War, and as done by the eminent translator Milan S. Nedić, who translated many classical works from English and French, such as Flaubert's *Madam Bovary* and Melville's *Moby Dick*. The other translation was done approximately thirty years later by Milan Predić who is less known for the translation of classical novels, although his translation conforms with contemporary language more than the first one.

Electronic versions of both texts were obtained in different ways. The French untagged text was obtained from *Association de Bibliophiles Universels* (<http://abu.cnam.fr/>), while both translations were retyped. All three texts were tagged according to TEI guidelines to the sentence level, with many phrase level elements tagged as well.



These texts were aligned using the Vanilla aligner (Danielsson and Ridings 1997), forming three pairs: French/Serbian (34), French/Serbian (63), and Serbian (63)/Serbian (34). The obtained results had to be manually corrected, as the sentence boundaries vary considerably from one text to another. The most prominent example is one French sentence aligned with eight sentences in Serbian translation. An example of the alignment for the pair French/Serbian (63) is given in Table 3.

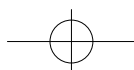
**Table 3. An example of French/Serbian (63) alignment** <sup>xi</sup>

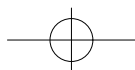
<p>*** Link: 1 – 2 ***</p> <p>&lt;seg id='KanFr.1.13.2.10'&gt;Candide dans l'instant tire la sienne et l'enfonce jusqu'à la garde dans le ventre du baron jésuite ; mais, en la retirant toute fumante, il se mit à pleurer : &lt;/seg&gt;.EOS</p> <p>&lt;seg id='Kan63.1.13.2.12'&gt; U tren oka Kandid izvu&amp;cy;e svoj ma&amp;cy; i zari ga baronu jezuiti do korica u trbuh. &lt;/seg&gt;&lt;seg id='Kan63.1.13.2.13'&gt; Ali kad je izvla&amp;cy;io o&amp;sx;tricu koja se jo&amp;sx; pu&amp;sx;ila od tople krvi, on zaplaka: &lt;/seg&gt;.EOS</p> <p>*** Link: 1 – 3 ***</p> <p>&lt;seg id='KanFr.1.13.2.11'&gt;«Hélas ! mon Dieu, dit-il, j'ai tué mon ancien maître, mon ami, mon beau-frère ; je suis le meilleur homme du monde, et voilà déjà trois hommes que je tue ; et dans ces trois il y a deux prêtres.» &lt;/seg&gt;.EOS</p> <p>&lt;seg id='Kan63.1.13.2.14'&gt; «Ah, bo&amp;zx;e moj!—uzviknu. &lt;/seg&gt; &lt;seg id='Kan63.1.13.2.15'&gt;—Ubio sam svog biv&amp;sx;eg gospodara, svog prijate&amp;l;x;a, svog &amp;sx;uraka. &lt;/seg&gt;&lt;seg id='Kan63.1.13.2.16'&gt;Nema bo&amp;l;x;eg &amp;cy;oveka na svetu od mene, a evo ubio sam ve&amp;cx; tri &amp;cy;oveka i me&amp;dx;u tom trojicom dvojica su sve&amp;sx;tenici.» &lt;/seg&gt;.EOS</p> <p>.EOP</p>	<p>Candide in an instant drew his rapier and plunged it up to the hilt in the Jesuit's body; but in pulling it out reeking hot, he burst into tears. "Good God!" cried he, "I have killed my old master, my friend, my brother-in-law. I am the best man in the world, and yet I have already killed three men, and of these three, two were priests."</p>
--	--

## 5. CONCORDANCES

The lemmatised concordances of French text were produced by the Intex system (Silberztein 1993). For instance, the entry <baron>, which corresponds to the inflective class N/A41 according to the LADL enumeration and representing the regular expression (n/ms + nne/fs + ns/ms + nnes/fs), has 56 occurrences in the text. A short extract from these concordances is given in Table 4.

If the lemmatisation of one of the translated texts is done according to the traditional Serbo-Croatian lexicographic practice, then the entry <baron> will encompass only the inflective forms of the noun *baron* listed in Table 2.



**Table 4. Keyword: <baron>, frequency=56, class: N/A41**

<seg id='KanFr.1.1.3.3'>	grasse, appétissante. Le fils du	<b>baron</b> paraissait en tout digne de son père.
<seg id='KanFr.1.27.9.1'>	Candide embrassa cent fois le	<b>baron</b> et Pangloss. « Et comment ne vous
<seg id='KanFr.1.4.1.19'>	qui voulait la défendre ; madame la	<b>baronne</b> a été coupée en morceaux ; mon
<seg id='KanFr.1.10.3.10'>	sur moi ; ajoutez que je suis née	<b>baronne</b> avec soixante et douze quartiers,
<seg id='KanFr.1.1.4.2'>	châteaux et madame la meilleure des	<b>baronnes</b> possibles. « Il est démonstré,
<seg id='KanFr.1.11.1.3'>	à lequel tous les châteaux de vos	<b>barons</b> allemands ne auraient pas servi de
<seg id='KanFr.1.27.8.3'>	Thunder-ten-tronckh, un des premiers	<b>barons</b> de l'Empire, et de M. Pangloss, le

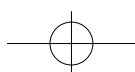
In lemmatised concordances of both translations, the entry <baron> that corresponds to the above extract from the e-dictionary has 44 in the translation from 1934 and 45 in the translation from 1963. The differences between frequencies in French and Serbian texts are due to the phenomenon of structural derivation. The French class <baron> is mapped to the Serbo-Croatian entries <baron>:Nm; <baronica>:Nf; <baronov>:Adj(Nm). The French form <baronette> which is not included in the class N/A41 is translated in both Serbian texts by <baroni-ica>:Dem(Nf) (diminutive of *baronica*). The excerpt from the lemmatised concordances of the Serbian translation from 1963 is given in Table 5.

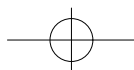
**Table 5. Serbian keyword: <baron>, frequency=45, class: N1**

<seg id='Kan63.1.27.9.1'>	vazduh proseca. Kandid zagrlj	<b>barona</b> i Panglosa po stotinu puta.
<seg id='KanFr.1.11.1.3'>	zamci va&sx;ih snova nema&cy;kih	<b>barona</b> ne bi mogli biti ni ko&nx;u&sx;ni
<seg id='KanFr.1.27.8.5'>	Tunder-ten-tronka, jednog od prvih	<b>barona</b> carstva, i za gospodina Panglosa,

The following three examples illustrate this phenomenon. To the French entry *baron*, which appears in the original text in the sentence identified by KanFr.1.1.3.3, in both translations corresponds the possessive adjective *baronov*: *le fils du baron* is either *baronov* sin or *sin baronov*, as can be seen in Table 6.

In contrast to the previous example, to the occurrence of the French entry *baron*, which appears in the original text in the sentence identified by KanFr.1.27.7.1, corresponds the noun *baron* in one translation and the possessive adjective *baronov* in the other, as shown in Table 7.





**Table 6. Mapping of a French noun *baron* to the Serbian possessive adjective *baronov***

<p>French/Serbian (63)            *** Link: 1 – 1 ***            Le fils du <b>baron</b> paraissait en tout digne de son père. .EOS            Sin <b>baronov</b> izgledao je u svemu dostojan svog roditelja. .EOS</p>
<p>Serbian(34)/Serbian (63)            *** Link: 1 – 1 ***  <b>Baronov</b> je sin izgledao u svemu dostojan svoga oca. .EOS            Sin <b>baronov</b> izgledao je u svemu dostojan svog roditelja. .EOS</p>
<p>The Baron's son seemed to be a youth in every respect worthy of the father he sprung from.</p>

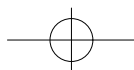
**Table 7. Mapping of a French noun *baron* to the Serbian noun *baron* and to the Serbian possessive adjective *baronov*.**

<p>French/Serbian (63)            *** Link: 1 – 1 ***            Au nom du <b>baron</b> et de Pangloss les deux forçats poussèrent un grand cri , s'arrêtèrent sur leur banc et laissèrent tomber leurs rames..EOS            Pri pomenu <b>baronova</b> i Panglosova imena, oba roba uzviknu&amp;#x27E8;e iz sveg glasa, zasta&amp;#x27E8;e nepomi&amp;#x27E9;ni na klupama i ispusti&amp;#x27E8;e vesla. .EOS</p>
<p>Serbian(34)/Serbian (63)            *** Link: 1 – 1 ***            Pri pomenu <b>barona</b> i Panglosa, ova dva roba vrisnu&amp;#x27E8;e, prestado&amp;#x27E8;e veslati sa svoje klupe i upusti&amp;#x27E8;e vesla. .EOS            Pri pomenu <b>baronova</b> i Panglosova imena, oba roba uzviknu&amp;#x27E8;e iz sveg glasa, zasta&amp;#x27E8;e nepomi&amp;#x27E9;ni na klupama i ispusti&amp;#x27E8;e vesla. .EOS</p>
<p>No sooner had Candide uttered the names of the Baron and Pangloss, than the two slaves gave a great cry, ceased rowing, and let fall their oars out of their hands.</p>

Similarly, the French compound <le roi des Bulgare> appears in the sentence KanFr.1.2.2.10. It is mapped in the text from 1934. to <kralj bugarski>, where *bugarski* is a relative adjective, and in text from 1963. to <kralj Bugara>, where *Bugarin* is a noun (Table 8).

The given examples show that in order to improve the extraction of meaning it is necessary to expand the inflective class so that it encompasses the cases of structural derivation too. This approach means that one textual word (word form) can be reduced to more lemma forms, depending to what level structural derivation has been included. Thus, the Serbo-Croatian entry <baron> can be considered in different ways:





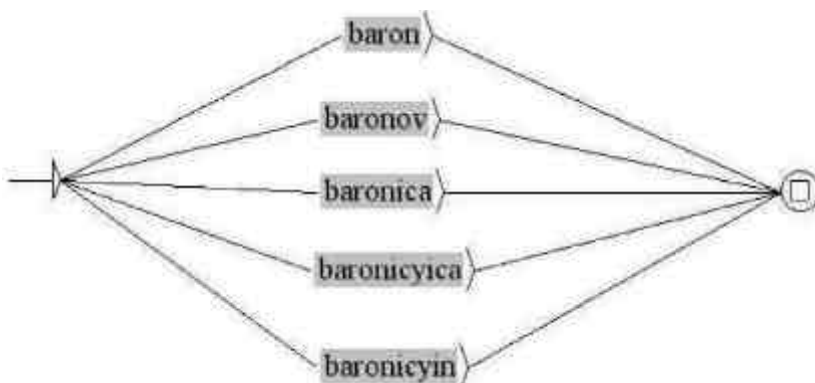
**Table 8. Mapping of a French compound *le roi des Bulgares* to the Serbian compound *kralj Bugara* and *bugarski kralj***

<p>French/Serbian (63)          *** Link: 1 – 1 ***          Il avait déjà un peu de peau et pouvait marcher, quand le <b>roi des Bulgares</b> livra bataille au roi des Abares. .EOS          Ve&amp;cx; be&amp;sx;e dobio malo ko&amp;zx;e i mogao je hodati, kad <b>kralj Bugara</b> zametnu boj sa kraljem Abara. .EOS</p>
<p>Serbian(34)/Serbian (63)          *** Link: 1 – 1 ***          Ve&amp;cx; je imao pomalo ko&amp;zx;e i mogao je i&amp;cx;i, kad se <b>kralj bugarski</b> upusti u bitku s kraljem Abara. .EOS          Ve&amp;cx; be&amp;sx;e dobio malo ko&amp;zx;e i mogao je hodati, kad <b>kralj Bugara</b> zametnu boj sa kraljem Abara. .EOS</p>
<p>His sores were now skimmed over and he was able to march, when the King of the Bulgarians gave battle to the King of the Abares.</p>

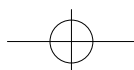
- (a) <baron> which corresponds to Serbo-Croatian DELAS entry baron.N01.01;  
 (b) <baron> which assembles Serbo-Croatian DELAS entries baron.N01.01 and baronica,N71.01, as in French;  
 (c) <baron> which assembles, besides (b) the corresponding forms of diminutives, augmentatives, possessive adjectives, etc.

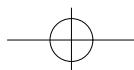
## 6. THE IMPROVEMENT OF SERBIAN CONCORDANCES

The lemmatisation with the entries augmented in the described way can be implemented in the Intex system by finite automata whose every node corresponds to the appropriate e-dictionary segment. Such an automaton for the Serbian entry <baron> is presented on the Figure 1.



**Figure 1. Finite automaton BaronPlus.grf**





This automaton covers the inflective paradigms of separate entries, as seen in traditional lexicography, into one entry named *BaronPlus.grf*. This new entry neutralises the effects of structural derivation. In Serbo-Croatian DELAF dictionary 56 different word forms correspond to the elements of the inflective paradigms of five separate entries from this automaton. Some of these word forms are:

barona,baron.N1:msg:msa:mpg  
 baronycicxe,baronycix.N21:mpa  
 baronicama,baronica.N71:fpd:fpi:fpl  
 baronycyice,baronycyica.N71:fsg:fsv:fpn:fpa:fpv  
 baronycyine,baronycyin.A2:qnmpa:qnfsg:qnfpn:qnfpa:qnfpv  
 baronove,baronov.A2:qnmpa:qnfsg:qnfpn:qnfpa:qnfpv  
 baronskim,baronski.A3:qomsi:qonsi:qompd:qompi:qompl:qonpd:qonpi:qonpl:qofpd:qofpi:qofpl

When the lemmatised concordances of translated text are produced using the given automaton, then the frequencies of the entry <baron> in the source and target texts become approximately equal. Lemmatisation of translations by Intex using the automaton *BaronPlus.grf* instead of <baron> yields 56 occurrences in text from 1934 and 57 occurrences in text from 1963. The excerpt from concordances produced in this way is given in Table 9.

**Table 9. Serbian keyword: *BronPlus.grf* (FS graph), frequency=57**

<seg id='Kan63.1.27.9.1'>	vazduh proseca. Kandid zagrlj	<b>barona</b> i Panglosa po stotinu puta.
<seg id='KanFr.1.11.1.3'>	zamci va&sx;ih snova nema&cy;kih	<b>barona</b> ne bi mogli biti ni ko&nx;u&sx;ni
<seg id='KanFr.1.27.8.5'>	Tunder-ten-tronka, jednog od prvih	<b>barona</b> carstva, i za gospodina Panglosa,
<seg id='KanFr.1.4.1.23'>	kad je hteo da je brani; gospo&dx;a	<b>baronica</b> je ise&cy;ena na komade; moj
<seg id='KanFr.1.10.3.14'>	Dodajte jo&sx; da sam se rodila	<b>baronica</b> sa sedamdeset dva pretka, a da
<seg id='KanFr.1.1.4.2'>	najbo&lx;a od svih mogu&cx;ih	<b>baronica</b> . "Dokazano je," govorio je on
<seg id='Kan63.1.1.3.3'>	i naoko primam&lx;iva. Sin	<b>baronov</b> izgledao je u svemu dostojan
<seg id='Kan63.1.2.1.1'>	skrivao u sebi najlep&sx;u od svih	<b>baroni&amp;cy;ica</b> . Le&zx;e da spava, bez

The exhibited procedure was applied to some other entries of source and target texts and appropriate lemmatised concordances produced. Their respective frequencies in the original text and the translation from 1963 are presented in

Table 10. In the column *SC class* the frequency of traditionally conceived Serbo-Croatian entries is given, in the column *Serbian FSA* the frequency of the entry described by the appropriate local grammar, and in the column *French Class* the frequency of the corresponding entry in the source text. The insignificant differences in frequencies between French lemma and Serbo-Croatian lemma defined by local grammar can have different sources. Some are not the consequences of the structural derivation at all (for instance, the French *actrice* is once translated by *pevacyica* instead of the direct translation *glumica*), and some are the results of different levels of involvement of this phenomenon in applied classes (for instance, the French diminutive *baronnette* is not in the class <baron>, while the Serbian counterpart *baronicyica* is in BaronPlus.grf).

**Table 10. Frequencies of some French entries and respective Serbian traditional and new entries**

English	SC class	F	Serbian FSA	F	French class	F
( <i>baron</i> )	<baron> Nm	45	BaronPlus.grf	57	<baron> N/A	56*
( <i>Bulgarian</i> )	<Bugarin> Nm	17	BugarinPlus.grf	30	<bulgaire> N/A	30
( <i>Christian</i> )	<hrisxcxanin> Nm	4	HristPlus.grf	7	<chrétien> N/A	7
( <i>actor</i> )	<glumac> Nm	4	GlumacPlus.grf	7	<acteur> N	8*
( <i>canon</i> )	<top> Nm	2	TopPlus.grf	6	<canon> Nm	6
( <i>Pope</i> )	<papa> Nm	4	PapaPlus.grf	6	<pape> Nm	6
( <i>lover</i> )	<lhubavnik> NmZ	3	LhubavnikPlus.grf	5	<amant> N	6*

The further improvement in neutralising the influence of structural derivation on the results of lemmatisation can be achieved if the automaton in Figure 1 is replaced by the finite transducer given in Figure 2 (Gross 1997). If a transducer is used in the process of lexical recognition to the textual words, the string attached to the graph node is associated instead of a dummy entry BaronPlus. For instance, to the form *baronice*, which in Serbo-Croatian DELAS can represent genitive singular of class <baronica>, lemma *baron* is associated with a semantic mark of feminine gender by gender motion.

## 7. CONCLUSION

The traditional definition of the inflective classes in lexicography of Serbo-Croatian cannot link the entries between which there exists a close structural derivation relationship. This relationship is on the dictionary level usually expressed by a stereotype definition that is based on the meaning of the basic entry. Several types of nominal modifiers of this kind are discussed in this paper: modifiers for

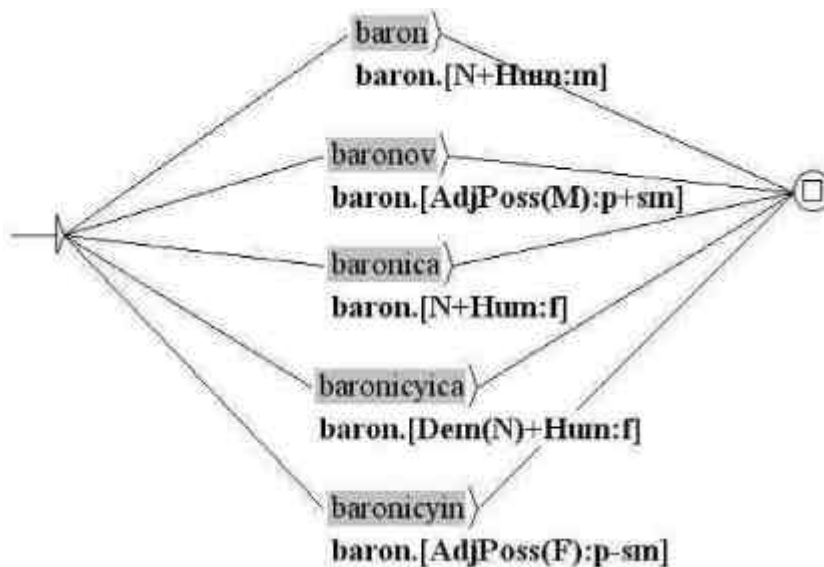
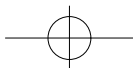


Figure 2. Transducer BaronPlus

the gender motion, modifiers for amplification (formation of diminutives and augmentatives), and modifiers for derivation of adjectives from nouns. It is shown that both modifiers form the new noun from the initial one according to the same model and in a similar way the inflective classes are described. The extension of the traditional nominal class can thus be seen as a tuple of the elementary classes that describe the derivatives of the basic entry.

## REFERENCES

- Courtois, Blandine & Silberztein, Max (eds).1990. Dictionnaires électroniques du français. *Langue française* 87. Paris: Larousse.
- Courtois, Blandine. 1992. *Dictionnaire électroniques des mots simples du français DELAS V07-E1*. Paris: LADL.
- Clemanceau, David. 1993. *Structuration du lexique et reconnaissance de mots dérivés*, Thèse de doctorat d'informatique fondamentale, Université Paris 7, LADL, CERIL.
- Gross, Maurice & Perrin, Dominique (eds.). 1989. *Electronic Dictionaries and Automata in Computational Linguistics*, Lecture Notes in Computer Science. Berlin: Springer Verlag.
- Gross, Maurice. 1997. "The Construction of Local Grammars". In Roche, Emmanuel & Yves Schabes, (eds.) *Finite State Language Processing*, Cambridge, Mass.: MIT Press.
- Danielsson, Pernilla & Ridings, Daniel. 1997. *Practical Presentation of a "Vanilla" Aligner*, Presentation held at the TELRI Workshop in Alignement and Exploitation of Texts, Ljubljana, February, 1–2
- Mota, Cristina. 1998/1999. "Enhancing the INTEX Morphological Parser with Lexical Constraints". In Shevalier, Jean-Claude, Maurice Gross & Christiane Leclere (eds). *Lingvisticae Investigationes*. Amsterdam: John Benjamins Publishing Company, 413–423



- Popović, Ljubomir. 1996. "Deux approches idéologique de la vernacularisation de la langue littéraire chez les Serbs à la fin du 18e et dans la première moitié du 19e siècle". In: *Langues et nation en Europe Centrale et Oriental du 19e siècle à nos iours*. Cahiers de IÖILSL, No.8: 209–240. Lausanne.
- Popović, Ljubomir. 2000. "From Standard Serbian to Standard Serbo-Croatian to Standard Serbian". In: *Proceedings of the Conference 'Languages in the Former Yugoslav Lands* (SSEES, London, 8–9 Septemebr 2000) (to appear)
- RMS-MH. 1967. *Rečnik srpskohrvatskoga književnog jezika*, vol. 1–6, Beograd-Zagreb: Matica Srpska, Matica Hrvatska.
- Silberztein, Max D. 1993. *Le dictionnaire électronique et analyse automatique de textes: Le système INTEX*, Paris:Masson.
- Vitas, Duško. 1993. *Matematički model morfologije srpskohrvatskog jezika (imenska fleksija)*, Beograd: Matematički fakultet, Univerzitet u Beogradu.
- Vitas, Duško, Krstev, Cvetana & Pavlović-Lažetić, Gordana. 2001. "Flexible Dictionary Entry". In Gerhild Zybatow (ed). *Current Issues in Formal Slavic Linguistics*. Frankfurt am Main: Peter Lang, 461–468.

