

Verb Valency Enhanced Croatian Lexicon

Kristina Vučković, Nives Mikelić Preradović, Zdravko Dovedan
kvuckovi@ffzg.hr, nmikelic@ffzg.hr, zdovedan@ffzg.hr
Faculty of Humanities and Social Sciences
University of Zagreb
Department of Information Sciences
Ivana Lucica 3, Zagreb, Croatia



The Plan

- **Our agenda?**
 - Increase # of unambiguous NPs
- **By means of?**
 - Existing chunker
 - Verb valency tags
- **Why?**
 - To raise the chunker performance to a higher level
 - Make preparations for a Croatian parser



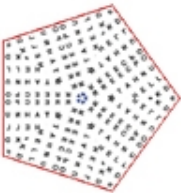
Overview

- Croatian verb valency lexicon
 - main characteristics
 - selected data
- .xml to .dic conversion
 - how we did it
- previous grammars for
 - <VP> | <NP> | <PP> selection
- new enhanced grammars
 - <VP+DCobl>
 - <VP+PCobl>
 - <VP+PCtyp>
- results comparison
 - precision,
 - recall,
 - f-measure



Croatian verb valency lexicon - CROVALLEX

- Formal description of verb valency frames
- 1739 verbs
 - selected from the Croatian frequency dictionary, 1999.
- 5118 valency frames (in average: 3 frames per verb)
- Each frame entry contains descriptions of
 - valence frame
 - frame attributes
 - frame attributes are either obligatory or optional i.e. obligatory or typical!



Selected data

1. Reflexive particle 'se'

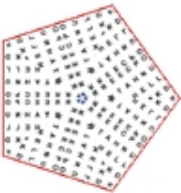
- if the verb is derived reflexive (e.g. *vratiti se*)
- reflexiva tantum (e.g. *smijati se*).



Selected data

1. Pure (prepositionless) case.

- 7 morphological cases in Croatian.
 - 0 - hidden nominative,
 - 1 - nominative,
 - 2 - genitive,
 - 3 - dative,
 - 4 - accusative,
 - 5 - vocative,
 - 6 - locative,
 - 7 - instrumental.

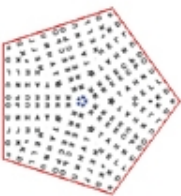


Selected data

1. Prepositional case.

- Lemma of the preposition and
- number of the required morphological case are specified, e.g.
 - od+2,
 - na+4,
 - o+6

CROVALLEX 2.0008 - *.xml



rmjt.ffzg.hr

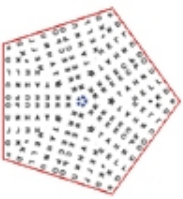
NooJ2008
Budapest
2008-06-08

```
firstobject XML Editor - [C:\Documents and Settings\Wiki\Desktop\Dovedan_konverter\CRVLL...
File Edit View Tools Window Help

<word_entry aspect="inf" freq="326">
  <headword_lemma>pjevati (pjevati)</headword_lemma>
  <frame_entry frame_index="1">
    <ID_valency>5616</ID_valency>
    <ID_accent>604</ID_accent>
    <frame_slots>
      <slot functor='AGT' type='obl'>
        <form type='direct_case' case='0_or_1'/>
      </slot>
      <slot functor='MANN' type='typ'>
        <form type='adverb' adverb_lemma='lijepo'/>
      </slot>
    </frame_slots>
    <gloss>glasati se ugodnim glasovima - o pticama pjevicama i nekim kukcima</gloss>
    <example>Pučice lijepo pjevaju</example>
    <class>speak_manner</class>
  </frame_entry>
  <frame_entry frame_index="2">
    <ID_valency>5618</ID_valency>
    <ID_accent>604</ID_accent>
    <frame_slots>
      <slot functor='AGT' type='obl'>
        <form type='direct_case' case='0_or_1'/>
      </slot>
      <slot functor='PAT' type='obl'>
        <form type='prepos_case' prepos_lemma='o' case='6'/>
      </slot>
    </frame_slots>
    <gloss>slaviti, veličati</gloss>
    <example>On pjeva o zgaženim i desetkovanim Rakocijevim buntovnicima</example>
    <class>judgement/praise</class>
  </frame_entry>
</word_entry>
```

Converting to *.dic

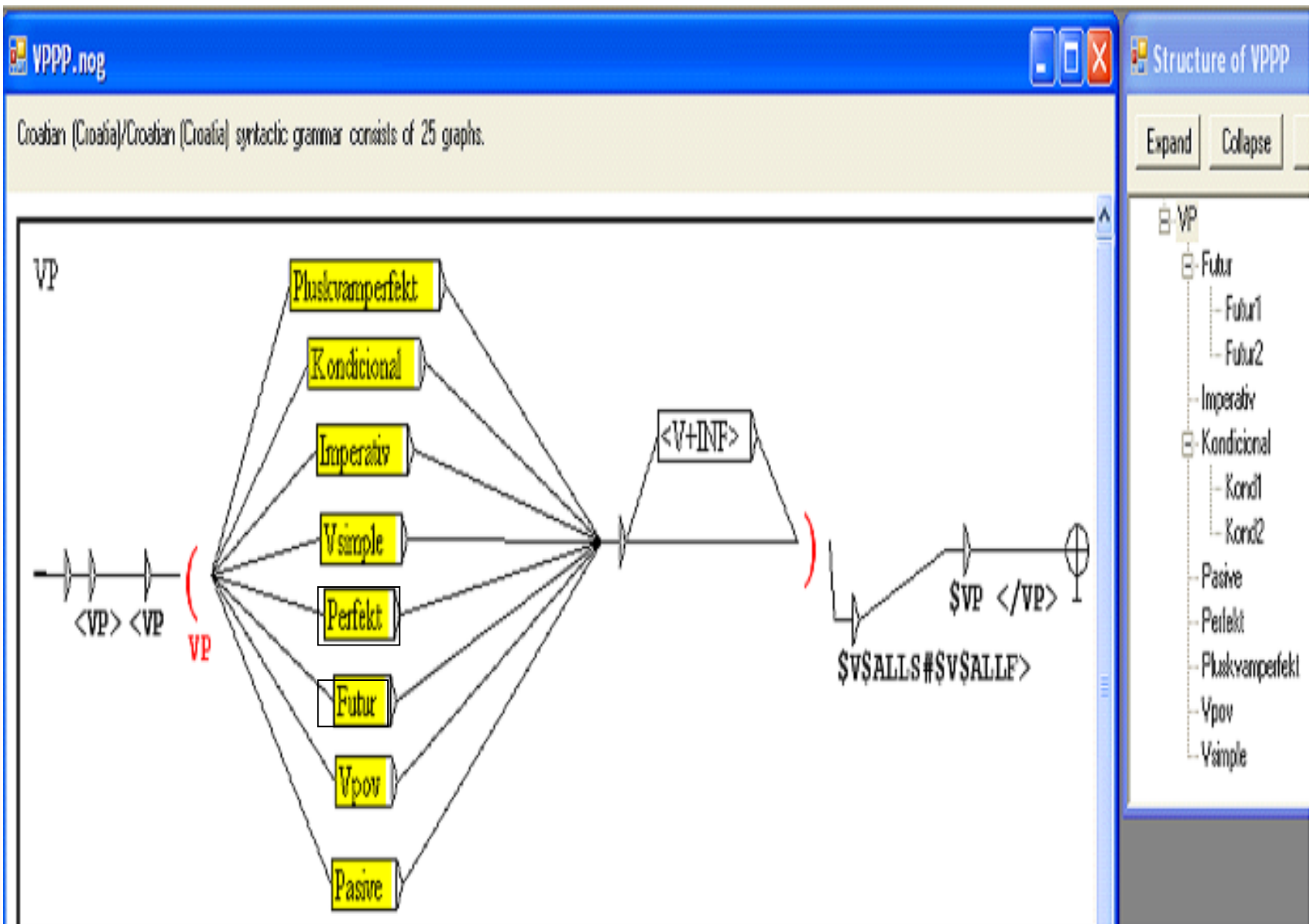
| Open XML File | Open DIC File | Match Lines | Non-matching lines |
|---|--|--|--------------------|
| zvoniti, aspect=inf+DC_obl=0+PC_typ=6+PC_typ=4+ zvučati, aspect=inf+DC_obl=0+AL_obl+DC_typ=3+A žaliti, aspect=inf+reflexive=YES+DC_obl=0+DC_obl= žaliti, aspect=inf+DC_obl=0+PC_typ=4 žariti, aspect=inf+reflexive=YES+DC_obl=0+DC_typ= žderati, aspect=inf+reflexive=YES+DC_obl=0+SC_typ žeći, aspect=inf+DC_obl=0+DC_obl=4+AL_typ+AL_t željeti, aspect=inf+DC_obl=0+DC_obl=4+DC_typ=3 žeti, aspect=inf+DC_obl=0+DC_obl=4+PC_typ=6 živjeti, aspect=inf+DC_obl=0+PC_typ=6+AL_typ+PC žmirkati, aspect=inf+DC_obl=0+AL_typ+PC_typ=2+P žrtvovati, aspect=dual+reflexive=YES+DC_obl=0+DC žuboriti, aspect=inf+DC_obl=0+PC_typ=2+PC_typ=4 žudjeti, aspect=inf+DC_obl=0+DC_obl=4+PC_obl=7+ zuriti, aspect=inf+reflexive=YES+DC_obl=0+DC_typ= | zijeivnuti, V+FLX=BLJESNUTI značiti, V+FLX=ŽAGORITI znati, V+FLX=ZNATI zračiti, V+FLX=BILJEŽITI zrcaliti, V+FLX=BRANITI zreti, V+FLX=ZRETI zvati, V+FLX=ZVATI zvučati, V+FLX=BOJATI žaliti, V+FLX=CIVILITI žalostiti, V+FLX=GOSTITI željeti, V+FLX=VOLJETI živjeti, V+FLX=ŽIVJETI žuboriti, V+FLX=ŽAGORITI žuriti, V+FLX=ŽAGORITI žvakati, V+FLX=ŽVAKATI | zavrtjeti, V+FLX=UGOJITI+aspect=fin+reflexive=YES+DC_obl=0+DC_ zaustaviti, V+FLX=CJEPITI+aspect=fin+reflexive=YES+DC_obl=0+DC_ zaustaviti, V+FLX=CJEPITI+aspect=fin+DC_obl=0+PC_typ=6+PC_typ zaustavljati, V+FLX=BIRATI+aspect=inf+reflexive=YES+DC_obl=0+DC_ zauzeti, V+FLX=UZETI+aspect=fin+reflexive=YES+DC_obl=0+DC_obl zauzeti, V+FLX=UZETI+aspect=fin+DC_obl=0+PC_obl=4 zauzimati, V+FLX=BIRATI+aspect=inf+reflexive=YES+DC_obl=0+DC_ zavesti, V+FLX=POJESTI+aspect=fin+DC_obl=0+DC_obl=4+PC_typ= zavladati, V+FLX=ZASJATI+aspect=fin+DC_obl=0+PC_typ=2+DC_obl završavati, V+FLX=BIRATI+aspect=inf+DC_obl=0+DC_typ=7+PC_obl završiti, V+FLX=UGOJITI+aspect=fin+DC_obl=0+PC_obl=6+DC_typ=7 zaželjeti, V+FLX=ZAŽELJETI +aspect=fin+reflexive=YES+DC_obl=0+I zgrabiti, V+FLX=CJEPITI+aspect=fin+DC_obl=0+PC_typ=4+DC_obl=4 zijeivati, V+FLX=SJATI+aspect=inf+DC_obl=0+PC_typ=2+PC_typ=6 značiti, V+FLX=ŽAGORITI+aspect=inf+DC_obl=0+DC_obl=4+PC_typ= znati, V+FLX=ZNATI+aspect=inf+DC_obl=0+SC_obl+DC_obl=4+PC_c zračiti, V+FLX=BILJEŽITI+aspect=inf+DC_obl=0+DC_obl=4+PC_typ=2 zreti, V+FLX=ZRETI+aspect=inf+DC_obl=0+PC_typ=6 zreti, V+FLX=ZRETI+aspect=inf+DC_obl=0+PC_typ=4+DC_typ=3 zvati, V+FLX=ZVATI+aspect=inf+reflexive=YES+DC_obl=0+DC_obl=4 zvučati, V+FLX=BOJATI+aspect=inf+DC_obl=0+AL_obl+DC_typ=3+A žaliti, V+FLX=CIVILITI +aspect=inf+reflexive=YES+DC_obl=0+DC_obl= žaliti, V+FLX=CIVILITI +aspect=inf+DC_obl=0+PC_typ=4 željeti, V+FLX=VOLJETI+aspect=inf+DC_obl=0+DC_obl=4+DC_typ=3 živjeti, V+FLX=ŽIVJETI +aspect=inf+DC_obl=0+PC_typ=6+AL_typ+PC žuboriti, V+FLX=ŽAGORITI+aspect=inf+DC_obl=0+PC_typ=2+PC_typ= žuriti, V+FLX=ŽAGORITI+aspect=inf+reflexive=YES+DC_obl=0+DC_ | |



Croatian lexicon

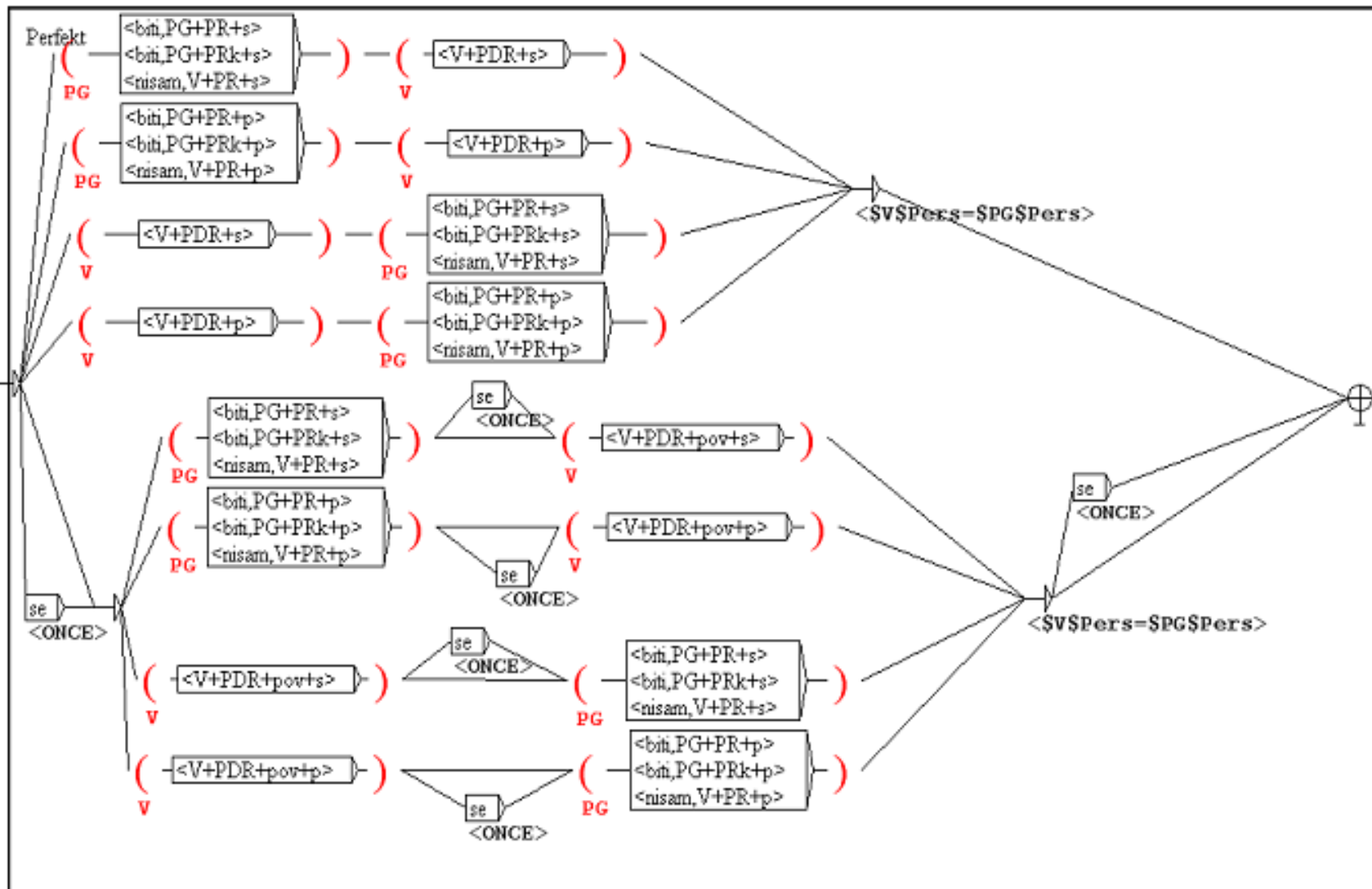
- **Nouns - 19 600**
- **Adjectives - 13 899**
- **Verbs - 1133**
- **Adverbs - 535**
- **Proper Nouns - 67 704**
- **S + C + Q + I + PRO - 363**

Previous grammars



Perfect

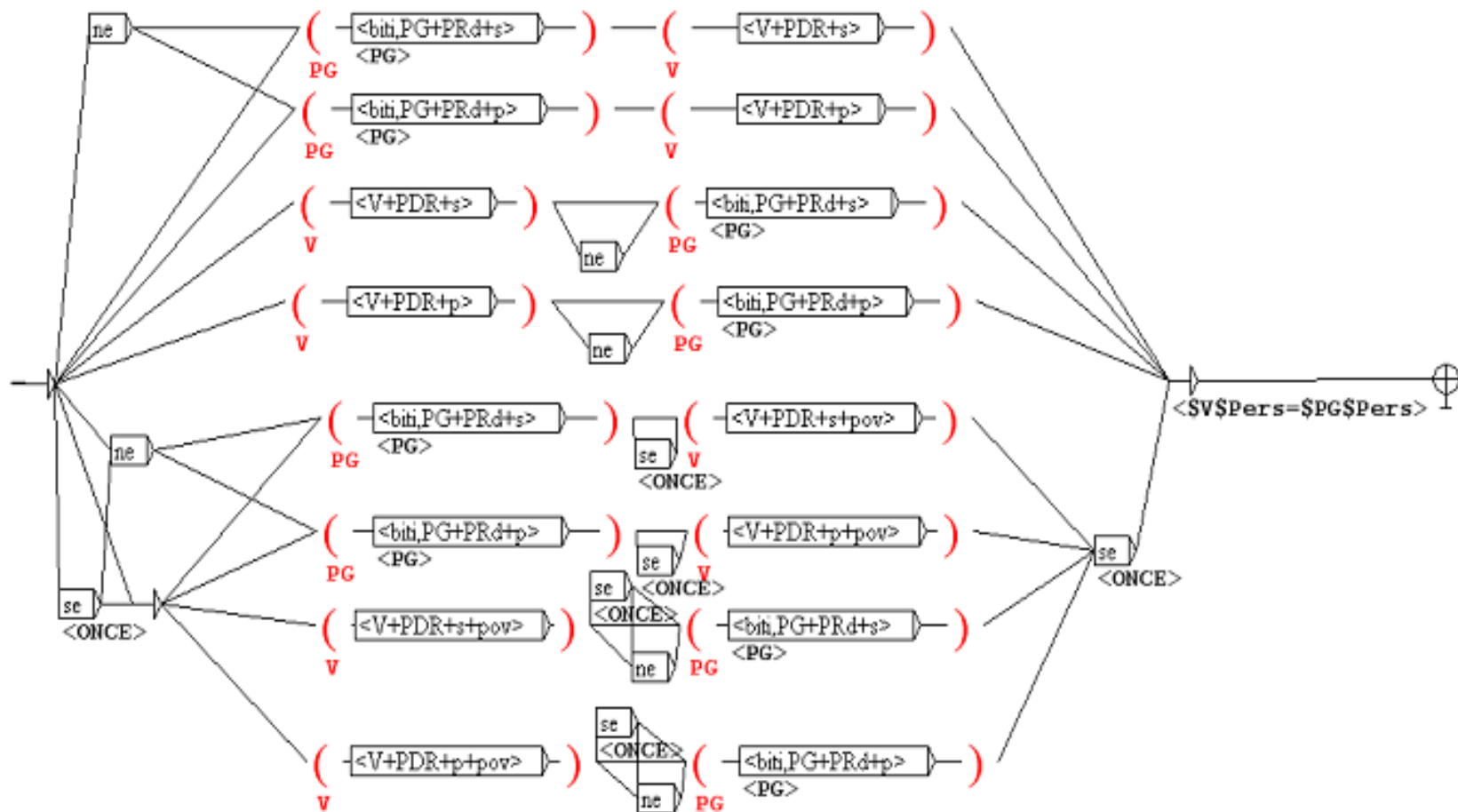
Croatian (Croatia)/Croatian (Croatia) syntactic grammar consists of 25 graphs.



II. Future

Croatian (Croatia)/Croatian (Croatia) syntactic grammar consists of 25 graphs.

Futur2



buđem pjevao; buđem se smijao; ne buđem pjevao; ne buđem se smijao

```

firstobject XML Editor - [xml1 *.]
File Edit View Tools Window Help
<S>
  <NP "A">
  <NP "N" Nb="s" Gender="m">Pakt</NP></NP> konačno
  <VP Nb="s" Pers="3" Gender="m">prohodao</VP>.
</S>
<S>
  <NP "A">
  <NP "N" Nb="s" Gender="m">Zagreb</NP></NP> nikada
  <VP Nb="s" Pers="3" Gender="m">moljaka</VP>
...
<S>U
  <NP "L">tim okvirima</NP>
  <NP "A">
  <NP "N" Nb="s" Gender="m">Bruxelles</NP></NP>
  <VP Nb="s" Gender="n" Pers="3">označava</VP>
...
  <VP>je</VP> još
  <NP "A">
  <NP "N" Nb="s" Gender="n">Graničevo ministarstvo</NP></NP>
  <VP Nb="s" Pers="3" Gender="n">pripremlilo</VP>
...
<S>
  <NP "G">
  <NP "N" Nb="s" Gender="f">Demokracija</NP></NP>
  <VP Nb="s" Pers="3" Gender="f">se isplati</VP>,
...
<S>
  <NP "G">
  <NP "N" Nb="s" Gender="f">konferencija</NP></NP>
  <VP FLX="UGOJITI" aspect="fin" DCobl="Acc" PCtyp="L" Tense="PDR" Nb="s" Pers="3" Gender="f">je polučila</VP>
  <NP "A">
  <NP "N" Nb="s" Gender="m">uspjeh</NP></NP>

```

```

firstobject XML Editor - [xml1 *.]
File Edit View Tools Window Help
<S>
  <NP "N" Nb="s" Gender="m">Pakt</NP> <R>konačno</R>
  <VP FLX="DODATI" aspect="fin" DCobl="Acc" Tense="PDR" Nb="s" Pers="3" Gender="m">prohodao</VP>
</S>
...
<S>
  <NP "N" Nb="s" Gender="m">Zagreb</NP> <R>nikada</R>
  <VP FLX="NISAM" aspect="dual" DCobl="O" Tense="PR" Nb="s" Gender="m" Pers="3">nije</VP><R>otvoreno</R>
  <VP FLX="BIRATI" aspect="inf" Type="pov" DCobl="Acc" Tense="PDR" Nb="s" Pers="3" Gender="m">moljaka</VP>
...
<S>
  <PP>U
  <NP "L">tim okvirima</NP></PP>
  <NP "N" Nb="s" Gender="m">Bruxelles</NP>
  <VP FLX="BIRATI" Aspect="inf" DCobl="Acc" Tense="IMT" Nb="s" Gender="n" Pers="3">označava</VP>
...
  <VP>je</VP> <Q>još</Q>
  <NP "N" Nb="s" Gender="n">Graničevo ministarstvo</NP>
  <VP FLX="CIJEPTI" Aspect="fin" Type="pov" DCobl="Acc" PCtyp="Acc" AL="typ" Tense="PDR" Nb="s" Pers="3" Gender="n">
...
<S>
  <NP "N" Nb="s" Gender="f">Demokracija</NP>
  <VP FLX="NAPATITI" Aspect="fin" Type="pov" DCobl="Acc" DCtyp="I" Tense="IMT" Nb="s" Gender="n" Pers="3">se isplati
...
<S>
  <NP "N" Nb="s" Gender="f">konferencija</NP>
  <VP FLX="UGOJITI" aspect="fin" DCobl="Acc" PCtyp="L" Tense="PDR" Nb="s" Pers="3" Gender="f">je polučila</VP><
  <NP "A" Nb="s" Gender="m">uspjeh</NP>

```


Verb + Obligatory DC

NooJ - [Concordance for Text hrCorpus_oznaceni_uzorci.not]

File Edit Lab Project Windows Info TEXT CONCORDANCE

Reset Display: 5 characters before, and 5 after. Display: Matches Results
 word forms

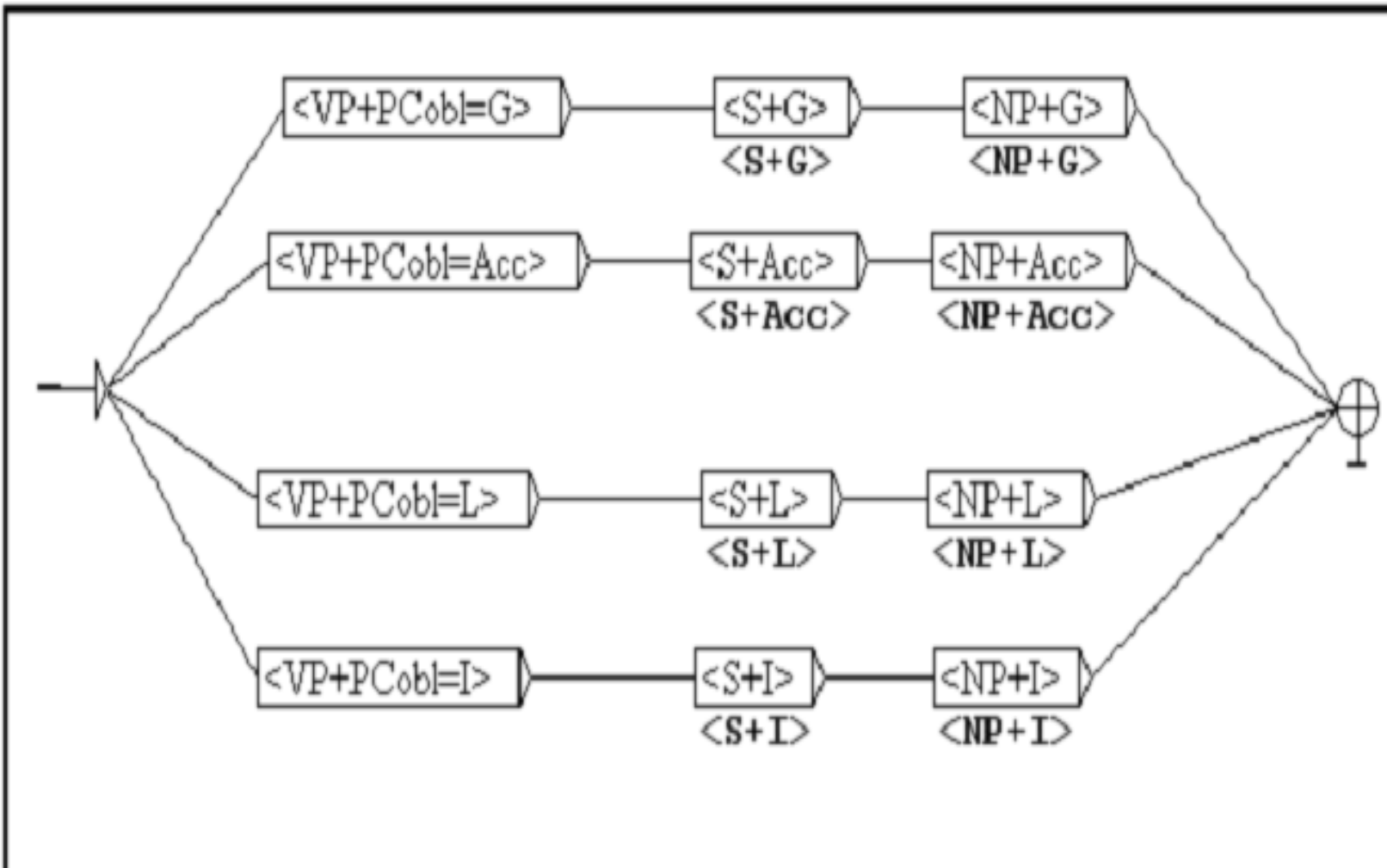
| Text | Before | Seq. | After |
|--|--|------|--|
| mandat na raspolaganje, istaknuvši da je Mesić. Premijer je istaknuo. " je do kratkih spojeva, jer Hrvata, ali među 50 tamošnjih policajaca Hrvata, ali među 50 tamošnjih policajaca jednom ponovili kako među njima Predsjednik države izjavio je da Mirka Galića o Kosovčevom razrješenju milijardi maraka, za koje danas obećanja koja se moraju ispuniti. službi postavlja Predsjednik Republike i politička smjena, kazao je Škrabalo. mandat. Glavni urednik HTV-a | ne osjeća krivnju/<NP+A> Ne vidim ozbiljne probleme/<NP+A> neki očito nisu shvatili/<NP+N> nema nijednog srpske nacionalnosti/<NP+A> nema nijednog srpske nacionalnosti/<NP+G> nema sukoba/<NP+G> nema sukoba/<NP+G> nije dobio natpolovičnu većinu/<NP+A> nitko ne zna/<NP+N> Nitko tada nije rekao/<NP+N> obećali smo sukreiranje/<NP+A> Obrad Kosovac je/<NP+N> Obrad Kosovac nije smijenjen/<NP+N> | | niti sram za ono što i sukobe između predsjednika, da je vrijeme predizborne kamp. Đukić je svjestan da za . Đukić je svjestan da za . Predsjednik države izjavio je d između njega, predsjednika Vla glasova prisutnih članova Vijeća gdje su i tko su da vojska može biti pod vanjske politike. Predsjednik R profesionalac koji je svoje profi na sjednici Vijeća HRT-a |

GRAM = Acp_OD 144/144

Cancel

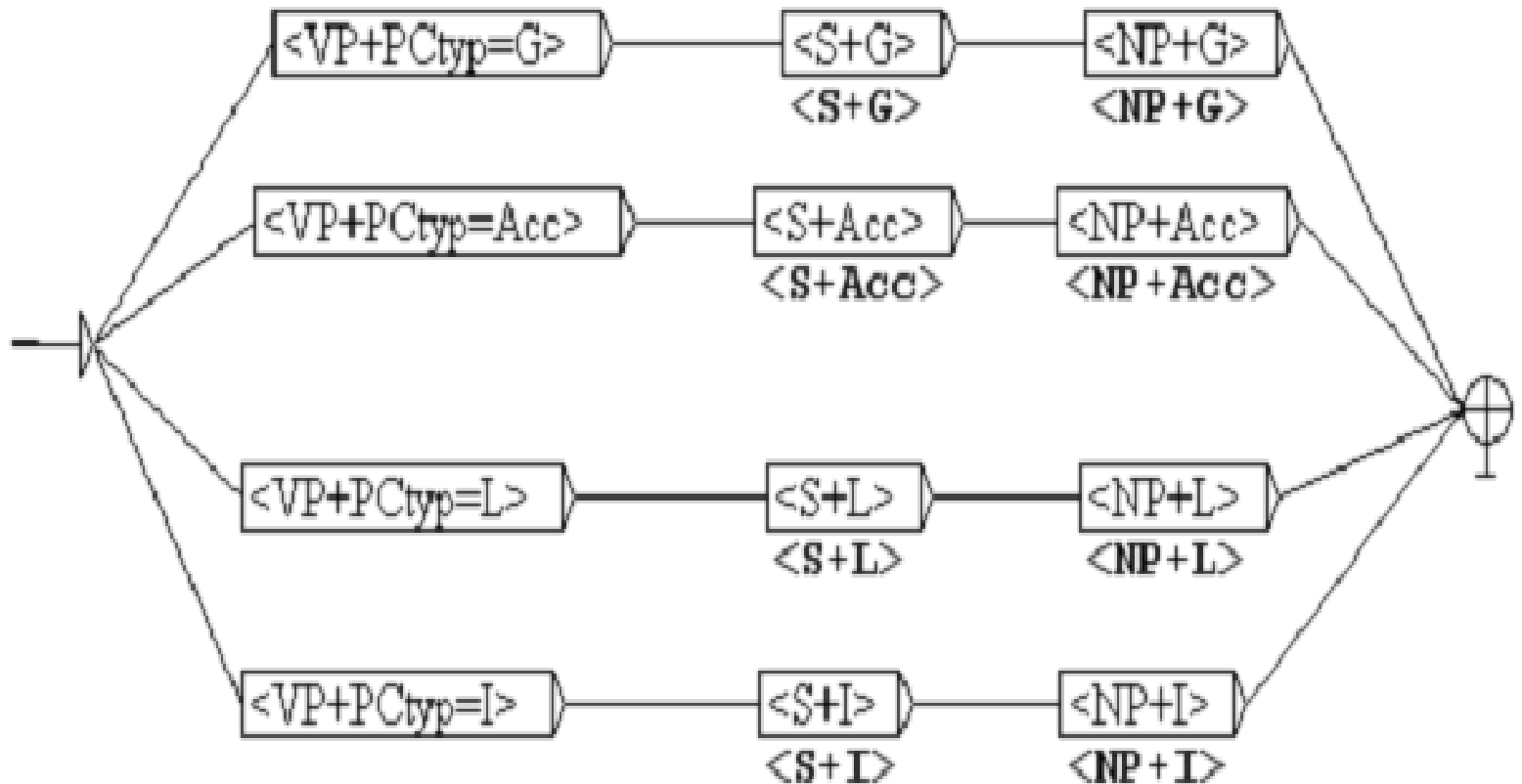
Verb + obligatory PC

Croatian (Croatia)/Croatian (Croatia) syntactic grammar.



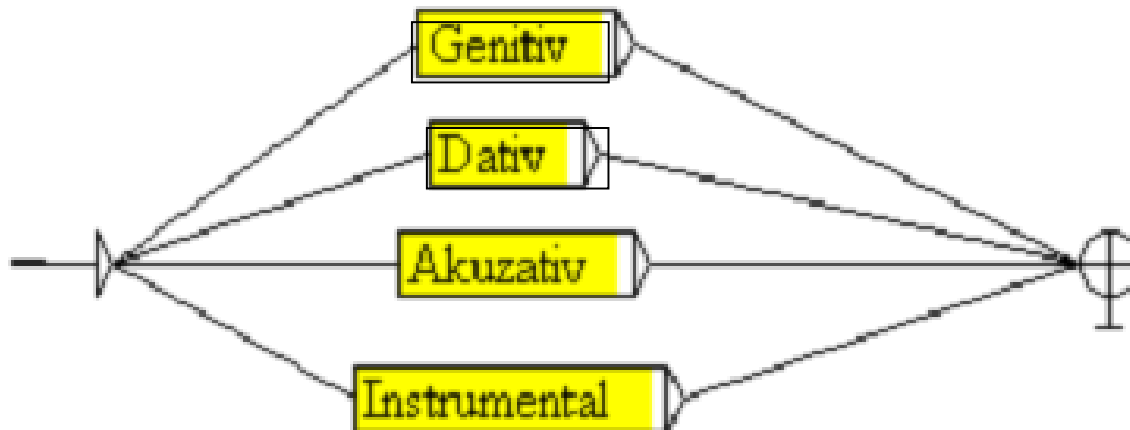
Verb + typical PC

Croatian (Croatia)/Croatian (Croatia) syntactic grammar.



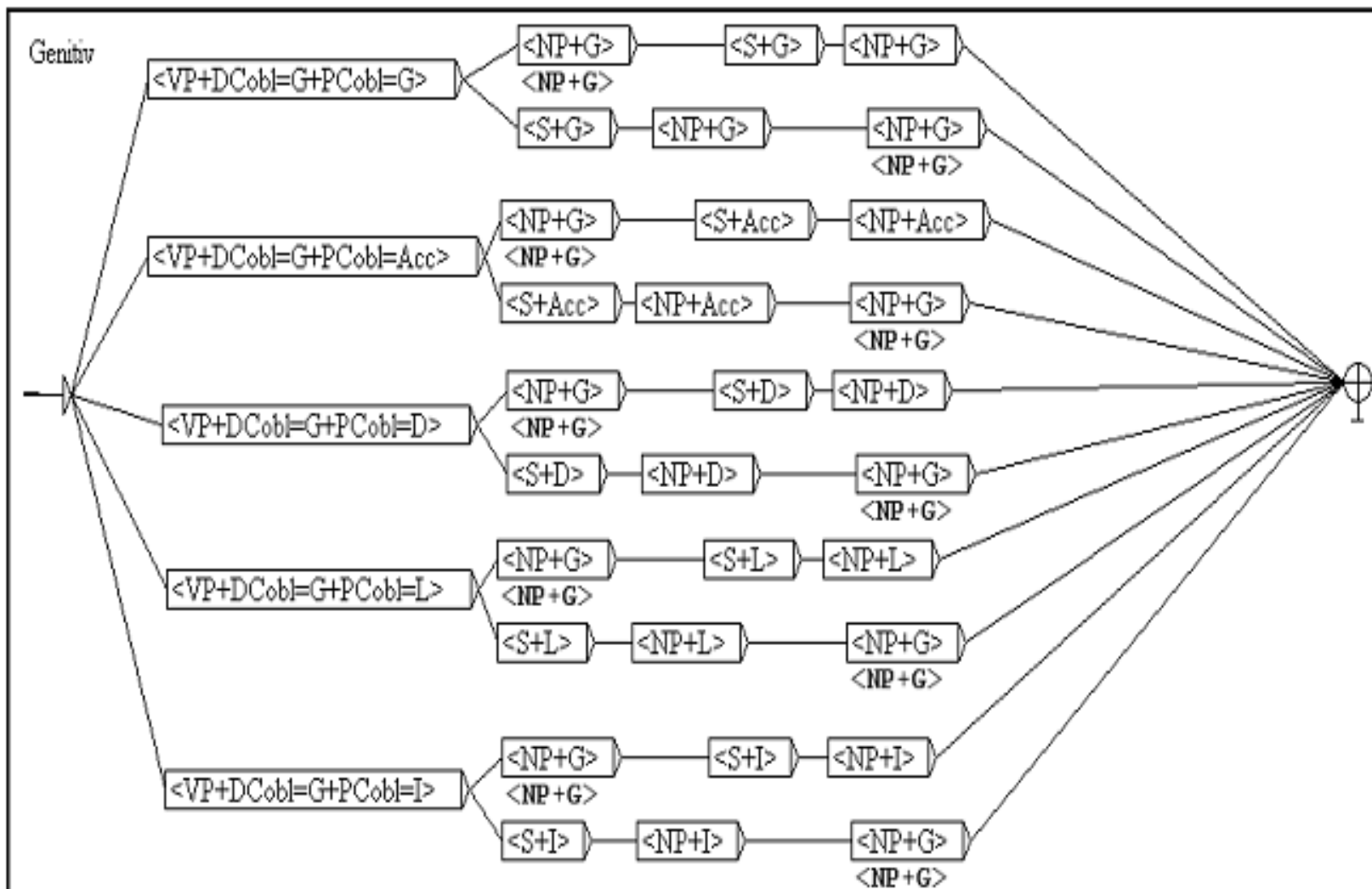
VP+DCobl=

Croatian (Croatia)/Croatian (Croatia) syntactic grammar consists of 5 graphs.



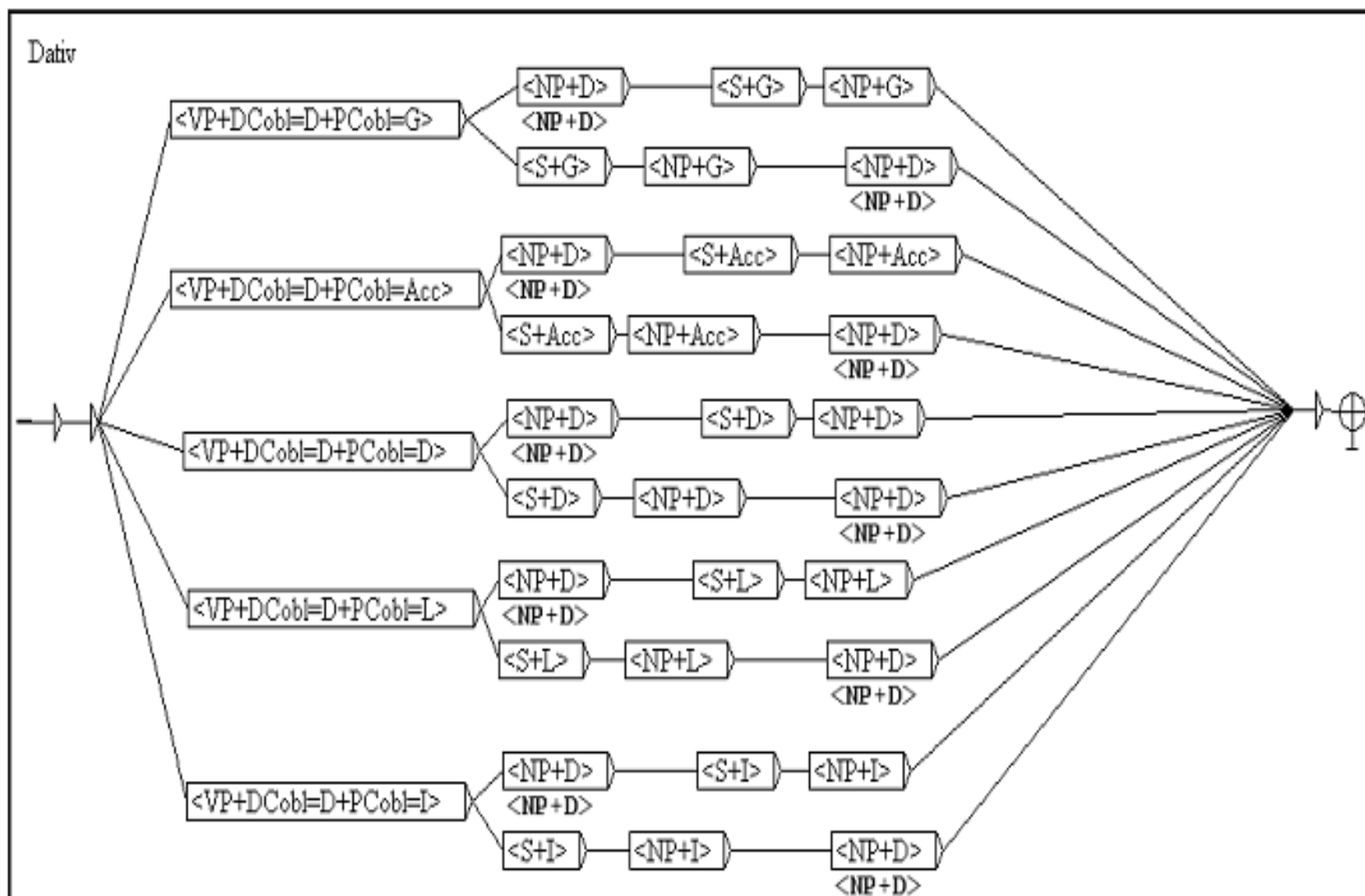
VP+DCobl=G+PCobl=G

Croatian (Croatia)/Croatian (Croatia) syntactic grammar consists of 5 graphs.



VP+DCobl=Dativ

Croatian (Croatia)/Croatian (Croatia) syntactic grammar consists of 5 graphs.





Results

| | By hand | Before CROVALLEX | After CROVALLEX |
|-----------------------|---------|------------------|-----------------|
| # of NP | 1150 | 1099 | 1070 |
| # of T unambiguous NP | | 601 | 729 |
| # of ambiguous NP | | 437 | 246+49 |
| # of F unambiguous NP | | | 26+20 |



P-R-F for unambiguous NPs

| | Before CROVALLEX | After CROVALLEX |
|-----------|---------------------|--------------------|
| Precision | 33,31 | 68,13 |
| Recall | 52,26 | 63,39 |
| F-measure | 40,69 | 65,68 |



Future work

- Subordinating conjunction.
- Infinitive construction can appear
 - with a preposition (e.g. 'nego+inf')
 - with the morphological case (e.g. 'inf+4').
- Construction with adjectives.
 - e.g. adj-7 ('Osjećam se osvježanim' - 'I feel fresh').
- Construction with adverbs.
 - e.g. adv-hrabro ('Osjećam se hrabro' - 'I feel brave').
- Construction with nominative predicate.
 - e.g. nom_pred ('Historija je postala legendom' - 'History has become a legend').