MUNI
ARTS

Automatic detection of zeugma

Mgr. Helena Medková

Zeugma detection

- continuation of the diploma thesis
- new online proofreader for Czech language
  - in development at Masaryk University within TAČR grant project in collaboration with Charles University, Czech Language Institute, Czech Academy of Sciences, and Seznam, a.s.

MUNI ARTS

Zeugma phenomenon I

- stylistic figure occurring across many languages
- forcing together two or more expressions with different meanings/grammatic structures
- typical example in Czech
  - ***Žiju*** *a* ***pocházím z Prahy****. (I* ***live*** *and* ***come from Prague****)*
  - coordinated verbs with syntactical (semantical) relation to common argument
  - their combination gives a defective sentence

MUNI ARTS

Zeugma phenomenon II.

- ***Navazuje*** *a* ***rozšiřuje publikaci…*** (***Follows up*** *and* ***extends publications****…*)

- *Mohou* ***navazovat*** *a* ***prohlubovat své kontakty*** *mezi sebou…*

  (*They can* ***establish*** *and* ***deepen their contacts*** *with each other…*)

MUNI ARTS

Zeugma phenomenon III.

– *byť Chrome umí **číst a pracovat s lokálními médii**… (although Chrome can read and work with local media…)*

– *Žáci v předchozí hodině **četli a pracovali s textem** na téma životní prostředí – třídění odpadů. (Pupils in the last lesson read and worked with the text about the environment - waste sorting.)*

MUNI ARTS

Rule-based detection of zeugma

- morphologic analysis: Majka, Desamb
- syntactic analysis: SET parser
- grammar with specific rules
- checking if there is a proper object in a proper form for a particular verb in the context of the sentence

MUNI ARTS

Automatically generated grammar

- the structure based on the manual grammar
  - created within the diploma thesis
  - involves 83 verbs
- containing together 5300 verbs
  - using lexical database VerbaLex
  - processing the right parts of the verbal patterns
  - +AG(kdo1;<person:1>;obl)+++VERB+++INFO(co4;<fact:1>;obl)
  - obligatory object on the first position
- generic, without special conditions

MUNI ARTS

Rule example

```
TMPL: bound $context* $verb (word a) (tag k5.*) $prep $noun $context*
rbound AGREE 2 4 mgn MARK 2 4 5 6 <zeugma>

$verb(lemma): rdousit pohněvat setřepat ověřit ověřovat
$verb(tag not): k5.*mN.*
$prep(tag): k7.*
$context*(tag not): k3.*yF.* .*c4.*
$noun(tag not): k3.*yF.* .*c4.*
$noun(tag): k[123].*
```


MUNI ARTS

Dataset

- sentences from czTenTen17 corpus
- annotated,1013 positive and 1681 negative cases (2694)
- 84 various verbs
- possible to verify quality of rules

MUNI ARTS

Results of the grammars comparison

	Precision	Recall	F - score
Manual grammar	0.982	0.381	0.549
Generated grammar	0.796	0.224	0. 350

testing manual grammar on the part of czTenTen17:

- precision score 0,633
- without mistakes in morphological tags 0,869

MUNI ARTS

Detection issues

- morphological analysis and desambiguation (nominativ, accusativ, genitiv):
  - *Řekl bych , že **věc chápe a rozumí** grafům.*
  - *(I suppose, he understand that thing and gets the graphs)*

rules deficiency:

- low recall
- lack of context reflection within compound sentence (marks bound, rbound)
- ellipsis, zero objects

MUNI ARTS

Possible improvements in future

- extending rules patterns
- detection of common argument
  - Po bulharské okupaci Makedonie dostal pozvání do probulharské vlády, ale odmítl a zapojil se do komunistického odboje.
  - After the Bulgarian occupation he got an invitation to probulharic government, but he refused and joined a communistic resistance.
- collocations
- machine learning

MUNI ARTS