

Návrh řešení – technické aspekty

Hana Skoumalová Alexandr Rosen

Ústav teoretické a počítačové lingvistiky
Filozofická fakulta Univerzity Karlovy v Praze

Žákovské korpusy a jejich využití ve výuce C2J
Technická univerzita v Liberci
9. září 2009

- 1 Cíle
- 2 Postup
- 3 Inspirace: projekt Falko
- 4 Prostředí pro opravy
- 5 Otázky

1 Cíle

2 Postup

3 Inspirace: projekt Falko

4 Prostředí pro opravy

5 Otázky

Cíle

Nástroje

- pro formální značkování textů
- pro opravy a chybové značkování textů
- pro gramatické značkování textů
- pro hledání v textech

Formální značkování

- identifikace textu a autora
- další údaje o textu a autorovi
- údaje o přepisu a korektuře
- identifikace odstavců a vět – striktně podle výchozího textu?

Opravy

- více možností – cílových hypotéz
- více úrovní oprav:
 - ▶ chyby týkající se jednoho slova: pravopisné, hláskoslovné, morfologické, lexikální
 - ▶ chyby týkající se více slov: syntaktické, stylistické

Gramatické značkování

- morfosyntaktická značka a lemma pro původní formy i opravy
- ? syntaktická funkce, syntaktická struktura, topologické pole

1 Cíle

2 Postup

3 Inspirace: projekt Falko

4 Prostředí pro opravy

5 Otázky

Postup

Zpracování textu

- 1 přepis textu
- 2 formální značkování
- 3 načtení do editoru
- 4 opravy jednotlivých slov s identifikací chyb
- 5 opravy posloupností slov s identifikací chyb
- 6 gramatické značkování
- 7 ? kontrola gramatického značkování
- 8 načtení do korpusového manažeru

1 Cíle

2 Postup

3 Inspirace: projekt Falko

4 Prostředí pro opravy

5 Otázky

Inspirace: projekt Falko

viz např.:

- [Anke Lüdeling & Adolphs(2005)]
- [Hagen Hirschmann & Lüdeling(2007)]
- software: EXMaRALDA

multi-layer standoff annotation:

- anotace na více rovinách
- nové roviny anotace se dají přidávat a zpracovávat nezávisle na existujících

automaticky anotované roviny:

- slovní druh (morfologická značka)
- lemma

roviny pro anotaci chyb:

- cílová hypotéza (jak to má být správně, zatím v *Zusammenfassungskorpus* a *Longitudinalkorpus*)
- anotace chyb na základě cílové hypotézy:
 - ▶ opravený slovní druh
 - ▶ topologická pole (v ZFK a LK)
 - ▶ chyby ve slovosledu (v ZFK a LK)

- 1 Cíle
- 2 Postup
- 3 Inspirace: projekt Falko
- 4 Prostředí pro opravy**
- 5 Otázky

Co je možné v Partitur Editoru EXMaRALDA

- import textu a anotace
- kopírování anotace na další úroveň
- spojování a rozdělování pozic ve větě
- určení typu chyby
- export textu s chybovou anotací a import gramaticky označovaného textu

- 1 Cíle
- 2 Postup
- 3 Inspirace: projekt Falko
- 4 Prostředí pro opravy
- 5 Otázky**

Otázky

- 1 V jediném slově může být chyba pravopisná, hláskoslovná, morfologická, lexikální, navíc slovo může být na nesprávném místě ve větě.
Měli bychom uvádět typ chyby a správný tvar (značku, lemma) „inkrementálně“, pro každou úroveň chyby?
např. opravit pravopis a koncovku lexikálně nevhodného slova?
(*okruhle očky – okruhle oči – okrouhlé oči – kulaté oči*)
- 2 Pro každou cílovou hypotézu uvádět všechny typy chyb?
- 3 Typy chyb zkratkou v buňce nebo pro každou chybu novou řádku?
- 4 Využít chybové anotace při gramatickém značkování?



Anke Lüdeling, Maik Walter, E. K. & Adolphs, P. (2005).
Multi-level error annotation in learner corpora.
In Proceedings of Corpus Linguistics 2005, Birmingham.



Hagen Hirschmann, S. D. & Lüdeling, A. (2007).
Syntactic annotation of non-canonical linguistics structures.
In Proceedings of Corpus Linguistics 2007, Birmingham.