

VONZATKERETEK A MAGYAR NEMZETI SZÖVEGTÁRBAN

Sass Bálint

joker@nytud.hu

Magyar Tudományos Akadémia
Nyelvtudományi Intézet
Korpusznyelvészeti Osztály

MSZNY2005

Szeged, 2005. december 8-9.

1 MOTIVÁCIÓ

2 ESZKÖZÖK ÉS ELJÁRÁSOK

- Vonzatkeretek
- Korpuszfeldolgozó eszköz
- A vonzatkeret-felismerés menete
- Jelenlegi állapot

3 EREDMÉNYEK

- Példák
- Esettanulmányok
- Mutatvány a vonzatkeret-gyakorisági szótárból

4 TÁVLAT

- Alkalmazás
- Fejlesztési lehetőségek

1 MOTIVÁCIÓ

2 ESZKÖZÖK ÉS ELJÁRÁSOK

- Vonzatkeretek
- Korpuszfeldolgozó eszköz
- A vonzatkeret-felismerés menete
- Jelenlegi állapot

3 EREDMÉNYEK

- Példák
- Esettanulmányok
- Mutatvány a vonzatkeret-gyakorisági szótárból

4 TÁVLAT

- Alkalmazás
- Fejlesztési lehetőségek

MOTIVÁCIÓ

CÉL

- vonzatkeretek felismerése
- vonzatkeret-gyakorisági szótár elkészítése

MIÉRT?

- szótárak, ontológiák „jelentéskincse” – empirikus alapokon
- a vonzatkeret-táblázat fejlesztése
- gépi szövegértés

1 MOTIVÁCIÓ

2 ESZKÖZÖK ÉS ELJÁRÁSOK

- Vonzatkeretek
- Korpuszfeldolgozó eszköz
- A vonzatkeret-felismerés menete
- Jelenlegi állapot

3 EREDMÉNYEK

- Példák
- Esettanulmányok
- Mutatvány a vonzatkeret-gyakorisági szótárból

4 TÁVLAT

- Alkalmazás
- Fejlesztési lehetőségek

VONZATKERETEK

- *forrás*: a Nyelvtudományi Intézetben készült lexikai adatbázis igei vonzatkeret-táblázata
- 9000 ige 18000 vonzatkerete
- teljesen kompozicionális szerkezetek nélkül

KORPUSZFELDOLGOZÓ ESZKÖZ

- a projekt keretében készülő általános célú eszköz
- elemzett, egyértelműsített korpuszból indul ki
- részletes morfológiai reprezentáció
- többszintű reguláris nyelvtan
- hasznos kiegészítő funkciók
 - pozíció megadása
 - tagadás
 - szerkezetek öröklik a fej tulajdonságait
 - többszintű tagek
 - annotáció törlése

A VONZATKERET-FELISMERÉS MENETE

- 1 Tagmondatokra bontás
- 2 Tulajdonnév felismerés
- 3 NP-felismerés
- 4 Vonzatkeret-illesztés

JELENLEGI ÁLLAPOT

TELJES FELISMERÉSI FOLYAMAT

Elemzett szöveg → nyers vonzatkeret-gyakorisági szótár

Egyszerűsítések:

- tesztkorpusz: írásjel nélküli, 9 szavas mondatok
- primitív tulajdonnév- és NP-felismerő
- egyszerűbb keretek:
adott igeő + névszói alany, tárgy, vonzatok;
megadható: szófaj, eset, konkrét szóalak, szótő

1 MOTIVÁCIÓ

2 ESZKÖZÖK ÉS ELJÁRÁSOK

- Vonatkeretek
- Korpuszfeldolgozó eszköz
- A vonatkeret-felismerés menete
- Jelenlegi állapot

3 EREDMÉNYEK

- Példák
- Esettanulmányok
- Mutatvány a vonatkeret-gyakorisági szótárból

4 TÁVLAT

- Alkalmazás
- Fejlesztési lehetőségek

PÉLDÁK

egybevet vmit vmivel:

„Az önellenőrzés során a dolgozó egybeveti a munkáját a követelményekkel.”

utasít vkit vmire:

„A Közgyűlés utasítja a Polgármestert a szükséges intézkedések megtételére.”

ESETTANULMÁNY: *vág*

- alulspecifikált:
vág vmit → *vág vmit vmire* (szeletekre, darabokra)
vág vmit vmibe → *nagy fába vágta a fejszáját*
elvág vmit → *torkát* ill. *vmit vmitől*
- elhagyandó: *vág vmiben*, *vág vkit vmin*
- nincs: *levág vmit vmiből*, *kettévág vmit*
- kiegészítendő:
pofát vág + *grimaszt* ill. *fát*
pofon vág + *kupán*

ESETTANULMÁNY: *vesz*

vonzatkeret	db	ÉKSz. jelentés
részt vesz vmi(be)n	227	5.
tudomásul vesz	23	16.
fordulatot vesz	8	18.
semmibe vesz	2	<i>nincs benne!</i>
feleségül vesz	1	11.

MUTATVÁNY

A VONZATKERET-GYAKORISÁGI SZÓTÁRBÓL

#	vonzatkeret	db
1.	részt vesz vmiben	124
2.	részt vesz vmin	103
3.	kérdést tesz fel	27
4.	tudomásul vesz	23
5.	győzelmet arat	16
6.	szert tesz vmire	16
7.	figyelmet fordít vmire	13
8.	hatást gyakorol vmire	12
9.	világra jön	9
10.	letartóztatásba helyez vkit	9

1 MOTIVÁCIÓ

2 ESZKÖZÖK ÉS ELJÁRÁSOK

- Vonzatkeretek
- Korpuszfeldolgozó eszköz
- A vonzatkeret-felismerés menete
- Jelenlegi állapot

3 EREDMÉNYEK

- Példák
- Esettanulmányok
- Mutatvány a vonzatkeret-gyakorisági szótárból

4 TÁVLAT

- Alkalmazás
- Fejlesztési lehetőségek

ALKALMAZÁS

- szótárak, ontológiák „jelentés-kincse” – empirikus alapokon
→ Magyar EuroWordNet Projekt
- a vonzatkeret-táblázat továbbfejlesztése
ritka kereteket (1) elhagyni vagy (2) jobban specifikálni
gyakori hiányzó kereteket hozzáadni
- gépi szövegértés
jelentéshozzárendelés
formailag azonos, eltérő jelentésű kereteket külön kódolni
- szintaktikai elemzők támogatása

FEJLESZTÉSI LEHETŐSÉGEK

- jobb/részletesebb nyelvtanok/felismerők
- tagmondatra bontó modul
- szemantikai jegyek feldolgozása
- szabad határozók kiszűrése

FEJLESZTÉSI LEHETŐSÉGEK

- jobb/részletesebb nyelvtanok/felismerők
- tagmondatra bontó modul
- szemantikai jegyek feldolgozása
- szabad határozók kiszűrése

Köszönöm a figyelmet!