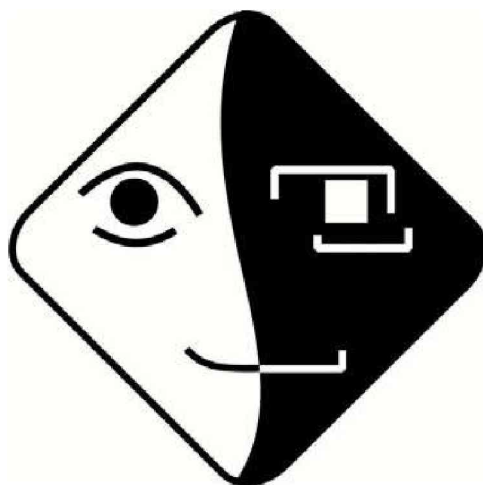


# **VIII. Magyar Számítógépes Nyelvészeti Konferencia**



## **MSZNY 2011**

Szerkesztette:

Tanács Attila  
Vincze Veronika

Szeged, 2011. december 1-2.

<http://www.inf.u-szeged.hu/mszny2011>

ISBN: 978-963-306-121-3

Szerkesztette: Tanács Attila és Vincze Veronika  
{tanacs, vinczev}@inf.u-szeged.hu

Felelős kiadó: Szegedi Tudományegyetem, Informatikai Tanszékcsoport  
6720 Szeged, Árpád tér 2.

Nyomtatta: JATEPress  
6722 Szeged, Petőfi Sándor sugárút 30–34.

Szeged, 2011. november

## Előszó

2011. december 1-2-án nyolcadik alkalommal rendezzük meg Szegeden a Magyar Számítógépes Nyelvészeti Konferenciát. Nagy örömet jelent számomra, hogy a rendezvény fokozott érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében. A konferencia fő célja – a hagyományokhoz híven – a nyelv- és beszédtechnológia területén végzett legújabb, illetve folyamatban levő kutatások eredményeinek ismertetése és megvitatása, mindemellett lehetőség nyílik különféle hallgatói projektek, illetve ipari alkalmazások bemutatására is.

A konferenciafelhívásra szép számban beérkezett tudományos előadások közül a programbizottság 40-et fogadott el az idei évben, így 28 előadás és 12 poszter-, illetve laptopos bemutató gazdagítja a konferencia programját. A programban a magyar számítógépes nyelvészet teljes palettájáról találhatunk előadásokat a beszédtechnológiától kezdve a számítógépes szemantika és pragmatika területén át az információkinyerésig és gépi fordításig.

A korábbi évekhez hasonlóan idén is tervezzük a „Legjobb Ifjú Kutatói Díj” odaítélését, mellyel a fiatal korosztály tagjait kívánjuk ösztönözni arra, hogy kiemelkedő eredményekkel járuljanak hozzá a magyarországi nyelv- és beszédtechnológiai kutatásokhoz. A díj felajánlásáért az MTA Számítástechikai és Automatizálási Kutatóintézetének tartozunk köszönettel.

Szeretnék köszönetet mondani a programbizottságnak: Vámos Tibor programbizottsági elnöknek, valamint Alberti Gábor, Gordos Géza, László János, Prószekey Gábor és Váradi Tamás programbizottsági tagoknak. Szeretném továbbá megköszönni a rendezőbizottság (Alexin Zoltán, Almási Attila, Vincze Veronika) és a kötet szerkesztők (Tanács Attila, Vincze Veronika) munkáját is.

Csirik János, a rendezőbizottság elnöke

Szeged, 2011. november



# Tartalomjegyzék

## I. Többynnyelvűség

Többynnyelvű dokumentum nyelvének megállapítása .....	3
<i>Pataki Máté, Vajna Miklós</i>	
Statisztikai gépi fordítási módszereken alapuló egynyelvű szövegelemző rendszer és szótövesítő .....	12
<i>Laki László János</i>	
Fordítási plágiumok keresése .....	24
<i>Pataki Máté</i>	
Soknyelvű páros gépi fordítás hatékony és megbízható kiértékelése .....	35
<i>Oravecz Csaba, Sass Bálint, Tihanyi László</i>	
Igei bővítménykeretek fordítási ekvivalenseinek kinyerése mélyen elemzett párhuzamos korpuszból .....	47
<i>Héja Enikő, Takács Dávid, Sass Bálint</i>	
Félig kompozicionális szerkezetek automatikus azonosítása magyar és angol nyelven .....	59
<i>Vincze Veronika, Nagy T. István, Zsibrita János</i>	

## II. Korpusz, ontológia

Jelentés-egyértelműsített szabadalmi korpusz .....	73
<i>Nagy Ágoston, Almási Attila, Vincze Veronika</i>	
Korpuszépítés ómagyar kódexekből .....	81
<i>Simon Eszter, Sass Bálint, Mittelholcz Iván</i>	
Nem lexikalizált fogalmak a Magyar WordNetben .....	90
<i>Vincze Veronika, Almási Attila</i>	
A Magyar szóelemtár megalkotása és a Magyar gyökszótár előkészítő munkálatai .....	102
<i>Kiss Gábor, Kiss Márton, Sáfrány-Kovalik Balázs, Tóth Dorottya</i>	

## III. Szintaxis, morfológia, névelem-felismerés

A sekély mondattani elemzés további lépései .....	113
<i>Recski Gábor</i>	

Közösségkeresés alapú felügyelet nélküli szófaji egyértelműsítés.....	119
<i>Berend Gábor, Vincze Veronika</i>	
Szófaji kódok és névelemek együttes osztályozása .....	131
<i>Móra György, Vincze Veronika, Zsibrita János</i>	
Magyar nyelvű klinikai dokumentumok előfeldolgozása .....	143
<i>Siklósi Borbála, Orosz György, Novák Attila</i>	

#### **IV. Beszédtechnológia**

Nyelvimodell-adaptáció ügyfélszolgálati beszélgetések gépi leiratozásához.....	155
<i>Tarján Balázs, Mihajlik Péter, Fegyő Tibor</i>	
Prozódiai változatosság rejtett Markov-modell alapú szövegfelolvasóval .....	167
<i>Csapó Tamás Gábor, Németh Géza</i>	
A szintaktikai szerkezet automatikus feltérképezése a beszédjel prozódiai elemzése alapján .....	178
<i>Szaszák György, Beke András</i>	
A HuComTech-korpusz és -adatbázis számítógépes feldolgozási lehetőségei. Automatikus prozódiai annotáció.....	190
<i>Szekrényes István, Csipkés László, Oravecz Csaba</i>	
A HuComTech audio adatbázis szintaktikai szintjének elvei és szabályrendszerének újdonságai .....	199
<i>Kiss Hermina</i>	

#### **V. Pszichológia, pragmatika, kognitív nyelvészet**

A csoportközi értékelés mint a csoporttrauma érzelmi feldolgozásának indikátora a nemzeti történelem elbeszéléseiben .....	211
<i>Csertő István, László János</i>	
Szemantikus szerepek vizsgálata magyar nyelvű szövegek narratív pszichológiai elemzésében.....	223
<i>Ehmann Bea, Lendvai Piroska, Fritz Adorján, Miháltz Márton, Tihanyi László</i>	
Paralingvisztikai jegyek a narratív pszichológiai tartalomelemzésben: a magabiztosság-krízis skála .....	231
<i>Puskás László</i>	
A multimodális pragmatikai annotáció jelentősége a számítógépes nyelvészetben .....	240
<i>Bódog Alexa, Abuczki Ágnes, Németh T. Enikő</i>	

Metaforikus kifejezések szerkezeti jellemzői .....	252
<i>Babarczy Anna</i>	

## VI. Szemantika

Az intenzionalitás számítógépes nyelvészeti kezelése – avagy a $\Re$ ALIS $\lambda$ szintfüggvénye .....	263
<i>Alberti Gábor</i>	
Tárgymodellváltozatok a $\Re$ ALIS nyelvi elemzéshez.....	276
<i>Kilián Imre</i>	
Interpretáció, intenzionalitás, modalitás – avagy a $\Re$ ALIS $\lambda$ függvényének implementációja felé .....	284
<i>Károly Márton</i>	
Kvantifikált kifejezések hatóköri többértelműségének szabályalapú kezelése .....	297
<i>Szécsényi Tibor</i>	

## VII. Poszterek és laptopos bemutatók

Interaktív formánsérték-módosító fejlesztése .....	309
<i>Abari Kálmán, Olasz Gábor</i>	
Korpuszalapú entrópiamértékek gating- és lexikai döntési kísérletekben .....	316
<i>Fazekas Judit, Németh Kornél, Pléh Csaba, Varga Dániel</i>	
Automatikusan előállított protoszótárak közzététele .....	319
<i>Héja Enikő, Takács Dávid</i>	
MASZEKER: szemantikus keresőprogram .....	321
<i>Hussami Péter</i>	
Interaktív fonetikai eszköz az artikulációs csatorna keresztmetszet-függvényének meghatározására.....	323
<i>Jani Máttyás, Björn Lindblom, Sten Ternström</i>	
Szabadalmak igénypontgráfiájának automatikus előállítása és hibaelemzése .....	329
<i>Kiss Márton, Vincze Veronika, Nagy Ágoston, Alexin Zoltán</i>	
Magyar NP-felismerők összehasonlítása .....	333
<i>Miháltz Márton</i>	
Javában taggelünk .....	336
<i>Novák Attila, Orosz György, Indig Balázs</i>	
A HunOr magyar-orosz párhuzamos korpusz .....	341
<i>Szabó Martina Katalin, Schmalcz András, Nagy T. István, Vincze Veronika</i>	

Magyar szóalak- és morfológiaelemzés-adatbázis .....	348
<i>Szidarovszky Ferenc P., Tóth Gábor, Tikk Domonkos</i>	
Lemmaasszociáció és morfológiai jegyek mesterséges neurális hálózatokban.....	354
<i>Tóth Ágoston, Csernyi Gábor</i>	
Fonológiai jegyek felügyelet nélküli tanulása fonemikus korpuszból .....	359
<i>Vásárhelyi Dániel</i>	
<b>Szerzői index, névmutató.....</b>	<b>362</b>