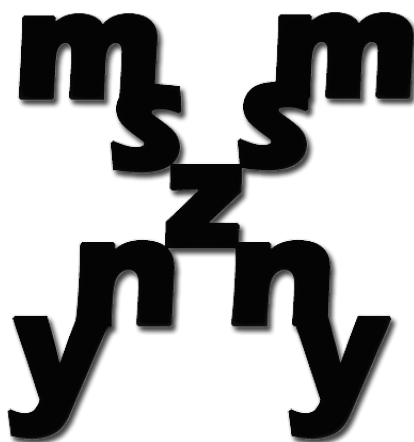


XX. Magyar Számítógépes  
Nyelvészeti Konferencia



Szerkesztette:  
Berend Gábor  
Gosztolya Gábor  
Vincze Veronika

Szeged, 2024. január 25–26.

**Szerkesztette:**

Berend Gábor, Gosztolya Gábor, Vincze Veronika  
{berendg,gabor,vinczev}@inf.u-szeged.hu

**Felelős kiadó:**

Szegedi Tudományegyetem  
TTIK, Informatikai Intézet  
6720 Szeged, Árpád tér 2.

**ISBN:** 978-963-306-973-8

Szeged, 2024. január

**Az MSZNY 2024 konferencia szervezője:**

HUN-REN–SZTE Mesterséges Intelligencia Kutatócsoport

## Előszó

2024. január 25–26-án immáron huszadik alkalommal kerül sor a Magyar Számítógépes Nyelvészeti Konferencia megrendezésére. A konferencia fő célkitűzése a kezdetek óta állandó: lehetőséget biztosítani a nyelv- és beszédtechnológia területén végzett kutatások eredményeinek ismertetésére és megvitatására, ezen felül pedig a különféle hallgatói projektek, illetve ipari alkalmazások bemutatására.

Az idei évben a 24 beküldött cikkből gondos mérlegelést követően 19 cikk került elfogadásra, melyek témája a nyelv- és beszédtechnológia számos szakterületét lefedi a legújabb nyelvi modellek bemutatásától kezdve a beszédtechnológia eredményein keresztül egészen a gépi fordításig.

Nagy örömet jelent számunkra, hogy Sebők Miklós elfogadta meghívásunkat, aki plenáris előadását *Nagy nyelvi modellek az összehasonlító politikatudományban: Közpolitikai témák klasszifikációja a Babel-rendszerrel* címmel fogja megtartani.

Az idei évben is különdíjjal jutalmazzuk a konferencia legjobb cikkét, mely a legjelentősebb eredményekkel járul hozzá a magyarországi nyelv- és beszédtechnológiai kutatásokhoz. Ezen felül immár hatodik alkalommal osztjuk ki a legjobb bíráló díját, amellyel a bírálók fáradtságos, ugyanakkor nélkülözhetetlen munkáját kívánjuk elismerni.

A szervezőbizottság nevében,

Ács Judit,

Berend Gábor,

Gosztolya Gábor,

Ligeti-Nagy Noémi,

Nemeskey Dávid Márk,

Novák Attila,

Simon Eszter,

Sztahó Dávid,

Vincze Veronika



# Tartalomjegyzék

<b>Előfeldolgozás, szintaxis</b>	<b>1</b>
3	Bírósági határozatok automatikus mondatszegmentálásának hatékonyságmérése <i>Csányi Gergely Márk, Lakatos Dorina, Üveges István, Vági Renátó, Megyeri Andrea, Fülöp Anna, Nagy Dániel, Vadász János Pál</i>
17	OCR-hibák kvantitatív elemzése több szövegváltozat összehasonlításával <i>Pethő Gergely, Sass Bálint, Simon László, Lipp Veronika</i>
31	Mi a manó! Egy sajátos szerkezet korpuszvezérelt vizsgálata <i>Vincze Veronika</i>
43	“A fatens felelt pedig. . .” – A Történeti Magánéleti Korpusz igei szerkezeteinek mozaik n-gram alapú feldolgozása <i>Bajzát Tímea Borbála, Indig Balázs, Kalivoda Ágnes</i>
<b>Nyelvmodellek, párbeszéd, gépi fordítás</b>	<b>59</b>
61	ParancsPULI: Az utasításkövető PULI-modell <i>Yang Zijian Győző, Dodé Réka, Héja Enikő, Laki László János, Ligeti-Nagy Noémi, Madarász Gábor, Váradi Tamás</i>
73	SHunQA: egy nyíltkérdés-megválaszoló rendszer <i>Berkecz Péter, Zombori Tamás, Banga Gergő, Szabó Gergő, Szántó Zsolt, Novák Attila, Farkas Richárd</i>
85	Neurális nyelvi modellek látens szemantikus információ alapján történő maszkolásmentes előtanítása <i>Berend Gábor</i>
97	Building high capacity machine translation models for knowledge distillation and production workflows <i>Csaba Oravecz, Bhavani Bhaskar, Katina Bontcheva, Bogomil Kovachev</i>
<b>Szemantika, pragmatika</b>	<b>115</b>
117	Személyes adatok azonosítása és automatikus lecserélése magyar nyelvű szövegekben <i>Novák Attila, Novák Borbála</i>
131	Kiskereskedelmi terméknevek kategorizálása Kombinált Nómenklatúra szerint <i>Ónozó Livia Réka, Putz Orsolya, Járási István, Gyires-Tóth Bálint</i>

- 145 Saving labeling cost by embracing Active Learning: a case study  
*István Üveges, Renátó Vági, Andrea Megyeri, Anna Fülöp, Dániel Nagy, János Pál Vadász, Gergely Márk Csányi*
- 159 Felszólításannotálás a MedCollect egészségügyi álhírkorpuszban  
*Szécsényi Tibor, Nagy C. Katalin, Németh T. Enikő*

## **Orvosi beszédfeldolgozás**

**171**

- 173 Comparative analysis of multiple speech tasks to recognise Parkinson's disease using pre-trained feature extractor embeddings  
*Attila Zoltán Jenei, Zalán Valárik, Dávid Sztahó*
- 187 Magyar nyelvű dizartriás beszéd automatikus elemzése - egy pilot kutatás eredményei  
*Oláh Julianna, Szabó Martina Katalin, Szőke Eszter, Plander Nóra, Hoffmann Ildikó*
- 201 Az egészség jele a szöveg EGÉSZsége? - Szövegkoherencia borderline személyiségzavarban  
*Felletár Fanni, Yang Zijian Győző, Babarczy Anna*

## **Poszter, laptopos bemutató**

**215**

- 217 Tagmondatok és megszakítatlan összetevők kinyerése függőségi elemzésből  
*Szécsényi Tibor*
- 229 A nagy nyelvi modell alapú szervezeti automatizáció lehetőségei és az autonóm ágensek kapcsolódó kihívásai  
*Vándor Péter, Csáki Csaba*
- 243 Magyar nyelvű utasításkövető korpusz építése Stanford Alpaca promptok fordításával és lokalizálásával  
*Yang Zijian Győző, Szlávik Szilárd, Ligeti-Nagy Noémi*
- 257 Kulcsszógenerálás magyar nyelvű, hosszú szövegekből nagy nyelvi modellekkel  
*Dodé Réka, Yang Zijian Győző*

## **Szerzői index, névmutató**

**269**