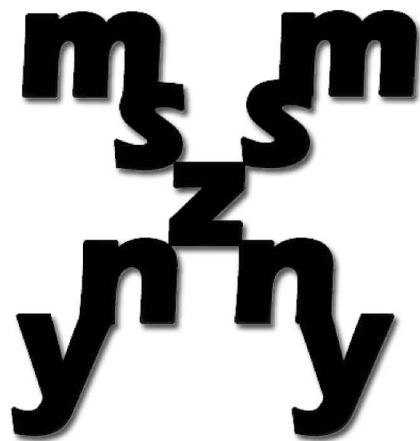


# XV. Magyar Számítógépes Nyelvészeti Konferencia



Szerkesztette:  
Berend Gábor  
Gosztolya Gábor  
Vincze Veronika

Szeged, 2019. január 24–25.

**Szerkesztette<sup>1</sup>:**

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

**Felelős kiadó:**

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

**Nyomtatta:**

JATEPress  
6722 Szeged, Petőfi Sándor sugárút 30–34.

Szeged, 2019. január

**Az MSZNY 2019 konferencia szervezője:**

MTA-SZTE Mesterséges Intelligencia Kutatócsoport

---

<sup>1</sup>a L<sup>A</sup>T<sub>E</sub>X's 'confproc' csomagjára támaszkodva

## Előszó

2019. január 24-25-én tizenötödik alkalommal rendezzük meg Szegeden a Magyar Számítógépes Nyelvészeti Konferenciát. A konferencia fő célkitűzése a kezdetek óta állandó: lehetőséget biztosítani a nyelv- és beszédtechnológia területén végzett kutató-sok eredményeinek ismertetésére és megvitatására, ezen felül a különféle hallgatói projektek, illetve ipari alkalmazások bemutatására. Nagy örömet jelent számunkra, hogy a hagyományokat követve a konferencia idén is nagyfokú érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében.

Az évek során hagyománnyá vált az is, hogy a mesterséges intelligencia vagy a számítógépes nyelvészet egy-egy kiemelkedő alakja plenáris előadást tart a konferencián. Az idei évben Turán György (MTA-SZTE Mesterséges Intelligencia Kutatócsoport és University of Illinois at Chicago) előadásában az interpretálhatóságról és annak számítógépes nyelvészeti vonatkozásairól lesz szó.

Az idei évben is szeretnénk különdíjjal jutalmazni a konferencia legjobb cikkét, mely a legkiemelkedőbb eredményekkel járul hozzá a magyarországi nyelv-és beszédtechnológiai kutatásokhoz. Továbbá idén először tervezzük bevezetni a "Legjobb Bírálók Díját" is, így elismerve a bírálók fáradtságos, ámde nélkülözhetetlen munkáját. A konferenciához idén is kapcsolódni fog egy kerekasztal-megbeszélés, ahol a főbb szakmai kérdéseket, a szakterület jelenlegi helyzetét és várható haladási irányát, valamint a konferenciához közvetlenül kapcsolódó kérdéseket vitatják meg a résztvevők.

Köszönettel tartozunk a LogMeIn-nek, a Neumann János Számítógéptudományi Társaságnak, valamint a Clementine-nak is, akik anyagi támogatásukkal járultak hozzá a konferencia sikeres lebonyolításához. Az előzőeken felül hálásak vagyunk az MTA-SZTE Mesterséges Intelligencia Kutatócsoportján és a Szegedi Tudományegyetem Informatikai Intézetének Szoftverfejlesztés Tanszékén dolgozó azon kollégáknak, akik a helyi szervezésben segédkeztek. Végezetül szeretnénk megköszönni a programbizottság és a szervezőbizottság minden tagjának áldozatos munkáját, ami nélkül nem jöhetett volna létre a konferencia.

A szervezőbizottság nevében,

Ács Judit

Berend Gábor

Novák Attila

Simon Eszter

Sztahó Dávid

Vincze Veronika



# Tartalomjegyzék

## Beszédtechnológia I.

1

- 3 Beszélőinvariáns akusztikus modellek létrehozása mély neuronhálók elleneséges multi-taszki tanításával  
*Tóth László, Gosztolya Gábor*
- 13 Autoenkóderen alapuló jellemzőreprezentáció mély neuronháló, ultrahang-alapú néma-beszéd-interfészekben  
*Pintér Ádám, Gosztolya Gábor, Tóth László, Grósz Tamás, Csapó Tamás Gábor, Markó Alexandra*
- 23 Ügyfélszolgálati beszélgetések nyelvmodellezése rekurrens neurális hálózatokkal  
*Tarján Balázs, Fegyó Tibor, Mihajlik Péter*

## Szemantika

35

- 37 CBOW/A: módosított CBOW algoritmus annotált szövegekből készített vektortérmodellek létrehozására  
*Novák Attila, Laki László János, Novák Borbála*
- 49 Interpretability of Hungarian embedding spaces using a knowledge base  
*Balogh Vanda, Berend Gábor, Dimitris Diochnos, Farkas Richárd, Turán György*
- 63 Mit hozott édesapám? Döntést – Idiomatikus és félig kompozicionális magyar igei szerkezetek azonosítása párhuzamos korpuszból  
*Novák Attila, Laki László János, Novák Borbála*
- 73 Neurális hálózat-alapú gépi fordítórendszer minőségének javítása domain adaptáció segítségével  
*Laki László János*
- 83 Egy magyar nyelvű kérdezőrendszer  
*Novák Attila, Laki László János, Novák Borbála, Dömötör Andrea, Ligeti-Nagy Noémi, Kalivoda Ágnes*

## Poszter, demó

97

- 99 Konverterek magyar morfológiai címkékészletek között  
*Vadász Noémi, Simon Eszter*
- 113 Named Entity Recognition in the Miskolc Legal Corpus  
*Üveges István*
- 123 End-to-end Convolutional neural networks for Intent Detection  
*Sevinj Yolchuyeva, Németh Géza, Gyires-Tóth Bálint*

- 135 An annotation tool for academic literature processing  
*Molnár Zsolt, Polgár Tímea, Vincze Veronika*
- 145 Formális fogalmak a jogi ontológiákban  
*Syi, Hamp Gábor, Markovich Réka, Grad-Gyenge Anikó, Héder Ákos, Nagy Krisztina, Vértesy László*
- 153 Kísérletek tudásbázis- és mondatkörnyezet-alapú beágyazásokkal magyar nyelvre  
*Kardos Péter, Berend Gábor, Farkas Richárd*
- 163 Szemantikai keretek felismerése neurális hálózatok és szódisztribúciós adatok felhasználásával  
*Tóth Ágoston*

### **Orvosi alkalmazások**

**175**

- 177 Információkinyerés magyar nyelv gerinc MR leletekből  
*Kicsi András, Pusztai Péter, Szabó Ledényi Klaudia, Szabó Endre, Berend Gábor, Vincze Veronika, Vidács László*
- 189 Szkizofrénia azonosítása spontán beszéd temporális paramétereire alapján – egy pilot kutatás eredményei  
*Bagi Anita, Gosztolya Gábor, Szalóki Szilvia, Szendi István, Hoffmann Ildikó*
- 203 Betegségek automatikus szétválasztása időben eltolt akusztikai jellemzők korrelációs struktúrája alapján  
*Sztahó Dávid, Kiss Gábor, Tulics Miklós, Vicsi Klára*

### **Morfológia, nyelvi elemzés**

**213**

- 215 PoS-tagging and lemmatization with a deep recurrent neural network  
*Ugray Gábor*
- 225 Hol ugat a kutya? Örömeiben. Helyhatározói esetragos névszók pontosabb annotációja  
*Ligeti-Nagy Noémi, Novák Attila*
- 235 emt sv – Egy formátum mind felett  
*Indig Balázs, Sass Bálint, Simon Eszter, Kundráth Péter, Vadász Noémi, Mittelholcz Iván*
- 249 The impact of inflection on word vectors  
*Lévai Dániel, Kornai András*

### **Beszédtechnológia II.**

**263**