

# DR. VINCZE VERONIKA

## ELÉRHETŐSÉG

MTA-SZTE Mesterséges Intelligencia Kutatócsoport

6720 Szeged, Tisza Lajos krt. 103.

Telefon: +36-62-54-41-40

Mobil: +36-70-22-99-108

Fax: +36-62-54-67-37

E-mail: vinczev@inf.u-szeged.hu

Honlap: <http://www.inf.u-szeged.hu/~vinczev>

## SZEMÉLYES ADATOK

Születési idő: 1981. július 1.

Állampolgárság: magyar

## ÉRDEKLŐDÉSI KÖR

számítógépes nyelvészet (elsősorban korpuszépítés, információkinyerés, számítógépes morfológia és szintaxis), lexikális szemantika, függőségi nyelvtanok

## VÉGZETTSÉG

- 2015 informatika PhD, Szegedi Tudományegyetem  
Doktori értekezés: *Uncertainty Detection in Natural Language Texts [Bizonytalanság azonosítása természetes nyelvű szövegekben]*
- 2012 nyelvtudomány PhD, Szegedi Tudományegyetem  
Doktori értekezés: *Semi-Compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistic Analyses [Félig kompozicionális főnév + ige szerkezetek: Elméleti kérdések és számítógépes nyelvészeti elemzések]*
- 2004 általános nyelvészet MA, Szegedi Tudományegyetem  
Szakdolgozat címe: *Funkcióigés szerkezetek elemzése lexikai függvények segítségével*
- 2004 angol nyelv és irodalom MA, Szegedi Tudományegyetem  
Szakdolgozat címe: *Complimenting Habits of Hungarian Students [Magyar diákok bókolási szokásai]*

## TANULMÁNYOK

- 2004-2007 elméleti nyelvészet PhD-program, Szegedi Tudományegyetem
- 1999-2004 általános nyelvészet BA&MA-program, Szegedi Tudományegyetem
- 1999-2004 angol nyelv és irodalom BA&MA-program, Szegedi Tudományegyetem
- 2001-2004 portugál nyelv és irodalom minor, Szegedi Tudományegyetem

## SZAKMAI TAPASZTALAT

- 2018- tudományos főmunkatárs – MTA-SZTE Mesterséges Intelligencia Kutatócsoport
- 2017-2018 QA Manager in Data Science – Black Swan Hungary
- 2015-2018 nyelvtechnológiai fejlesztő – Lionbridge Technologies Inc.
- 2013- adjunktus – Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Nyelvtechnológiai Csoport

- 2012-2018 tudományos munkatárs – MTA-SZTE Mesterséges Intelligencia Kutatócsoport
- 2005-2011 tudományos segédmunkatárs – Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Nyelvtechnológiai Csoport
- 2009 tudományos segédmunkatárs (részmunkaidőben) – MTA Nyelvtudományi Intézet, Nyelvtechnológiai és Élőnyelvi Osztály, Nyelvtechnológiai Kutatócsoport
- 2000-2005 annotátor – Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Nyelvtechnológiai Csoport

## PROJEKTEK

- 2015-2016 *INFRA: Nyílt, integrált magyar nyelvtechnológiai kutatási infrastruktúra fejlesztése.* A Magyar Tudományos Akadémia támogatásával.  
Státusz: kutató, számítógépes nyelvész szakértő
- 2013-2017 *PARSEME: PARSing and Multi-word Expressions. Towards linguistic precision and computational efficiency in natural language processing.* ICT COST akció az Európai Unió támogatásával.  
Státusz: kutató, vezető testületi tag, fiatal kutatók képviselője
- 2013-2017 *Számítógépes eszközök a veszélyeztetett finnugor nyelvek nyelvi revitalizációjáért (FinUgRevita)* a Finn Tudományos Akadémia és az OTKA támogatásával.  
Státusz: kutató, számítógépes nyelvész szakértő
- 2013-2015 *Telemedicina fókuszú kutatások orvosi, matematikai és informatikai tudományterületeken (TOMI)* az Európai Unió és az Európai Szociális Alap támogatásával.  
Státusz: kutató, nyelvész szakértő
- 2012-2014 *Infokommunikációs technológiák és a jövő társadalma (futurICT),* az Európai Unió és az Európai Szociális Alap támogatásával.  
Státusz: kutató, nyelvész szakértő
- 2009-2012 *Modellalapú szemantikus keresőrendszer,* a Nemzeti Innovációs Hivatal támogatásával.  
Státusz: kutató, nyelvész szakértő
- 2007-2008 *Hatékony tudásmenedzsment eszköz kialakítása nyelvészeti és gráfelméleti eszközök integrációjával a hazai és közösségi vámhatóságok részére,* a Nemzeti Kutatási és Fejlesztési Hivatal támogatásával.  
Státusz: kutató, nyelvész szakértő
- 2005-2007 *Magyar ontológia építése és alkalmazása információkinyerő rendszerekben,* a Nemzeti Kutatási és Fejlesztési Hivatal támogatásával.  
Státusz: kutató, nyelvész szakértő
- 2005-2007 *Magyar–angol gépi fordítórendszer,* a Nemzeti Kutatási és Fejlesztési Hivatal támogatásával.  
Státusz: kutató, nyelvész szakértő
- 2002-2004 *Mondatszintaxis gépi tanulása (gépi tanulási módszerek a magyar nyelv szintaktikai szabályainak létrehozására),* az Oktatási Minisztérium támogatásával.  
Státusz: vezető annotátor, nyelvész
- 2001-2003 *Automatikus információkinyerés rövid (politikai, üzleti, piaci) hírekből,* az Oktatási Minisztérium támogatásával.  
Státusz: vezető annotátor, nyelvész
- 2000-2002 *Magyar nyelvi szófaji egyértelműsítő módszer fejlesztése gépi tanulási algoritmusok felhasználásával,* az Oktatási Minisztérium támogatásával.  
Státusz: vezető annotátor, nyelvész

**DÍJAK, KITÜNTETÉSEK, ÖSZTÖNDÍJAK**

- 2017-2018 Fiatal Oktatói, Kutatói Ösztöndíj, Új Nemzeti Kiválósági Program.
- 2017 Nők A Tudományban Kiválósági Díj (információs technológia kategóriában).
- 2016 Herman József Fiatal Kutatói Díj.
- 2015 vendégkutató a Stanford University egyetemen, CSLI, Stanford, USA (2,5 hónap), a PARSEME COST Action keretében.
- 2015 Legjobb Ifjú Kutatói Díj, a XI. Magyar Számítógépes Nyelvészeti Konferencián.
- 2014 Erdős Pál Fiatal Kutatói Ösztöndíj, a *Nemzeti Kiválóság Program – Hazai hallgatói, illetve kutatói személyi támogatást biztosító rendszer kidolgozása és működtetése konvergencia program* keretében.
- 2012 DAAD fiatal kutatói ösztöndíj az Universität Trier egyetemen, Trier, Németország (5 hónap).
- 2008 Második helyezés az *Obesity Challenge* versenyen, a Szeged csapat tagjaként.
- 2005 Negyedik helyezés a Nemzeti Kulturális Örökség Minisztériuma által kiírt *Nyelv és nyelvhasználat a családban* pályázaton.
- 2003-2004 ERASMUS ösztöndíj az Université de Paris III, Sorbonne Nouvelle egyetemen, Párizs, Franciaország (4 hónap).
- 2002-2004 Köztársasági ösztöndíj.
- 2001-2003 Balatonfüred város ösztöndíja.

**OKTATÁSI TAPASZTALAT**

- 2017 *Introduction to Computational Linguistics* – új kurzus PhD-hallgatók részére, angol nyelven, Szegedi Tudományegyetem, Bölcsészettudományi Kar.
- 2016 *Introduction to Computational Linguistics* – új kurzus BA-hallgatók részére, angol nyelven, Szegedi Tudományegyetem, Bölcsészettudományi Kar.
- 2014 *Mesterséges intelligencia* – BA és BSc-hallgatók részére, Szegedi Tudományegyetem.
- 2014- *Programozás nyelvészeknek - Java* – új kurzus MA-hallgatók részére, Szegedi Tudományegyetem, Bölcsészettudományi Kar.
- 2013 *Nyelvészet az informatikában – informatika a nyelvészetben* – új kurzus MSc-hallgatók részére, Szegedi Tudományegyetem, Természettudományi és Informatikai Kar.
- 2013- *Korpuszok a nyelvészeti kutatásban* – új kurzus BA-hallgatók részére, Szegedi Tudományegyetem, Bölcsészettudományi Kar.
- 2013 *A számítógépes nyelvfeldolgozás alapjai* – új kurzus BA-hallgatók részére, Szegedi Tudományegyetem, Bölcsészettudományi Kar.
- 2012- *How to Write Scientific Papers?* – új kurzus PhD-hallgatók részére, angol nyelven, Szegedi Tudományegyetem, Természettudományi és Informatikai Kar, Általános Orvostudományi Kar.
- 2011 *Syntax-Morphology Interface and Natural Language Processing*. Thematic Training Course on Processing Morphologically Rich Languages, MTA Nyelvtudományi Intézet – Helsinkii Egyetem.

- 2004-2007 *English syntax* – BA-hallgatók részére, angol nyelven, Szegedi Tudományegyetem, Angol–Amerikai Intézet.
- 2004- Angol nyelv, magánórák.
- 2004 Angol nyelv, Ságvári Endre Gyakorlógimnázium, Szeged (tanítási gyakorlat).

A 2017/18-as tanévben 3 PhD-hallgató témavezetője vagyok a Szegedi Tudományegyetemen.

## BÍRÁLÓI TEVÉKENYSÉG

Konferenciák: ACL-2013, ACL-MWE 2016, RANLP-2013, EACL-MWE-2014, NAACL-2016, NAACL-2018, NAACL-MWE-2013, NAACL-MWE-2015, MoMa 2015, PolTal-2014, Baltic NLP 2018, Law-CxG-MWE 2018, CoNLL-2013 Shared Task, CoNLL-2010 Shared Task

Folyóiratok: *Journal of Language Modelling* (szerkesztőbizottsági tag), *Language Resources and Evaluation*, *Natural Language Engineering*, *Studies in Computational Intelligence* (Springer series), *BMC Research Notes*, *Applied Linguistics*, *Nyelvtudomány*

## KONFERENCIASZERVEZÉS

- 2017-2018 The PARSEME Shared Task on Identifying Verbal Multiword Expressions
- 2017 The 13th Workshop on Multiword Expressions (MWE 2017)
- 2016 Second International Workshop on Computational Linguistics for Uralic Languages
- 2010 The CoNLL-2010 Shared Task
- 2008 The Fourth Global WordNet Conference (GWC2008)
- 2007- Magyar Számítógépes Nyelvészeti Konferenciák
- 2005-2006 Nyelvészdoktoranduszok 9. és 10. Országos Konferenciája

## PUBLIKÁCIÓK

164 publikáció (87 angol nyelven, 77 magyar nyelven, 27 egyéni, 38-ban első szerző)

1 könyv, 4 könyvfejezet, 15 folyóiratcikk, 2 doktori értekezés, 130 konferenciakötetben megjelent cikk, 12 szerkesztett kötet

h-index = 14, független hivatkozás > 700, Erdős-szám = 3

## NÉHÁNY KIVÁLASZTOTT PUBLIKÁCIÓ<sup>1</sup>

Savary, Agata; Ramisch, Carlos; Cordeiro, Silvio; Sangati, Federico; Vincze, Veronika; QasemiZadeh, Behrang; Candito, Marie; Cap, Fabienne; Giouli, Voula; Stoyanova, Ivelina; Doucet, Antoine 2017. The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In: Markantonatou, Stella; Ramisch, Carlos; Savary, Agata; Vincze, Veronika (eds.) *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*. ACL, Valencia, pp. 31-47.

Vincze, Veronika; Gosztolya, Gábor; Tóth, László; Hoffmann, Ildikó; Szatlóczki, Gréta; Bánréti, Zoltán; Pákáski, Magdolna; Kálmán, János 2016. Detecting Mild Cognitive Impairment by Exploiting Linguistic Information from Transcripts. In: Erk, Katrin; Smith, Noah A. (eds.): *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Berlin, pp. 181-187.

<sup>1</sup> A teljes publikációs lista elérhető:

<https://vm.mtmt.hu/search/docres.php?lang=0&filter=4&SCTrue=1&SCFalse=1&SCNull=1&AuthorID=10027782&tmtbe=4&oszt=0#>

Vincze, Veronika 2014: Uncertainty Detection in Hungarian Texts. In: *Proceedings of COLING 2014*, Dublin, pp. 1844-1853.

Vincze, Veronika; Nagy T., István; Farkas, Richárd 2013: Identifying English and Hungarian Light Verb Constructions: A Contrastive Approach. In: *Proceedings of ACL 2013 (Volume 2: Short Papers)*, pp. 255-261.

Vincze, Veronika; Zsibrita, János; Nagy T., István 2013: Dependency Parsing for Identifying Hungarian Light Verb Constructions. In: *Proceedings of IJCNLP 2013*, pp. 207-215.

Szarvas, György; Vincze, Veronika; Farkas, Richárd; Móra, György; Gurevych, Iryna 2012. Cross-Genre and Cross-Domain Detection of Semantic Uncertainty. *Computational Linguistics – Special Issue on Modality and Negation* 38(2):335-367.

Farkas, Richárd; Vincze, Veronika; Schmid, Helmut 2012. Dependency Parsing of Hungarian: Baseline Results and Challenges. In: *Proceedings of the Thirteenth Conference of the European Chapter of Association for Computational Linguistics (EACL-2012)*, Avignon, France.

Vincze, Veronika; Csirik, János 2010. Hungarian Corpus of Light Verb Constructions. In: *Proceedings of COLING 2010*, Beijing, China.

Farkas, Richárd; Vincze, Veronika; Móra, György; Csirik, János; Szarvas, György 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In: *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task*, Uppsala, Sweden, pp. 1-12.

Vincze, Veronika; Szauter, Dóra; Almási, Attila; Móra, György; Alexin, Zoltán; Csirik, János 2010. Hungarian Dependency Treebank. In: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Vincze, Veronika; Szarvas, György; Farkas, Richárd; Móra, György; Csirik, János 2008. The BioScope Corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* 9 (Suppl 11):S9 doi:10.1186/1471-2105-9-S11-S9.

## **NÉHÁNY KIVÁLASZTOTT KORPUSZ, ADATBÁZIS ÉS ESZKÖZ**

magyarlanc: nyelvi előfeldolgozó lánc magyar nyelvű szövegekre

Szeged (Dependencia) Treebank: magyar nyelvű, kézzel egyértelműsített szövegek (szófaji egyértelműsítés, összetevős és függőségi elemzés, névelemek és félig kompozicionális szerkezetek)

Magyar WordNet

BioScope: biológiai szövegek, melyekben kézzel meg vannak jelölve a bizonytalanságot jelölő nyelvi elemek és azok hatóköre

CoNLL-2010 Shared Task korpuszai: biológiai és Wikipedia-szövegek, melyekben kézzel meg vannak jelölve a bizonytalanságot jelölő nyelvi elemek

Wiki50: Wikipedia-szövegek, melyekben névelemek és többszavas kifejezések vannak kézzel jelölve

S-GAP: az enyhe kognitív zavar felismerésére kifejlesztett, beszélt nyelvi sajátosságokon alapuló szűrő eljárás

## **TUDOMÁNYOS ISMERETTERJESZTÉS**

2018 Résztétel és bemutatkozó előadás tartása a Nők a Tudományban Egyesület által szervezett *Lányok Napja* rendezvényen az MTA székházában, középiskolás lányok részére.

2017 Résztétel számítógépes nyelvészeti szakértőként *A tudomány hangjai* című tudományos ismeretterjesztő rádióműsorban, Kossuth Rádió.

- 2014- Részvétel és bemutatkozó előadás tartása a Szegedi Tudományegyetem Informatikai Intézet által szervezett *Lányok Napja* rendezvényen, középiskolás lányok részére.
- 2014 Részvétel számítógépes nyelvészeti szakértőként a *Többet ésszel* című tudományos ismeretterjesztő rádióműsorban, Klubrádió.
- 2012- A szegedi Nyelvtechnológiai Csoport számítógépes nyelvészeti kutatásainak bemutatása egyetemi hallgatók, programozók és nyelvészek részére az NLP Meetup keretében, Budapest (több alkalommal).
- 2011 Részvétel számítógépes nyelvészeti szakértőként a *Mindentudás Egyeteme 2.0* tudományos ismeretterjesztő televíziós műsorban, Magyar Televízió.
- 2008-2010 Részvétel a Beszéd- és Nyelvtechnológiai Platform tudományos ismeretterjesztő tevékenységében.
- 2008 A szegedi Nyelvtechnológiai Csoport számítógépes nyelvészeti kutatásainak bemutatása egyetemi hallgatók részére a Móra Ferenc Kollégiumban megszervezett tudományos előadás-sorozat keretében, Szeged.

## **SZÁMÍTÓGÉPES ISMERETEK**

Windows

Linux

Microsoft Office

LaTeX

Java

HTML

## **NYELVISMERET**

magyar: anyanyelv

angol: felsőfok (egyetemi diploma)

francia: felsőfok (felsőfokú C típusú nyelvvizsga)

német: középszint (alapfokú C típusú nyelvvizsga)

portugál: alapszint

olasz: alapszint

spanyol: alapszint

latin: alapszint

kínai: kezdő