

KALIVODA ÁGNES

**IGEKÖTŐS SZERKEZETEK
A MAGYARBAN**

Doktori (PhD) értekezés

TÉZISFÜZET

Pázmány Péter Katolikus Egyetem
Bölcsészet- és Társadalomtudományi Kar
Nyelvtudományi Doktori Iskola
Nyelvtechnológia Műhely

Témavezető:

Dr. Prószéky Gábor

egyetemi tanár, az MTA doktora

Budapest

2021

1. Célkitűzés

A disszertációm célja a magyar nyelv igekötős szerkezetének minél teljesebb leírása adatközpontú szemlélettel. Három fő témája (1) az igekötők állományának meghatározása, (2) az igekötős szerkezetek szórendi mintázatainak leírása, valamint (3) az igekötők produktív kapcsolódási mintáinak feltárása. Célja továbbá olyan, szabadon elérhető erőforrások létrehozása, amelyekre további nyelvészeti kutatások épülhetnek, és amelyek nyelvtchnológiai feladatokban is hasznosíthatók.

2. A kutatás módszere

A munkám során korpuszvezérelt megközelítést alkalmaztam, amely szerint a kutatás nagy mennyiségű szöveganyag automatikus elemzéséből indul ki, és elsődleges célja az introspekcióval nem feltárható jelenségek kimutatása.

Korpuszok: A legtöbb korpuszmérésem alapjául a Magyar Nemzeti Szövegtár 2.0.4 verziója szolgált (Oravecz et al. 2014), amely a 20–21. századi magyar nyelv kiegyensúlyozott korpuszának tekinthető. Emellett három további korpuszt vontam be a kutatásba egy diakrón korpuszvizsgálat keretében. Az ómagyar kori adatok (896–1526) forrása az Ómagyar Korpusz, amely tartalmazza az összes fennmaradt ómagyar és néhány középmagyar szöveg-empléket (Simon és Sass 2012). A középmagyar kori (1526–1772) szövegek vizsgálatához a Történeti Magánéleti Korpuszt használtam (Dömötör et al. 2017), amelyben az élőnyelvhez közel álló írásos anyagok érhetőek el: boszorkányperek során készült jegyzőkönyvek, valamint misszilisek, levelek. Az újmagyar kori (1772-től napjainkig tartó) szöveganyag a Magyar Történeti Szövegtárból származik (Ittész 2009).

Nyelvtchnológiai eszközök: Több esetben szükség volt a meglévő korpuszannotáció javítására vagy bővítésére, ehhez az emtsv (Indig et al. 2019; Váradi et al. 2018), valamint a magyarlanc 3.0 (Zsibrita et al. 2013) nyelvfeldolgozó rendszereket használtam. A legnagyobb módszertani kihívást az igekötők produktív kapcsolódási mintái jelentették. Itt szükségessé vált a különféle igeképzők pontos azonosítása, az igekötős igék vonzatkereteinek kinyerése, valamint a szemantikailag hasonló szócsoportok felismerése. Az első feladathoz az emMorph morfológiai elemző rendszert (Novák et al. 2016, 2017), a másodikhoz Sass (2011) módszerét, a harmadikhoz word2vec szóbeágyazást (Siklósi és Novák 2016) alkalmaztam. Bizonyos nyelvi jelenségekkel kapcsolatban saját algoritmusokat dolgoztam ki: ilyen például az elvált igekötő pozíciójának megállapítása az igezőhöz képest, valamint a különféle hangzássémák azonosítása.

3. A disszertáció felépítése és főbb tézisei

A dolgozat egy rövid bevezetéssel kezdődik, amelyet a kutatás módszertani háttérének ismertetése követ. Ezután egy-egy nagy fejezetben foglalkozom a főbb kutatási kérdéseimmel, amelyek a következők: (1) Mely lexikai elemeket soroljuk az igekötők közé, és mi alapján végezzük a besorolást? (2) Milyen mintázattípusokat mutatnak az igekötős szerkezetek, mikor és mennyire távolodhat el az igekötő az igétől, igenévtől, igei származéktól? (3) Hogyan írhatjuk le az igekötők produktív kapcsolódási mintáit, és ez alapján milyen megállapítások tehetők az igekötők szemantikájáról? Végül egy rövid, de a mondanivalóját tekintve nagyon lényeges fejezetben visszatérek a dolgozat elején bevezetett megközelítés kiértékelésére, majd ennek a problémáit látva felvázolok egy másik megközelítést, amelyben az igekötős szerkezetek kutatása nem az igekötőkből, hanem a szerkezetekből indul ki. A disszertációt a kutatás főbb eredményeinek összefoglalásával és a téziseim megfogalmazásával zárom. Az alábbiakban részletezem azokat a fejezeteket, amelyekhez a téziseim kapcsolódnak.

A második, módszertani fejezetben körüljáróm a korpuszvezéreltség fogalmát, majd ismertetem azt a két erőforrást, amelyet a dolgozat egészében felhasználok. Az egyik a Magyar Nemzeti Szövegtár – MNSZ2 – módosított verziója, amelyből kiszűrtem a verseket, a duplumokat és az idegen nyelven írt mondatokat. A másik a PREVLEX táblázat, amelyhez az első téziseim kapcsolódik:

1. Az MNSZ2 korpusz felhasználásával létrehoztam a PREVLEX-et, amely az igekötős igeik jelenleg legbővebb (53 535 szavas), manuálisan ellenőrzött, nyílt hozzáférésű táblázata. Részét képezik a hapaxok – egyszer előforduló szavak –, valamint azok a szavak is, amelyek a korpuszban UNKNOWN címkével jelennek meg. Az igekötős igeik mellett a korpuszban mért tokengyakoriságuk is szerepel.

A harmadik fejezetben az igekötő-állomány meghatározásának problémájával foglalkozom. Abból a megfigyelésből indulok ki, hogy az igekötők és egyéb, pusztán névszói igemódosítók között nem húzható éles határ, ezért olyan modellre van szükség, amely ezt az átmenetiséget képes megragadni. Feltételezem, hogy a prototípus-elmélet megfelel erre a célra. Ennek keretében olyan automatikusan mérhető, morfológiai és gyakorisági jegyeket veszek sorra, amelyek segíthetnek körülhatárolni az igekötők csoportját. A jegyek értékét korpuszméréssel határozom meg 235 igekötőjelölt esetében. Az adatgyűjtés eredményéből Igekötő \times Jegy mátrixokat hozok létre, amelyek

elsősorban a jegyértékek ábrázolásában különböznek. A mátrixok alapján megvizsgálom az egyes jegypárok közötti korrelációkat. Ennek eredménye, a kapcsolódó szakirodalmak, és végül tagadhatatlanul a saját intuícióm alapján a *meg*-et jelölöm ki prototipikus igekötőnek, majd az igekötőség tipikus jellemzőit a *meg* tulajdonságai alapján határozom meg. Végül bemutatok három módszert, amelyek alkalmasak lehetnek az igekötők jegyalapú osztályozására. Ezek összehasonlítása után Smith et al. (1988) módszere mellett döntök, amely alapján négy, leginkább a tárgyalást segítő kategóriát nevezek meg: a prototipikus (pl. *meg, el*), a centrális (pl. *szét, vissza*), a félperiférikus (pl. *agyon, félbe*) és a periférikus (pl. *szénné, létre*) igekötőket. Főbb eredményeim a következők:

2. Megállapítottam és korpuszon kimértem 10 olyan jegyet, amellyel az igekötők jellemezhetők – ugyanakkor jeleztem, hogy nem minden jegy egyformán fontos. A korpuszmérések eredményét nyílt hozzáférésű Igekötő × Jegy mátrixokban rögzítettem.
3. Az abszolút gyakoriságot tartalmazó Igekötő × Jegy mátrixon kiszámoltam az egyes jegypárok közötti korrelációkat. Ezáltal mennyiségi mutatókkal jellemeztem a grammatikalizáció folyamatát az igekötők esetében. A produktivitás erős pozitív korrelációban áll a gyakorisággal, míg a szótagszám és a szerkesztettség ezekkel negatív korrelációt mutat – a gyakori és produktív igekötők jellemzően röviddek és morfológiai szempontból bonthatatlanok. A korreláció-számítást bináris adaton is elvégeztem, és kimutattam, hogy a jegyek közti alapvető összefüggések nem változnak attól, ha a gyakoriságra vonatkozó információt elhagyjuk. Ezt azzal magyaráztam, hogy a gyakoriság történetileg olyan szorosan együtt jár más jegyekkel – éppen a grammatikalizációs folyamat révén –, hogy akkor is érezhető a hatása a vizsgált jegyekben, ha külön jegyként nem vesszük figyelembe.
4. Smith et al. (1988) módszerét alapul véve felállítottam egy tipikalitási skálát, amelynek egyik végpontján a prototipikus igekötők, másik végpontján a puszta névszói igemódosítók helyezhetők el.

A negyedik, mondattani tárgyú fejezetben először egy szinkrón korpuszvizsgálatot mutatok be, amely az MNSZ2 korpusz 20–21. századi szövegein alapul. Főként az olyan szerkezetekre fókuszálok, amelyekben az igekötő elválhat az igétől, igenévtől, vagy ritkábban egyéb igei származéktól. Ezeket elsősorban az igekötőik eloszlási mintái alapján jellemzem. Ezután egy diakrón korpuszvizsgálat eredményeiről számolok be, amely során a prototipikus

igekötők mondatbeli pozícióját mérem ki a finit igéhez viszonyítva az ómagyar kortól napjainkig terjedő szöveganyagon. A kutatásom főbb eredményei a következők:

5. Kimutattam, hogy a prototipikus igekötők relatív gyakoriságukat tekintve közel maradnak a finit igéhez, míg a periférikusabb igekötők jobban eltávolodhatnak. Ehhez kapcsolódva megneveztem két olyan tényezőt, amely valószínűleg hatással van arra, hogy a hátravetett igekötő mennyire távolodik el a finit igétől. Az egyik a szöveg szerkesztettségi foka: szerkesztetlen – főleg szóban elhangzott – szövegben nagyobb annak a valószínűsége, hogy az igekötő távol kerül a finit igétől, mint gondozott, szerkesztett szöveg esetében. A másik az ige utáni mondatösszetevők, köztük a hátravetett igekötő fonológiai súlya. A méréseim alapján az a tendencia rajzolódik ki, amelyet a növekvő összetevők törvényének vagy Behaghel-törvénynek is neveznek, és É. Kiss (2007) szerint a magyarban a finit ige utáni mondatszakaszra érvényes: a rövidebb összetevő megelőzi a hosszabbat, ha ezt szintaktikai szabály nem gátolja. Ez azzal is egybevág, hogy az egy szótagú, prototipikus igekötők ritkábban távolodnak el a finit igétől, mint a több szótagú, periférikusabb igekötők.
6. Korpuszméréssel igazoltam, hogy az infinitívusztól az igekötője preverbálisan messzire kerülhet, de csak akkor, ha segédigeszerű elem – elsősorban finit ige – férkőzik be a szerkezetbe, és az igekötő az ehhez tartozó igemódosítói pozíciót foglalja el (pl. *össze lehetne őket többékevésbé objektív módon is mérni*). Ez alátámasztja azt a megfigyelést, hogy bár az igekötő az infinitívusszal alkot szótári egységet, szórendi és prozódiai szempontból szorosabban kapcsolódik a segédigeszerű elemhez (Kálmán C. et al. 1989). Megállapítottam, hogy ezzel rokon szintaktikai viselkedést mutat a létigével passzív szerkezetet alkotó határozói igenév is. Ennek az igekötője akkor kerül távoli preverbális helyzetbe, ha a létige előtti igemódosítói pozíciót foglalja el (pl. *ki vannak ezek a marketinges dolgok találva*).
7. Megállapítottam, hogy a határozói igenév igekötője csak akkor vethető hátra, ha az igenév állapot- vagy módhatározói szerepet tölt be (pl. *ezzel szoktatva át*). A létigével alkotott passzív konstrukcióban az igekötő mindig preverbális (pl. *el van intézve*).
8. A melléknévi igevek kapcsán a következő megfigyeléseket tettem: (1) Ha a -hAtÓ végződésű igenév állítmányi helyzetben van, és a tagmondatban nem hangzik el finit ige, akkor a -hAtÓ teljes mértékben át-

veszi ennek a helyét (pl. *az mindig akkor vonható már csak le*). Az igekötője ugyanúgy elválhat, mint egy finit igéé, az igekötők disztribúciója is hasonlóan alakul, emellett a szerkezetbe férköző szavakat tekintve is párhuzam látszik a finit igei szerkezettel. Mindez az ellen szól, hogy a -hAtÓ-t melléknévképzőnek tekintjük – ahogy például Kiefer (2003) teszi –, mivel az ezzel képzett szónak jellegzetes igei tulajdonságai vannak. (2) Az általános szakirodalmi véleménnyel szemben kimutattam, hogy az -AndÓ végződésű melléknévi igenevek állhatnak állítmányi helyzetben, és ekkor az igekötőjük is hátravethető (pl. *nem tévesztendő össze*), bár a jelenség kétségtelenül ritka (az esetek 1,85%-ára jellemző, ez a vizsgált korpuszban 1624 példát jelent).

9. Kimutattam, hogy a megszakított szórend nagyon elterjedt mintázattípus, amely számos deverbális főnév, melléknév és határozószó esetében is megfigyelhető. Ekkor a *nem, sem, se, is* elemek valamelyike férközik a szerkezetbe (pl. *el is várhatóan, legösszenemillőbb*).
10. Megvizsgáltam azokat a szerkezeteket, amelyekben egy kötőmódú ige igekötője vagy egy kötőmódú ige non-finit igei bővítményének az igekötője egy finit modális ige – tipikusan a *kell* – elé kerül (pl. *el kell, hogy menjek – el kell menjek*). A fontosabb megfigyeléseim a következők: (1) A *hogy* kötőszót tartalmazó és nem tartalmazó változatok gyakorisága közel áll egymáshoz, függetlenül attól, hogy az igekötő finit kötőmódú igéhez, infinitívuszhoz vagy határozói igenévhez tartozik. (2) A finit modális ige és a kötőszó közé kerülhetnek rövid elemek, a finit modális ige és az igekötő igeje között álló kifejezések pedig hasonlóságot mutatnak az infinitívuszi beférközések szerkezetekkel.
11. Megvizsgáltam azokat a szerkezeteket, amelyekben egy igekötős ige infinitívuszi vagy határozói igenévi alakban topikalizálva van, és emellett finit igeként is megjelenik (pl. *feljeleneni azért feljelenítik*). Megállapítottam, hogy tagadó, illetve fókuszos mondatokban a hátravetett igekötő jelenléte opcionális (pl. *betanulni nem tanultam semmit*). A topikalizációs szerkezeteken belül külön vizsgáltam az olyan elliptikus szerkezetek tulajdonságait, amelyekben az igekötő újabb előfordulását egy segédigeszzerű elem követi (pl. *kiírva ki van*).
12. A diakrón vizsgálat keretében kimutattam a hátravetett igekötőt tartalmazó, nem-semleges mondatok arányának növekedését az ómagyar kortól napjainkig. Ez a tendencia egyrészt azzal magyarázható, hogy a fordított szórendű tagadás (pl. *nem mondom meg*) fokozatosan teret nyert az archaikus, megszakított szórendű tagadással szemben (pl.

meg nem mond). Másrészt az is lehetséges magyarázat, hogy az ómagyar kor óta folyamatosan növekszik azoknak a szerkezeteknek az aránya, ahol a fókuszhasználat kötelezővé vált.

Az ötödik fejezetben az igekötők produktív kapcsolódási mintáival foglalkozom. Ezek azonosítására olyan módszert dolgozok ki, amely ‘igekötő – képző – vonzatkeret’ hármassok (pl. *el-X-elődik valami, tönkre-X-el valamit*) korpuszvezérelt vizsgálatára épül. Ismertetem az igealkotás három leggyakoribb módját: a névszói és az igei bemenetű igeképzést, valamint a hangzásséma-alapú igealkotást. Ezután bemutatom a PREVCONS adatbázist, amely 21 038 igekötős igei hapaxot tartalmaz, és a fentebb említett hármassok kereshetőségével lehetővé teszi a produktív kapcsolódási minták feltárását. Végül beszámolok egy olyan kísérletemről, amelynek célja az igekötők jelentéseinek és a jelentések közti viszonyoknak a hálózatszerű ábrázolása a PREVCONS alapján, egy ontológia formájában. A fejezethez kapcsolódó téziseim a következők:

13. Kidolgoztam egy algoritmust a hangzássémára illeszkedő igék (pl. *mormog, dörmög, csemcseg*) azonosítására, amely egyúttal sémátípusokba is rendezi az érintett igéket. Ezáltal kimutattam, hogy bár a nyelvészeti szakirodalom kevés jelentőséget tulajdonít az igealkotás ezen módjának, a hangzássémás igék aránya nem elhanyagolható: ezek jelentik az igekötős igei hapaxok közel egytizedét (9,4%).
14. Kimutattam, hogy az új igekötős igék létrehozásában a névszói bemenetű igeképzés a legjelentősebb. Ez a hapaxok 35,2%-a esetében mutatható ki. Ugyanakkor mindössze 62 – az alakváltozatok összevonása után 56 – igekötő kapcsolódik így képzett igéhez.
15. Létrehoztam a PREVCONS adatbázist, amely az igekötős szerkezetek vizsgálatára alkalmas, nyílt hozzáférésű kutatóeszköz. 21 038 igekötős igei hapaxot tartalmaz a morfológiai szerkezetükre, vonzatkeretükre, szemantikájukra és szövegbeli előfordulásukra vonatkozó információval együtt.
16. Létrehoztam egy nyílt hozzáférésű ontológiát, amely 56 igekötő jelentéseit és a jelentések közti viszonyokat jeleníti meg. Entitásként szerepelnek benne az igekötők és a jelentések, relációként pedig három hagyományos jelentésviszony, a szinonímia, az antonímia és a hipe-ronímia. Az ontológiát síkba rajzolható gráfként ábrázoltam.

A disszertáció végén, **a hatodik fejezetben** visszatérek a kiinduló feltevésemhez, miszerint az igekötőnek nevezett szóosztály a prototípus-elmélet segítségével jól megragadható. Megvizsgálom, hogy a főként morfológiai és

gyakorisági jegyek alapján felállított igeekötő-kontinuum nem változik-e meg lényegesen akkor, ha disztribúciós és szemantikai jegyeket is figyelembe veszünk. A konklúzió az, hogy a skála két végpontja csak minimálisan változik, de a köztes mezőben jelentős eltérések mutatkoznak. Az eredeti megközelítésben tapasztalt bizonytalanság elvezet egy új szemlélethez, amely Kálmán László opponensi véleményéből indul ki, és a lényege az, hogy az egyes lexikai elemekről áthelyezi a fókuszot az azokat tartalmazó szerkezetekre. Főbb eredményeim a következők:

17. Felvázoltam a szerkezetek felőli megközelítés koncepcióját, és rámutattam két előnyére a lexikai elemek felől induló, prototípus-elméleti keretet használó megközelítéshez képest: (1) Nem kell bevezetni olyan, nem szilárd empirikus alapon nyugvó döntéseket, mint amelyek korábban szükségesek voltak. (2) Azáltal, hogy előzetesen nem kategorizáljuk a lexikai elemeket, sokkal kisebb az információvesztés, illetve annak a veszélye, hogy túláltalánosítunk.
18. Létrehoztam a PREVDISTRO adatbázist, amely 49 szerkezettypus korpuszbeli előfordulásait tartalmazza, minden előfordulás esetében feltüntetve külön az igeekötőt és az igelemmát, az igeekötő pozícióját az ige-tőhöz képest, valamint a szerkezetbe beférkőző egyéb szavakat. Emellett elérhető a szerkezetek nagyobb kontextusa – a teljes mondat – is. A 41,5 millió rekordból álló adatbázis nyílt hozzáférésű.

A disszertáció elméleti és gyakorlati szempontból is számos új tudományos eredményt tartalmaz. A gyakorlati haszna elsősorban a PREVLEX, a PREVMATRIX, a PREVCONS, a PREVONTO és a PREVDISTRO létrejötte, amelyek önmagukban is értékes, szabadon elérhető erőforrások. Elméleti téren számos olyan tendenciát mutat ki, amelyek korpuszvezérelt módszer hiányában észrevétlenül vagy sejtés szintjén maradtak volna. Emellett felhívja a figyelmet olyan jelenségekre, amelyek egyáltalán nem ritkák, mégis csak kevés nyelvész érdeklődésére tarthattak eddig számot. A dolgozatban leírt módszerek, ötletek hasznosnak bizonyulhatnak nemcsak az igeekötős szerkezetek, hanem más nyelvi jelenségek adatvezérelt vizsgálatában is.

4. A témában végzett publikációs tevékenység

Publikációk:

- Kalivoda Ágnes 2021. Az igekötők produktív kapcsolódási mintái. *Argumentum* 17: 56–82. <https://doi.org/10.34103/ARGUMENTUM/2021/4>
- Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igekötős igére. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szegedi Tudományegyetem, TTIK, Informatikai Intézet. Szeged. 331–344.
- Kalivoda Ágnes 2018a. Hungarian particle verbs in a corpus-driven approach. In: Gelbukh, Alexander (ed.): *Computational Linguistics and Intelligent Text Processing: 18th International Conference (CICLing 2017), Budapest, Hungary, April 17–23, 2017, Revised Selected Papers, Part I*. Springer International Publishing. Cham. 159–176.
- Kalivoda Ágnes 2018b. Az igekötős igék szintaxisa korpuszvezérelt megközelítésben. In: Scheibl György (szerk.): *Lingdok 17.: Nyelvészdoktoranduszok dolgozatai*. Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. Szeged. 159–176.
- Kalivoda Ágnes – Vadász Noémi – Indig Balázs 2018. MANÓCSKA: A Unified Verb Frame Database for Hungarian. In: Sojka, Petr – Horák, Aleš – Kopeček, Ivan – Pala, Karel (eds.): *Proceedings of the 21st International Conference on Text, Speech and Dialogue (TSD)*. Springer-Verlag. Brno. 135–143.
- Vadász Noémi – Kalivoda Ágnes – Indig Balázs 2018. Egy egységesített magyar igei vonzatkerettár építése és felhasználása. In: Vincze Veronika (szerk.): *XIV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2018)*. Szegedi Tudományegyetem, Informatikai Intézet. Szeged. 3–15.
- Kalivoda Ágnes 2017. Az igekötők gépi annotálásának problémái. In: Ludányi Zsófia (szerk.): *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2017: XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia*. MTA Nyelvtudományi Intézet. Budapest. 100–109.
- Vadász Noémi – Indig Balázs – Kalivoda Ágnes 2017. Ablak által világosan – Vonzatkeret-egyértelműsítés az igekötők és az infinitívuszi vonzatok segítségével. In: Vincze Veronika (szerk.): *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)*. Szegedi Tudományegyetem, Informatikai Tanszékcsoport. Szeged. 3–12.

Indig Balázs – Vadász Noémi – Kalivoda Ágnes 2016. Decreasing Entropy: How Wide to Open the Window? In: Martín-Vide, Carlos – Mizuki, Takaki – Vega-Rodríguez, Miguel A. (eds.): *Theory and Practice of Natural Computing: 5th International Conference*. Springer International Publishing. Cham. 137–148.

Konferencia-szereplések:

Ackerman, Farrell – Kalivoda, Ágnes – Malouf, Robert 2021. A network analysis of Hungarian preverb constructions. *5th American International Morphology Meeting (AIMM5)*. Online, 2021. aug. 26–29. (poszter)

Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igekötős igére. *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged, 2019. jan. 24–25. (előadás)

Kalivoda Ágnes – Vadász Noémi – Indig Balázs 2018. Manócska: A Unified Verb Frame Database for Hungarian. *21st International Conference on Text, Speech and Dialogue (TSD 2018)*. Brno, 2018. szept. 11–14. (poszter)

Vadász Noémi – Kalivoda Ágnes – Indig Balázs 2018. Egy egységesített magyar igei vonzatkerettár építése és felhasználása. *XIV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2018)*. Szeged, 2018. jan. 18–19. (előadás)

Kalivoda Ágnes 2017a. Hungarian verbal particles in a corpus-driven approach. *13th International Conference on the Structure of Hungarian (ICSH13)*. Budapest, 2017. jún. 29–30. (poszter)

Kalivoda Ágnes 2017b. Hungarian particle verbs in a corpus-driven approach. *Computational Linguistics and Intelligent Text Processing – 18th International Conference (CICLing 2017)*. Budapest, 2017. ápr. 17–23. (poszter)

Kalivoda Ágnes 2017c. Az igekötők gépi annotálásának problémái. *XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia (Alknyelvdok 2017)*. Budapest, 2017. febr. 3. (előadás)

Vadász Noémi – Indig Balázs – Kalivoda Ágnes 2017. Ablak által világosan – Vonzatkeret-egyértelműsítés az igekötők és az infinitívuszi vonzatok segítségével. *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)*. Szeged, 2017. jan. 26–27. (előadás)

Indig Balázs – Vadász Noémi – Kalivoda Ágnes 2016. Decreasing Entropy: How Wide to Open the Window? *5th International Conference on the Theory and Practice of Natural Computing (TPNC 2016)*. Sendai, 2016. dec. 12–13. (előadás)

Kalivoda Ágnes 2016. Az igekötős igék szintaxisa korpuszvezérelt megközelítésben. *Nyelvészdoktoranduszok 20. Országos Konferenciája (LingDok 17)*. Szeged, 2016. nov. 30. – dec. 1. (előadás)

Hivatkozások

- Dömötör Adrienne – Gugán Katalin – Novák Attila – Varga Mónika 2017. Ki-
útkeresés a morfológiai labirintusból – korpuszpépítés ó- és középmagyar
kori magánéleti szövegekből. *Nyelvtudományi Közlemények* 113: 85–110.
- É. Kiss Katalin 2007. Az ige utáni szabad szórend magyarázata. *Nyelvtudo-
mányi Közlemények* 104: 124–152.
- Indig Balázs – Sass Bálint – Simon Eszter – Mittelholcz Iván – Kundráth Pé-
ter – Vadász Noémi 2019. emtsv – Egy formátum mind felett. In: Berend
Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számí-
tógépes Nyelvészeti Konferencia (MSZNY 2019)* Szegedi Tudományegye-
tem, TTIK, Informatikai Intézet. Szeged. 235–247.
- Ittész Nóra 2009. A magyar nyelv nagyszótára. In: Fábíán Zsuzsanna
(szerk.): *Szótárírás és szótárírók*. Lexikográfiai füzetek 4. Akadémiai
Kiadó. Budapest. 65–80.
- Kálmán C. György – Kálmán László – Nádasy Ádám – Prószekey Gábor
1989. A magyar segédigék rendszere. *Általános Nyelvészeti Tanulmá-
nyok* XVII: 49–103.
- Kiefer Ferenc 2003. Alaktan. In: É. Kiss Katalin – Siptár Péter – Kiefer
Ferenc (szerk.): *Új magyar nyelvtan*. Osiris. Budapest. 127–199.
- Novák Attila – Rebrus Péter – Ludányi Zsófia 2017. Az emMorph morfoló-
giai elemző annotációs formalizmusa. In: Vincze Veronika (szerk.): *XIII.
Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)*. Szegedi
Tudományegyetem Informatikai Intézet. Szeged. 70–78.
- Novák Attila – Siklósi Borbála – Oravecz Csaba 2016. A New Integrated
Open-source Morphological Analyzer for Hungarian. In: Calzolari, Ni-
coletta – Choukri, Khalid – Declerck, Thierry – Goggi, Sara – Grobelnik,
Marko – Maegaard, Bente – Mariani, Joseph – Mazo, Hélène – Moreno,
Asunción – Odijk, Jan – Piperidis, Stelios (szerk.): *Proceedings of the
Tenth International Conference on Language Resources and Evaluation
(LREC 2016)*. European Language Resources Association (ELRA). Por-
torož. 1315–1322.
- Oravecz Csaba – Váradi Tamás – Sass Bálint 2014. The Hungarian Giga-
word Corpus. In: Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thi-
erry – Loftsson, Hrafn – Maegaard, Bente – Mariani, Joseph – Moreno,
Asuncion – Odijk, Jan – Piperidis, Stelios (szerk.): *Proceedings of the 9th*

- International Conference on Language Resources and Evaluation (LREC 2014)*. European Language Resources Association (ELRA). Reykjavik. 1719–1723.
- Sass Bálint 2011. *Igei szerkezetek gyakorisági szótára – egy automatikus lexikai kinyerő eljárás és alkalmazása*. Doktori értekezés. Pázmány Péter Katolikus Egyetem. Budapest.
- Siklósi Borbála – Novák Attila 2016. Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra. In: Tanács Attila – Varga Viktor – Vincze Veronika (szerk.): *XII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2016)*. Szegedi Tudományegyetem, TTIK, Informatikai Intézet. Szeged. 3–14.
- Simon Eszter – Sass Bálint 2012. Nyelvtudomány és kulturális örökség, avagy korpuszépítés ómagyar kódexekből. *Általános Nyelvészeti Tanulmányok* 24: 243–264.
- Smith, Edward E. – Osherson, Daniel N. – Rips, Lance J. – Keane, Margaret 1988. Combining prototypes: A selective modification model. *Cognitive Science* 12: 485–527.
- Váradai Tamás – Simon Eszter – Sass Bálint – Mittelholcz Iván – Novák Attila – Indig Balázs 2018. E-magyar – A Digital Language Processing System. In: Calzolari, Nicoletta – Choukri, Khalid – Cieri, Christopher – Declerck, Thierry – Goggi, Sara – Hasida, Koiti – Isahara, Hitoshi – Maegaard, Bent – Mariani, Joseph – Mazo, Hélène – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios – Tokunaga, Takenobu (szerk.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA). Miyazaki. 1307–1312.
- Zsibrita János – Vincze Veronika – Farkas Richárd 2013. magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In: Angelova, Galia – Bontcheva, Kalina – Mitkov, Ruslan (szerk.): *Recent Advances in Natural Language Processing (RANLP 2013)*. INCOMA Ltd. Sumen. 763–771.