

## Magyar szerkeztár demó

Sass Bálint<sup>1</sup>, Indig Balázs<sup>2</sup>, Kalivoda Ágnes<sup>1</sup>, Lagos Cortes Mátyás<sup>3</sup>, Lipp Veronika<sup>1</sup>, Makrai Márton<sup>4</sup>, Pethő Gergely<sup>5</sup>, Simon László<sup>1</sup>, Vadász Noémi<sup>6</sup>

<sup>1</sup>HUN-REN NYTK Lexikológiai Intézet, <sup>2</sup>ELTE IK Mesterséges Intelligencia Tanszék, <sup>3</sup>HUN-REN NYTK Általános és Magyar Nyelvészeti Intézet, <sup>4</sup>HUN-REN TTK, <sup>5</sup>Debreceni Egyetem Egy Egészség Intézet és Regensburgi Egyetem Romanisztikai Intézet, <sup>6</sup>Magyar Nemzeti Levéltár

sass.balint@nytud.hu, indig.balazs@inf.elte.hu, kalivoda.agnes@nytud.hu, lagos.matyas@nytud.hu, lipp.veronika@nytud.hu, makrai.marton@ttk.hu, petho.gergely@etk.unideb.hu, simon.laszlo@nytud.hu, vadasz.noemi@mnl.gov.hu

**Kivonat** A konstruktikonok (szerkeztárak) egy nyelv konstrukcióit veszik számba. Az utóbbi években számos nyelvre (például: svéd Lyngfelt és mtsai (2018b), orosz Bast és mtsai (2021), észt Vainik és mtsai (2024)) készülnek ilyen lexikai erőforrások, ebbe a trendbe illeszkedik a magyar változat. A Magyar szerkeztár első verziójának a működését és a <https://szerkeztar.hu> címen szabadon elérhető felületét ismertetjük laptopos bemutató formájában.

**Kulcsszavak:** konstrukció, szerkezet, konstruktikon, szerkeztár

## 1. Háttér

### 1.1. A szavak is konstrukciók

A konstrukciós nyelvtan (Goldberg, 2006) szerint minden olyan nyelvi egység konstrukció (szerkezet), ami felfogható forma–jelentés (vagy forma–funkció) párként, amiben a jelentés vagy funkció nem kikövetkeztethető, hanem megtanulandó. Ide tartoznak mindenekelőtt a fix és szabad elemekkel is bíró összetett egységek, mint például az *áll valamiből*, a *kísérletet tesz valamire* vagy az *őrizetbe vesz valakit valami miatt*. De ide tartoznak a skála egyik oldalán az absztrakt nyelvtani szabályok (konstrukciós sémák (Diessel, 2023)), a skála másik oldalán pedig az egyes szavak, sőt morfémák is. Ma is vannak egymással szembenálló vélemények arról, hogy a szavakat konstrukcióknak tekintsük-e. Hilpert (2014, 2. oldal) híres megfogalmazása szerint „nyelvi tudásunk semmi másból nem áll, kizárólag konstrukciókból” („a person’s knowledge of language consists of nothing but constructions”), tehát a szavak is azok. Janda és mtsai (2020) vagy Haspelmath (2023) viszont amellet érvel, hogy csak a fenti összetett konstrukciókra érdemes koncentrálni.

A Magyar szerkeztár (röviden: Szerkeztár) koncepciójának kiindulópontja, hogy *minden* konstrukciót számba vegyen a morfémáktól a nyelvtani szabályokig. Éppen az a nagy előnye a konstrukciós nyelvtani keretnek, hogy ezt lehetővé teszi, hogy általa a nyelv egésze írható le. Nincs elkülönülés az általában

„szótár”-nak és „nyelvtan”-nak nevezett részek között, mivel egy skálára helyezve azonos típusú entitásként kezel minden nyelvi jelenséget a teljesen konkrétól a teljesen absztrakttig: konstrukcióként. Ennek köszönhetően egy lexikai erőforráson belül ábrázolni tudjuk a nyelvi egységek hálózatában lévő összes kapcsolatot, és a felhasználónak is kényelmes, hogy nem kell különböző lexikai erőforrások, szótárak között váltania, mert a szerkezetárban megtalál mindent.

Ahogy azt Janda és mtsai (2020) is megállapítják, a nyelv ilyenfajta teljes leírására való törekvés joggal tűnik utópisztikusnak. A Magyar szerkezetár projekt is annak tudatában tűzi ki ezt a célt, hogy csak valamilyen mértékű megközelítésére lesz képes.

## 1.2. Szótárból szerkezetár

A szerkezetárak sok esetben az adott nyelv FrameNet-jén (Lyngfelt és mtsai, 2018a) alapulnak. Magyar FrameNet hiányában, illetve mivel a mi megközelítésünk fókusza a konstruktikon felépítése, hálózatos struktúrája, azt az utat választottuk, hogy egy meglévő szótárból kiindulva alakítjuk ki a Magyar szerkezetár tartalmát.

Ez olyan módon történik, hogy a szótári szócikkek mélyén megtalálható kifejezésként vagy szókapcsolatként (azaz speciális jelentéssel bíró egységként) megjelölt szerkezeteket önálló egységekként kiemeljük, és a továbbiakban a szavakkal, illetve a többi konstrukcióval egyenrangú módon kezeljük (1. ábra). A *címszó* terminus mintájára nevezhetjük ezeket az önálló egységeket *címszerkezet*nek. Kiinduló szótárunk az Értelmező kéziszótár (Pusztai, 2003), a fentieknek megfelelően lesz az eredetileg a *fehér* címszó belsejében található *fehér asztal*, *fehér holló* stb. kifejezésekből saját jogú egység.

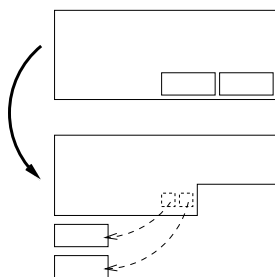
Azt a döntést tehát, hogy mit tekintünk szerkezetnek, a kiinduló szótárra hagyjuk, a szótárban manifesztálódó döntést elfogadjuk. Hangsúlyozzuk, hogy egyfelől a Szerkezetárban jelenleg szereplő egyszavas és többszavas egységek egyaránt kizárólag a kiindulásul szolgáló Értelmező kéziszótárból származnak; másfelől pedig minden ilyen egységet elfogadunk speciális jelentésűnek, azaz konstrukciónak, akkor is, ha merülnek fel ezzel kapcsolatban kétségek, mint az *asztal-fiók* esetében. A fentieknek köszönhetően minden konstrukcióhoz eleve megvan a hozzá tartozó szótári definíció, és van egy bár nem tökéletes, de megalapozott érvünk arra, hogy mit tekintünk konstrukciónak.

A fenti két fejezet részletesebb kifejtését az (Sass, 2023) cikkben olvashatjuk.

## 1.3. Általánosított online felület

A szerkezetárak általában bizonyos szintű nyelvészeti ismeret meglétét feltételezik a felhasználó részéről (vö. Lyngfelt és mtsai, 2018b, 92. oldal). A Magyar szerkezetár célközönsége ezzel szemben a laikusokat is magában foglalja: az a cél, hogy bárki tudja használni akár egyfajta egynyelvű szótárként, akár a nyelvben lévő hálózatok felfedezésére, mindenfajta konstrukciós nyelvtani tudás nélkül.

Ennek megfelelően alakítjuk ki a felületet: a felhasználó egy szövegdobozt lát, ebbe beírhat egy szót, néhány szót vagy egy rövid magyar szövegrészletet, és



1. ábra: Az eredeti szócikk (fent) mélyén található szerkezeteket önálló egységekként kiemeljük, belőlük az eredeti szócikkkel azonos módon kezelendő egységeket képezünk (lent). A szótárbeli egy darab szócikkból a szerkezetárban három egyenrangú konstrukció lett.

a rendszer feladata, hogy ebből az *összes* szerkezetet kihüvelykezze, és egyenként, kapcsolataikkal együtt bemutassa a felhasználó számára. Ez az újfajta felület több szempontból is a hagyományos online szótári felület általánosításának tekinthető az alábbiak miatt.

Egy online szótári felület alapvetően egy egyszavas inputot vár, azt szótári címszónak tekintve kikeresi az adatbázisból, és bemutatja. Az elvárást, hogy az input kanonikus szótári alak legyen, annyival szokták megkönnyíteni a szótárak, hogy a rendhagyó alakokat külön kezelik. A szabályos ragozott alakokat viszont nem, vagy nem teljes lefedettséggel, ezért azokra sokszor nem kapunk választ. Azt sem szokták megtenni az online szótárak, hogy több címszó beírása esetén mindet külön-külön visszaadják eredményként (vö. *black dog* az OALD-ben (Oxford University Press, 2023)).

A Szerkezetár felületén valamilyen módon nyilván meg kell engedni többszavas bemenetet, tekintve, hogy számos konstrukció többszavas. Mivel nem szeretnénk elvárni a felhasználóktól valamiféle kanonikus alak ismeretét – a konstrukciók esetleges kanonikus alakjának definiálása eleve nem triviális kérdés –, a Magyar szerkezetár tetszőleges szöveges inputot megenged. Egy ilyen, bemenetként megadott rövid szövegrészletben viszont több szerkezet is megjelenhet egymással összefonódva. Ezt az esetet is kezeljük: bemutatjuk az összes megjelenő konstrukciót, valamint kezeljük a szabályos alakokat is, felfedve a bennük rejlő konstrukciókat (ld. a (3) példát a 3. részben).

Az általánosítás tehát abban ragadható meg, hogy egyrészt nem várunk el kanonikus alakot, másrészt megengedünk többszavas bemenetet, harmadrészt az inputban megjelenő összes konstrukciót kezeljük.

## 2. Hogyan működik?

Jelenleg az egyszavas és a többszavas bemenetek esetében két külön eljárás működik az összes konstrukció azonosítása érdekében. Az előbbihez tokenizálást (Mittelholcz, 2017) és morfológiai elemzést (Orosz és Novák, 2013; Novák és mt-

sai, 2016) használunk, utóbbihoz szintaktikai elemzést is (Straka és mtsai, 2016), mindezt az *e-magyar* (Indig és mtsai, 2019) rendszerben integrálva.

## 2.1. Egyszavas egységek

Egy önálló, speciális jelentéssel bíró toldalékolt szó egy konstrukció (például: *valójában*), egy kompozicionális jelentésű toldalékolt szó viszont több konstrukcióból áll: minden morféma saját funkcióval bíró külön konstrukció (például: *asztalában*). Az egyszavas egységek kezelését egy elemekre bontásból és összevonásból álló kétlépéses algoritmus valósítja meg. Először morfológiai elemzéssel morfémákra bontjuk a szóalakot, majd az egymás melletti morfémák potenciális összevonásával alakítjuk ki a végleges konstrukciókat. Két vagy több egymás melletti morfémát természetesen akkor tudunk összevonni, ha a szerkezetárban megtalálható konstrukcióként az összevont forma. Az összevonás balról jobbra, a lehető leghosszabb összevont egység kialakításával történik. Az ún. „szavankénti” algoritmus lépései tehát a következők:

1. ha a lekérdezés megtalálható a szerkezetárban, akkor konstrukció, különben
2. morfológiai elemzést végzünk;
3. a morfémákat lehetőség szerint konstrukciókká vonjuk össze;
4. a megmaradó morfémák is önálló konstrukciók lesznek.

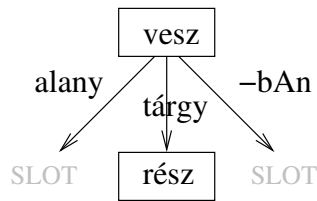
Az *asztalfiókba* (ld. az (5) példát a 3. részben) bemenetet a morfológiai elemzés három elemre bontja, de mivel az eredeti szótárban – és ennek nyomán a Szerkezetárban – önállóan, speciális jelentéssel bíró egységként szerepel az *asztalfiók*, ezért az *asztal* és *fiók* elemeket összevonva ezt és a *-ba/-be* ragot kapjuk meg eredményül.

## 2.2. Többszavas egységek

A többszavas – sok esetben nem fix, nem rögzített szórendű vagy nem folytonos – konstrukciókat függőségi fákhhoz hasonló struktúrák formájában tároljuk az adatbázisban, és ugyanilyen függőségi elemzésnek vetjük alá az input szöveget. Kulcsfontosságú, hogy a Szerkezetárban nyilvántartott szerkezetekben a bennük lévő szabad slotokat beazonosítsuk. Ez jelenleg annyit jelent, hogy a konstrukciók eredeti szótári bejegyzéseiben lévő *valaki/valami* elemet a feldolgozás során szabad slotná alakítjuk. A függőségi elemzés *nmod:obl* (oblikvusz) címke néven egy kalap alá veszi az összes esetraggal megjelenő bővítményt. Számunkra elegendhetetlen ennek felosztása az egyes esetragoknak megfelelően. A céljainknak megfelelő, fentiek szerint módosított függőségi fákat slot-filler fákna nevezzük. Így lesz az eredeti *részt vesz vmiben* alakból a 2. ábrán látható reprezentáció.

Miután közös reprezentációban rendelkezésre állnak a címszerkezetek és az input szöveg, az a feladat, hogy e kettőt egymásra illesszük. Ez a következő ún. „*fatöredék*” rekurzív algoritmussal történik:

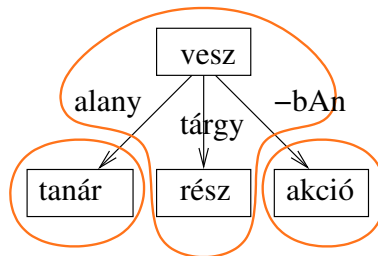
1. végighaladva az adatbázison megnézzük, hogy



2. ábra: A *részt vesz valamiben* szerkezet slot-filler fa reprezentációja a Szerkezetárban.

2. mely szerkezetek illeszkednek az input szöveg fájának gyökerére;
3. ezek közül kiválasztjuk a leghosszabbat (a legtöbb csomópontból állót);
4. a kiválasztottat kivonjuk (eltávolítjuk) a fából;
5. a fa így kisebb megmaradó részekre esik szét;
6. ezeken a részeken rekurzívan újból lefuttatjuk az algoritmust.

A második pontbeli illeszkedés azt jelenti, hogy a fix elemeknek egyezni kell, a címszerkezetek szabad slotjainak megfelelő helyen pedig bármi szerepelhet az inputban. Az algoritmus működésére egy példa a 3. ábrán látható. Látjuk, hogy az algoritmus – helyesen – nem fogja eredményül adni sem a tárgyas *vesz* igét, sem a *rész* szót. Előbbit azért nem, mert van hosszabb illeszkedő szerkezet, utóbbi pedig azért nem, mert ez a bizonyos hosszabb illeszkedő szerkezet már lefedte. Az eljárás elnevezése onnan ered, hogy az egyes szerkezetek végeredményben az eredeti bemeneti fa-reprezentáció részeinek, töredékeinek felelnek meg.



3. ábra: Szerkezetek az *a tanár részt vesz az akcióban* bemenetben. A gyökérnél illeszkedő leghosszabb címszerkezet éppen a 2. ábrán látható szerkezet. A névelőket az egyszerűség kedvéért nem tüntettük fel.

A működés részletesebb bemutatása a (Sass, 2023, 2024) cikkekben olvasható. Azt gondoljuk, hogy a konstrukciók túlnyomó többsége kezelhető a fenti módon. A gyökérben slotot tartalmazó, valamint a függőségi fák formájában esetleg egyáltalán nem megragadható konstrukciók kezelése a jövő feladata.

Megjegyezzük, hogy a fenti két algoritmus struktúrája absztrakt szinten meglehetősen hasonló: mindkettőben van egy elemekre bontó és egy összevonó lépés. A fő különbség, hogy ez az első esetben egy lineáris, a második esetben pedig egy fa-struktúrán történik. A jövőben ez lehetőséget adhat egy egyesített algoritmus kialakítására, ami közös reprezentációban kezeli a morfológiát és a szintaxist.

### 3. Demó

Az alábbiakban példákon keresztül mutatjuk be a Magyar szerkezetár működését.

- (1) *asztal*
- (2) *asztalos*
- (3) *asztalokra*
- (4) *faasztal*

Az (1-2) példák önálló szócikként szerepelnek az eredeti szótárban, egyszerűen az eredeti szótárból származó szócikküket kapjuk eredményül. Utóbbi esetben morfológiai elemzés és összevonás után (vö. a „szavankénti” algoritmus a 2.1. részben). A (3) példa természetesen nem szerepel az eredeti szótárban, a morfológiai elemzésnek köszönhetően a benne lévő három morféma három külön konstrukcióként – három önálló funkcióval bíró egységként – fog megjelenni az eredményben: *asztal* +  $-k$  +  $-rA$ . A (4) esetében hasonlóan: *fa* + *asztal*.

- (5) *asztalfiókba*

Az (5) példát a morfológiai elemzés három elemre bontja, de mivel az *asztalfiók* az eredeti szótárban önálló szócikk, ezért az első két elemet összevonjuk, így az eredmény két konstrukció: *asztalfiók* +  $-bA$ .

- (6) *fehér asztal*
- (7) *sárga asztal*

Említettük, hogy ami a kiinduló szótárban kifejezésként vagy szókapcsolatként szerepel, azt elfogadjuk konstrukciónak. A (6) az eredeti szótárban egy önálló jelentéssel bíró konstrukció, amit kiemeltünk (vö. 1.2), jelentése: ‘étkezésre terített asztal’; (7) ezzel szemben két konstrukció szabad kombinációja: *sárga* + *asztal*. Ez tehát két, formailag azonos, de konstrukciók szempontjából eltérő példa.

Az alábbi példákon érdemes összevetni az eredeti szótár és a Szerkezetár működését (vö. 1.3). Elérhetőségek a 5. részben.

- (8) *avokádó*
- (9) *fűbe harap*
- (10) *fűbe*

(11) *fű*

(12) *aki fűbe harap*

(13) *nem harapott fűbe*

(14) *fűvet harap*

Az eredeti szótárban csak a (8)-ra és a (11)-re kapunk eredményt, a Szerkezetárban mindre. Az előbbi esetében teljesen azonos az kimenet, utóbbi esetében viszont látjuk, hogy a szerkezeteket – köztük (9)-et – önálló egységekként kiemeltük, és csak kereszthivatkozást hagytunk a helyükön (vö. 1. ábra). A (9) természetesen egy konstrukció, szemben a (10)-zel, ami kettő, és a (14)-gyel, ami három. A (12) mutatja, hogy szöveg részeként is felismerjük a szerkezeteket, a (13) pedig azt, hogy ragozott alakban, más szórenddel és nem folytonos formában is. Utóbbihoz – és az alább következő összes példához – már szükséges 2.2. részben ismertetett „fatöredék” algoritmus.

(15) *a tanár részt vesz az akcióban*

(16) *a tanár kenyeret vesz az üzletben*

(17) *az alapító tag részt vesz az akcióban*

(18) *az okos gyerek részt vesz az akcióban*

A fentebb (vö. 3. ábra) ismertetett példát látjuk (15)-ben. A „fatöredék” algoritmusnak köszönhetően felismerjük a *részt vesz valamiben* szerkezetet. A (16) formailag azonos, de itt minden elem külön konstrukció lesz. Hasonlóan a (6-7) példákhoz, az *alapító tag* önálló egység az eredeti szótárban, tehát számunkra egy konstrukció, szemben az *okos gyerek* jelzős főnévvel, ami kettő. Ennek megfelelően ad eredményt a rendszer a (17-18) példákra. A (17) példa azt is demonstrálja, hogy a „fatöredék” algoritmus a rekurzivitásának köszönhetően a slot-filler fa alsóbb részein is képes a többszavas konstrukciók feltárására, egyben példa a konstrukcióknak az 1.2. részben említett összefonódására.

(19) *tanár részt vett az akcióban*

(20) *tanár vesz részt az akcióban*

(21) *tanár vesz az akcióban részt*

(22) *tanár részt vesz a munkában*

(23) *tanár kollégáival részt vesz az akcióban*

A (19-21) példákban látszik, hogy a különböző ragozott és más szórendű változatokat is felismeri az algoritmus, a (22-23) példák pedig azt mutatják, hogy természetesen az eltérő szabad elemek és az esetleges további bővítmények sem zavarják a felismerést.

## 4. Tervek

A számos továbbfejlesztési lehetőség közül az alábbiakat tervezzük megvalósítani először.

A kiinduló szótárnak köszönhetően szavakat és morfémákat, a kiemelésnek köszönhetően pedig fix és szabad elemekkel is bíró összetett szerkezeteket tartalmaz a Szerkezzettár. Viszont egyelőre hiányoznak az absztrakt nyelvtani konstrukciók, mint például a jelző+főnév, a birtokos+birtok vagy az ige+tárgy. Tervben van az ezekkel – vagy az ezek közül a leggyakoribbakkal – való kiegészítés. Ehhez a feladathoz hasznos forrás lehet a Croft (2022)-es kötetének végén található nyelvfüggetlennek és teljesnek szánt gyűjtemény.

Látjuk, hogy egyszavas input esetén, amennyiben ragozott a szó, külön szerkezetekként megkapjuk az egyes morfémákat (ld. a (3) példát), viszont ha összetett szerkezetet azonosítunk be, akkor nem (ld. a (13) és (19) példát). Tervben van ennek a megoldása. A kulcs minden bizonnyal az összetett szerkezet gyökerében lévő szónak a morfológiai elemzése lesz, mivel általánosnak tűnik, hogy az összetett szerkezet morfológiai jellegzetességeit a gyökér hordozza. Ez lényegében a 2. fejezet végén említett algoritmus megvalósítását jelentené.

A kiinduló szótárból első körben az XML struktúrában kifejezésként vagy szókapcsolatként megjelölt elemeket emeltük ki konstrukcióként. Az a tapasztalat, hogy a szótárban példaként megjelenő egységek is legtöbb esetben konstrukciók. Példában találjuk meg többek között a *kísérletet tesz vmire* és a *szívesen vesz* konstrukciót. Ezt meg kell vizsgálni, és ha bebizonyosodik, hogy valóban így van, akkor érdemes a példákat is hozzávenni teljes jogú szerkezetként a Szerkezzettár adatbázisához, jelentősen növelve ezzel lefedettségét.

## 5. Hozzáférhetőség

A felület szabadon elérhető a <https://szerkezzettar.hu> címen minden érdeklődő számára. A működés tanulmányozása során érdemes összevetni a Szerkezzettár által adott kimenetet azzal, amit az ÉKSz. (Pusztai, 2003) <https://eksz.nytud.hu> címen szintén szabadon elérhető online felületén kapunk.

Reményeink szerint ez az újfajta szótárszerű, szótárhelyettesítő lexikai erőforrás hasznos lesz a nyelvészeti vizsgálatokon túl a magyart mint idegen nyelvet tanulók számára, illetve az iskolai magyarórákon is, határon innen és túl.

## 6. Köszönetnyilvánítás

Az ebben a publikációban ismertetett kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hitaval (NKFIH) OTKA (K 147452) projektje támogatta. A K 147452 számú projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a K\_23 pályázati program finanszírozásában valósult meg.



## Hivatkozások

- Bast, R., Endresen, A., Janda, L.A., Lund, M., Lyashevskaya, O., Mordashova, D., Nessel, T., Rakhilina, E., Tyers, F.M., Zhukova, V.: The Russian Constructicon. An electronic database of the Russian grammatical constructions (2021), <https://constructicon.github.io/russian>
- Croft, W.: Morphosyntax. Constructions of the World's Languages. Cambridge University Press (2022)
- Diessel, H.: The Constructicon. Cambridge University Press (2023)
- Goldberg, A.E.: Constructions at work: The nature of generalization in language. Oxford University Press (2006)
- Haspelmath, M.: On what a construction is. *Constructions* 15(1) (2023)
- Hilpert, M.: Construction Grammar and Its Application to English. Edinburgh University Press (2014)
- Indig, B., Sass, B., Simon, E., Mittelholcz, I., Vadász, N., Makrai, M.: One format to rule them all – the `emtsv` pipeline for Hungarian. In: Proceedings of the 13th Linguistic Annotation Workshop. pp. 155–165. Association for Computational Linguistics, Florence, Italy (aug 2019), <https://www.aclweb.org/anthology/W19-4018>
- Janda, L., Endresen, A., Zhukova, V., Mordashova, D., Rakhilina, E.: How to build a constructicon in five years: The Russian example. *Belgian Journal of Linguistics* 34, 162–175 (12 2020)
- Lyngfelt, B., Borin, L., Ohara, K., Torrent, T.T. (szerk.): Constructicography: Constructicon development across languages. John Benjamins, Amsterdam (2018a)
- Lyngfelt, B., Bäckström, L., Borin, L., Ehrlemark, A., Rydstedt, R.: Constructicography at work: Theory meets practice in the Swedish Constructicon. In: Lyngfelt és mtsai (2018a), pp. 41–106
- Mittelholcz, I.: emToken: Unicode-képes tokenizáló magyar nyelvre [emToken: Unicode-aware tokenizer for hungarian]. In: Berend, G., Gosztolya, G., Vincze, V. (szerk.) MSZNY 2017. pp. 61–69. Szegedi Tudományegyetem, Informatikai Intézet (2017)
- Novák, A., Siklósi, B., Oravecz, Cs.: A new integrated open-source morphological analyzer for Hungarian. In: Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odiijk, J., Piperidis, S. (szerk.) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Paris, France (may 2016)
- Orosz, G., Novák, A.: PurePos 2.0: a hybrid tool for morphological disambiguation. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2013). pp. 539–545 (2013)
- Oxford University Press: Oxford advanced learner's dictionary (2023), <https://www.oxfordlearnersdictionaries.com>
- Pusztai, F. (szerk.): Magyar Értelmező Kéziszótár. Akadémiai Kiadó (2003)
- Sass, B.: From a dictionary towards the Hungarian Constructicon. In: Medved', M., Měchura, M., Tiberius, C., Kosem, I., Kallas, J., Jakubíček, M., Krek,

- S. (szerk.) Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference. pp. 534–544. Lexical Computing CZ s.r.o. (2023)
- Sass, B.: The “dependency tree fragments” model for querying a constructicon. In: Š. Despot, K., Ostroški Anić, A., Brač, I. (szerk.) Lexicography and Semantics. Proceedings of the XXI EURALEX International Congress. pp. 275–282. Institute for the Croatian Language (2024)
- Straka, M., Hajič, J., Straková, J.: UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. In: Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (szerk.) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). pp. 4290–4297. European Language Resources Association (ELRA), Portorož, Slovenia (May 2016), <https://aclanthology.org/L16-1680>
- Vainik, E., Paulsen, G., Sakhai, H., Kallas, J., Tavast, A., Koppel, K.: From a dictionary to a constructicon: Putting the basics on the map. In: Š. Despot, K., Ostroški Anić, A., Brač, I. (szerk.) Lexicography and Semantics. Proceedings of the XXI EURALEX International Congress. pp. 209–216. Institute for the Croatian Language (2024)