

KALIVODA ÁGNES – PALÁGYI LÁSZLÓ

Szimultán igekötőzés és igeképzés – konstrukcióegyesítés a magyarban¹

Kivonat

Tanulmányunkban a szimultán igekötőzést és igeképzést mint konstrukcióegyesítést vizsgáljuk adatvezérelt megközelítéssel. Azt feltételezzük, hogy a konstrukcióegyesítés folyamatát a különböző szóképzési konstrukciók hálózatos szerveződése teszi lehetővé, valamint hogy a morfológiai konstrukciók produktivitasában és kidolgozásában a sémáknak és az analógiának egyaránt szerepe van. Kutatásunk alapja az MNSZ2-UD korpusz, amelyből kinyerjük az összes releváns, -(s)Odik és -(s)Ul képzős derivátumot a kontextusukkal, metaadataikkal és különféle formai jellemzőikkel együtt. Ez összesen 1,1 millió példányt jelent, amelyeket szabadon elérhető adattábla formájában is közzéteszünk. A kutatásunk célja az, hogy feltérképezzük az igekötőket és a passzív állapotváltozást kifejező igeképzőket egyesítő konstrukciók jellemző példányait, valamint a konstrukciók globális produktivitasát.

Kulcsszavak: konstrukcióegyesítés, igekötők, igeképzők, magyar nyelv, adatvezérelt megközelítés

¹ Szeretnénk köszönetet mondani a tanulmányunk két bírálójának, Ladányi Máriának és Indig Baláznak a cikk gondos átolvasásáért, a sok hasznos javaslatért. Köszönet illeti Sass Bálintot a módszertani kérdésekben nyújtott segítségéért. Kalivoda Ágnes kutatása a PD 142317 számú projekt keretében, az Innovációs és Technológiai Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a PD_22 pályázati program finanszírozásában valósult meg. Palágyi László kutatását az NKFIH K 129040 számú pályázata (A magyar nyelv igei konstrukciói. Használatalapú konstrukciós nyelvtani kutatás) támogatta.

1. Bevezetés

Noha az igekötőzésnek számos szemantikai (lásd Szilágyi N. 1996; Szili 2005a, 2005b; Szili 2009; Tolcsvai Nagy 2013), morfológiai (lásd Kiefer–Ladányi 2000) és szintaktikai (többek között É. Kiss 2004) megközelítése ismeretes, abban osztoznak a különböző megközelítések, hogy az igekötős ige szerkezetét alapvetően egy igekötő és egy ige összeépüléseként értelmezik. Azonban az „igekötő+ige” morfológiai felépítés kapcsán valójában nem mindig tudunk redukcionizmussal élni, a szerkezetet nem tudjuk pusztán az igekötő és az ige összeépüléseként értelmezni, ha az igekötőzés nem létező igrén megy végbe. A jelen tanulmány olyan igekötős igék vizsgálatát célozza, amelyeknek a bázisigéje ún. potenciális szó², más szóval, amelyeknek valójában nincs bázisigéjük. A *bedühödik*, *feldühödik* (?dühödik), *elhidegül* (?hidegül), *bepipul* (?pipul), *elérzékenyül* (?érzékenyül), *elmosolyodik* (?mosolyodik), *elszomorodik* (?szomorodik) igék arra utalnak, hogy a magyar beszélők tudásának részét képezheti egy igekötőt (*fel*, *el*, *meg*, *be*), melléknévi jelentést (*dühös*, *hideg* 'idegen, közönyös', *pipa* 'dühös', *érzékeny*) és ahhoz járuló passzív igeképzőket (-(s)*Odik*/(s)*Ul*)³ egységben tartalmazó szerkezeti mintázat. Használatalapú megközelítésben ez a mintázat konstrukcióként, azaz sematikus forma–jelentés párként írható le. Ennek a konstrukciónak a sajátossága, hogy a hálózatosan gyakran összekapcsolódó konstrukciók (igekötőzés, igeképzés) egyesülése (unification) révén áll elő (Booij 2010a: 41–50). Ugyanakkor pusztán kvalitatív elemzéssel, lexikai alapon nem lehetséges a mintázat teljes feltérképezése, a feltételezett konstrukció adatokkal való alátámasztása nyelvtechnológiai kihívásként is értelmezhető. A tanulmány célja az igekötőket és passzív igeképzőket egyesítő konstrukció megvalósulásainak (konstruktumainak) kinyerése

² A potenciális szó fogalmával Ladányi Mária a tanulmányhoz fűzött véleményében részletesen foglalkozik. E helyen annyit jegyezni meg, hogy a fogalom nem csupán a kompetencia és a performancia, hanem a szinkronia és a diakronia dichotomikus elkülönítését is támogatja, mivel eliminálja a nyelvi változást a szóképzést illetően. Véleményünk szerint azonban a szóképzésben a nyelvi működés és változás elválaszthatatlanul van jelen, és egy új szó létrejötte valamilyen termékeny, szabályszerűséget mutató általánosítás alapján tipikus példája a nyelvi változásnak is. Minthogy – használatalapú keretben – a szóképzési sémák a meglévő (és a sémát felidéző) derivátumokról való beszélői tudásra épülnek, egy új derivátum megerősíti (vagy részben módosítja) magát a sémát is séma és megvalósulás dinamikus viszonyrendszerében.

³ A tanulmányban használt fonológiai, morfológiai és korpusznyelvészeti rövidítések jegyzéke az I. számú mellékletben található.

a korpuszból, azok kvalitatív elemzése, valamint a konstrukció produktivitásának kimutatása a hapaxok vizsgálatával. A tanulmány második részében a kutatásnak elsősorban a szóképzésre és az igekötőzésre vonatkozó elméleti keretét ismertetjük. A harmadik részben a kutatás módszertanát mutatjuk be. A negyedik részben azt tárgyaljuk, hogy a kvantitatív kutatás alapján az egyesített konstrukciónak milyen mintázatai vannak az egyes igekötőket és képzőket illetően. Az ötödik részben a konstrukció produktivitásával foglalkozunk, a hatodik részben pedig a tanulmány összeggel zárul.

2. Elméleti háttérfeltevések

A hagyományos morfológiai leírás alapegysége a morféma, a szavak szerkezetességéért a morfémák kombinációja a felelős, az általánosítások (szabályok vagy sémák) a morfémákra vonatkoznak, a szavak konstituenciája a morfémák mint önálló reprezentációval bíró építőelemek révén állandó (rész → egész). Ezzel szemben az igekötőket és igeképzőket egyesítő konstrukció vizsgálatában leírásunk alapegysége a szó(alak) és a morfológiai konstrukció (Booij 2010a). A szóalak egy fonológiai és egy szemantikai struktúra szimbolikus egysége (Langacker 1987), a szóalak morfológiai jellege a forma és a jelentés szó szintű együttes variálódásával áll elő (vö. Haspelmath–Sims 2010: 3), a kötött morfémák a szóalakok egymáshoz való viszonyából általánosított nyelvi egységek, amelyeknek nincs önálló, lexikai elemektől független reprezentációjuk (egész → rész) (Blevins 2006). A morfológiai konstrukció szó szintű sematikus forma–jelentés pár, amely ugyancsak szóalakokból elvont általánosítás eredménye (Booij 2010a). A szó szóalakok hálózata, míg a (forma és jelentés együttes variálódásához hozzájáruló, vagyis morfológiai) szavak szóképzési hálózatot képeznek, amelyek a szóalakhálózatokhoz hasonlóan többirányú kapcsolódások mentén épülnek föl, és a derivátumok morfológiai jellege szintén konstrukciós sémák révén mutatkozik meg.

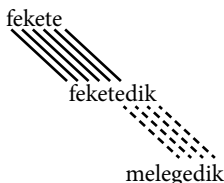
2.1. A szóképzési konstrukciók működése és egyesülése

A konstrukciós szemléletű, szóalapú morfológia a derivátumok struktúrájának elemzését illetően is alternatíváját jelenti a morfémaalapú leírásoknak. Minthogy a szavak derivációs hálózatot képeznek, a szóképzési művelet mindig túlmutat az alapszó és a derivátum viszonyán. Eszerint például az *erősít* derivátum a $[[N-s]_{ADJ-ít}]_V$ általánosítás eredményeként éppúgy leírható

(lásd *erő* → *erős* → *erősít*), mint a [ADJ-ít]v általánosítás eredményeként (*erős* → *erősít*) attól függően, hogy a derivátumot a derivációs hálózat mely részére vetítjük rá. A deriváció hálózatos jellege magyarázatul szolgál a morfémaalapú disztribúciós elemzések azon problémájára, hogy „sokszor nem könnyű eldönteni, hogy két képző egymásutánjáról vagy egy összetett képzőről van-e szó” (Keszler 2000: 315), illetve arra, hogy egy összetett képzési művelet hogyan mehet végbe nem létező szavakon (lásd *tanú* → **tanús* → *tanúsít*). Utóbbi jelenség halmozódása annak is az indikátora, hogy két (vagy több) szóképzési séma az egyesülés (unification) irányában alakul (Booij 2010a: 41–50).

A szóképzési konstrukciók forrásalapú (source-oriented) és kimenetalapú (product-oriented) általánosításokra épülnek (Bybee 2001). A forrásalapú általánosítás információt tartalmaz az alapszó jellegéről, illetve az alapszó és a derivátum közötti formai és szemantikai leképezés módjáról. A kimenetalapú általánosítás viszont csupán azt rögzíti, hogy a szóképzési művelet eredményének milyen formai és szemantikai sémát kell megvalósítania. A forrásalapú általánosítás tehát különböző sémák hálózatos, asszociatív viszonyára épül, például meghatározhatja a konstrukció fonológiai és/vagy szemantikai alkalmazási tartományát/hatókörét, formai és jelentésbeli kötöttségeit (lásd Ladányi 2017: 550–554). A szóképzési konstrukcióban a forrásalapú és kimenetalapú általánosítások különböző mértékben játszanak szerepet. Mind alaki, mind szemantikai kidolgozás szempontjából kimenetalapú sémát képviselnek például a *battyog* és *billeg* típusú, [(C)(C)VC:Og] alaki mintázatot kidolgozó onomatopoeitikus igék (Kalivoda 2021), amelyeknek gyakran nincs is alapszavuk, vagyis a jelentésük és alakjuk pusztán kimenetalapú általánosítások felől értelmezhető. Ebben az esetben a hangzásséma hangsúlyozottan globális (Szilágyi N. 2013) abban az értelemben, hogy a szerkezet részletezőbb jelentésű alkotórészeiről való tudásunk (például *batty*-, *bill*-) sem választható el a derivátumtól, csupán a teljes szerkezetben érvényesül. A forrásalapú és kimenetalapú általánosítások egyaránt fontosak lehetnek a konstrukcióról való tudásunkat illetően. A tanulmányban tárgyalt, passzív igeképző konstrukciót képviselő *sárgul*, *barnul*, *szürkül*, *kékül*, *zöldül*, *őszül* igék szemantikai kidolgozása során fontos szerepet játszik az a forrásalapú általánosítás, hogy a derivátum a névszói komponensek (például színnevek) által felidézett állapotba kerülést jelöli. Az ige alaki kidolgozása során ugyanakkor az a kimenetalapú általánosítás is fontos szerepet játszik, hogy a derivátum két vagy három szótagos, illetve *-CUL* szótagra végződik. Az egy szótagos melléknevek és két szótagos, mássalhangzóra végződő melléknevek tehát egyszerűen kiegészülnek az *-UL* végződéssel

(*kék* → *kékül*, *zöld* → *zöldül*, *halvány* → *halványul*), a két szótagos, magánhangzóra végződő melléknevek viszont úgy dolgozzák ki a konstrukciót, hogy az alapszó második szótagjának magja helyett a toldalék *u* vagy *ü* hangja szerepel, így illeszkedve a mintázat -*CUl* végződéséhez: *sárga* → *sárgul*, *szürke* → *szürkül*. Az adott színt tehát a *sárg-*, *barn-*, *szürk-* *kék-*, *zöld-* tövek egyaránt felidézik, a sémát a két szótagos, magánhangzóra végződő és az egy szótagos, mássalhangzóra végződő melléknevek különbözőképpen, de hasonló eredménnyel dolgozzák ki: *sár\$gul* ~ *zöl\$dül* ~ *ké\$kül* ~ *ö\$szül* (a \$ jel szótaghatárt jelöl). A magyar színnevek tehát a következő kimenetalapú általánosítást dolgozzák ki ebben a konstrukcióban: [(C)V(C)\$CUl]. A fentebb tárgyalt, passzív állapotváltozást kifejező igei konstrukció forrásalapú általánosítása részben átfedésben van a kimenetalapú általánosítása szerint (egy-egy harmadik személyben, jelen időben és kijelentő módban ikes) legalább négy szótagú, felső szótagszám-határt pedig nem tartalmazó [Adj-*Od(ik)*]v sémával. A három szótagú, vagy két szótagú, de mássalhangzóra végződő szavak nem mindig illeszkednek a [(C)V(C)\$CUl] sémába (*meleg* → **melegül*, *beteg* → **betegül*, *fekete* → ?*feketül*), kidolgozhatják viszont ugyanebben a jelentésben a szótagstruktúrát illetően nem specifikált [ADJ-*Od(ik)*]v sémát, például *melegedik*. Mind az [ADJ-*Od(ik)*]v, mind az [ADJ-*Ul*]v szóképzési séma jelentése: 'az alapszóban megadott tulajdonságúvá válik'. A folyamat nem kötődik cselekvő ágenshez, szereplője az átélő, az *experiens*, például *öreg-ed(ik)*, *sötét-ed(ik)*, *meleg-ed(ik)*, *vékony-od(ik)*. A kimenetalapú általánosítás az [ADJ-*Od(ik)*]v sémánál is a tő „csonkulásához” vezet az alapszóhoz képest, az -*ü*, -*ü* végű névszók a szóvégi magánhangzó nélkül tesznek eleget annak az általános követelménynek, hogy a derivátum alapalakjának -*COd(ik)* végződésűnek kell lennie: (*savanyú* →) *sava\$nyodik* ~ *mele\$gedik* (← *meleg*). Ugyanakkor például a *fekete* alapszó szóvégi -*e* magánhangzója átfedésben van a konstrukció hangzássémájával, ezáltal a *feketedik* harmadik -*e* magánhangzója a részletező jelentésű alapszót és a képzőt is felidézi attól függően, hogy a hálózat mely részeire vetítjük rá, vagyis többszörösen motivált, ahogy az 1. ábrán is látható.



1. ábra. A *feketedik* ige harmadik *-e-* elemének többszörös motiváltsága

Az [ADJ-*Od(ik)*]v konstrukció az *-s* melléknévképzős konstrukcióval egyesített formában is termékeny. Az egyesülés (unification) kiindulópontja, hogy a derivációs hálózatban az *-s* melléknévképzős séma és a passzív állapotváltozást kifejező igei konstrukció paradigmaticusan összekapcsolódik. Az [ADJ-*Od(ik)*]v konstrukciót ugyanis számos *-s* képzős melléknév kidolgozza: például *erős-ödik*, *rozsdás-odik*, *szőrös-ödik*, *poros-odik*, *felhős-ödik*, *házas-odik*. Az egyesülés hátterében feltételezhetően a két képző tömbösödése (chunking) áll, amely folyamat számos *-(V)sOdik* végű igében végbe-mehetett. Az *-s* gyakori tövégi előfordulása révén egyesült az igeképzővel és sematizálódott a konstrukcióban, amelyet ezáltal az alapszó transzparensabb módon dolgozhat ki: *egyszerű-södik* (vö. nincs sem **egyszer-edik*, sem **egyszerű-s* szó a magyarban, azaz a konstrukcióegyesítés nélkül az ige felépítése nem magyarázható meg). A konstrukcióegyesítés leírható úgy is, mint egy többszörös forrásalapú általánosítás kimenetalapúvá válása. A többszörös forrásalapú általánosítás információt tartalmaz a leképezési lánc módjáról. Az *erő* – *erő-s* – *erő-s-ödik* hálózatrész leírható a [[N-(V)s]Adj-*Od(ik)*]v többszörös forrásalapú általánosítás eredményeként. A kimenetalapú általánosítás viszont csupán azt rögzíti, hogy az ige formai oldalon *-sOd(ik)* végződésű, szemantikai oldalon pedig a szó névszói, részletező jelentésével asszociált állapotba kerülést jelzi: [állapot-*sOd(ik)*]v. A konstrukcióegyesítésben tehát fontos szerepet játszik a konstrukciók közeledése a derivációs hálózatban (networking) többszörös forrásalapú általánosítások mentén (lásd [[N-(V)s]Adj-*Od(ik)*]v), ez vezet a különböző konstrukciók egyesüléséhez egy olyan kimenetalapú általánosítás jegyében, amelyet számos, eredetileg különbözőképpen létrejövő derivátum képvisel. Pusztán ebben a(z) alapszó alakját háttérbe helyező) megközelítésben állítható párhuzamba például a *karcsú-sodik* (**karcsús-*) és az *erősödik* szó alakja (lásd még **korszerűs-*, *korszerű-södik*, **ézszerűs-*, *ézszerűsödik*, **nagykorús-*, *nagykorúsodik*, **egyértelműs-*, *egyértelmű-södik*). A számítógépes nyelvészeti eszköztár lehetőséget nyújt

a sémához csak kimenetalapúan kapcsolódó, vagyis vélhetően egyesített konstrukciót képviselő szavak feltérképezésére.

2.2. Az igekötős konstrukció

Mint hogy az igekötő és az ige szemantikai egysége nem, illetve nem feltétlenül jár együtt fonológiai egységességgel (az igekötős ige nem fonológiai szó), az igekötős igék a szó és a szónál nagyobb (szintaktikai) szerkezetek között helyezhetők el a lexiko-grammatikai kontinuumon. Az igekötők az esetragokhoz hasonlóan számos szintaktikai konstrukcióhoz hozzájárulnak, amelyekben a jelentésük erősen konstrukciófüggő. Például az ún. okozott mozgás konstrukcióban az igekötők a tárgyeset, a delatívus és (opcionálisan) az instrumentális mintázatában a morfológiai esetekhez hasonló funkciót töltenek be, jelentésük a teljes konstrukció felől értelmezhető. A *lerúg* ige igekötője például nem feltétlenül a rúgás irányát, hanem a (rúgással) okozott mozgás irányát jelöli a *lerúgta a virágot az erkélyről (a labdával)*, vö. *ledobta a labdát az erkélyről* (Palágyi 2022).

Az igekötős igei konstrukciók feltűnően dinamikusak abban az értelemben, hogy némely jelentésük – lásd irányjelentés, aspektuális jelentés – kifejezetten magas általánossági szinten kezelhető, míg más jelentéseik kidolgozásában nyilvánvaló a már meglévő példányok és az analógia szerepe. Az igekötős ige prototipikus kompozitumszerkezete aszimmetrikus: a sematikusabb szemantikai szerkezetű komponens, az igekötőt kidolgozza egy részletesebb jelentésű komponens, az ige (Tolcsvai Nagy 2013). Azonban az igekötős sémák gyakran a vártnál nagyobb mértékben módosítják az igék jelentését. A példányok szerepét mutatja, hogy az új igekötős igék jelentése pusztán az igekötő + ige kapcsolódásból nem mindig vezethető le, a megértéséhez szükség van analógiás csoportként a hasonló jelentésű példányok aktiválására is: *jár : repül : autózik : utazik ~ bejárta : beutazta : beautózta : berepülte* (+ accusativus, például *Európát*). Ezek a gyakran metaforikus jelentések jól elkülönülnek egymástól, a *le* igekötő például hozzájárul a tranzitív 'felülmúl' (*legyőz*) és az intranszitiv, egy szereplőt profiláló 'passzíválódik/passzívál' jelentéshez is (*leáll*) (Palágyi 2022). Jellemző, hogy az igen specifikus konstrukciós jelentések sokszor viszonylag kevés, ugyanakkor viszonylag gyakori igekötős igére épülnek, mégis produktívak. A következő igék például a [*le-V*] konstrukció 'felülmúl' sematikus jelentését dolgozzák ki: *lefutja/leússza az ellenfelét, ledumálja/leérveli a vitapartnerét*.

Az igekötős igei konstrukciók tehát jól mutatják, hogy a séma mint általánosítás és az analógia mint példány szerepének megkülönböztetése mögött valójában csupán szemléleti különbség rejlik. A mi megközelítésünk szerint a kidolgozás sémaalapúsága (szabályszerűsége) és analógiás (egyedi) jellege pusztán a konstrukciót képviselő mintázatok eltérő méretéből adódó, látszólagos megkülönböztetés (vö. Booij 2010b). Az analógiának is központi motívuma az (akár egyetlen példány valamely tulajdonságán alapuló) általánosító művelet, miképpen az általánosítások (sémák) mind alakilag, mind szemantikailag lexikai nyalábok sematizációi, amelyekkel dinamikus séma-megvalósulás viszonyt tartanak fenn a sematizáció után is (Bybee 2010), vagyis nem a példányaitól független reprezentációk. A sémák alapjául szolgáló lexikai nyalábok feltérképezése korpuszalapon válik lehetségessé.

3. A kutatás módszere

A kutatásunk során adatvezérelt, kvantitatív megközelítést alkalmazunk. A munkánk reprodukálhatósága érdekében elérhetővé tettünk a https://github.com/kagnes/odik_ul repozitóriumban minden olyan scriptet és szöveges segédfájlt, amelyet az adatgyűjtéshez és adatelemzéshez készítettünk. Nyilvánosan elérhetővé tettük az adatgyűjtés eredményeként létrejött adat-táblát is, erről bővebben lásd a 3.4. alfejezetet.

3.1. A felhasznált korpusz

A kutatásunk alapja az MNSZ2-UD korpusz (Kalivoda et al. 2024). Ez a megjelenés előtt álló korpusz az 1,04 milliárd szavas, 20–21. századi szövegeket tartalmazó Magyar Nemzeti Szövegtár 2 (Oravecz et al. 2014) legújabb változata. Számos korpusztisztítási lépésen esett át, amelyek közül a legfontosabbak a paragrafus-szintű duplumszűrés, a nem magyar nyelvű bekezdések automatikus azonosítása és törlése, valamint az 500 tokennél hosszabb mondatok kiszűrése. Az utóbbiak jellemzően nem valódi mondatok, hanem kódrészletek, vagy hibásan feldolgozott weboldalak. Ezáltal a szöveg minősége jobb, de a korpusz mérete kisebb az MNSz2-höz képest (776,9 millió szövegszó).

Az MNSZ2-UD kapcsán fontos még kiemelni, hogy új elemzési szintekkel gazdagodott: az emMorph (Novák et al. 2016) által létrehozott, részletes morfológiai elemzéssel, valamint függőségi elemzéssel, amely a Universal

Dependencies formalizmusát követi (de Marneffe et al. 2021). A kutatásunk számára alapvető fontosságú ez a két új annotációtípus. A részletes morfológiai elemzés abban segít, hogy könnyen megtaláljuk a *-(s)Odik* és *-(s)Ul* képzős igéket, és pontosan azonosítani tudjuk a szóképzés bázisául szolgáló névszókat. A függőségi elemzés pedig abban nyújt segítséget, hogy az igétől elvált igekötőt a megfelelő igetőhöz tudjuk kapcsolni, valamint könnyen meg tudjuk állapítani az ige vonzatkeretét is.

3.2. Az adatgyűjtés folyamata

A korpuszból lekérdeztünk minden olyan mondatot, amely legalább egy *-(s)Odik* vagy *-(s)Ul* képzős finit igét illetve infinitívuszt tartalmaz. Ezeket a részletes morfológiai elemzésben a [_AdjVbz_Ntr/V] címke keresésével lehetett azonosítani. A címke feloldása: olyan igeképző, amely melléknévből intranszitiv igét képez. Az (1) példa a *pirosodik* igeen keresztül illusztrálja azt, hogy a keresett címke hogyan jelenik meg a részletes morfológiai elemzésben.

(1) piros[/Adj|col]odik[_AdjVbz_Ntr/V][Prs.NDef.3Sg]⁴

Azt tapasztaltuk, hogy a mai magyar nyelvre kifejlesztett emMorph elemző bizonyos alakulatokban már nem azonosítja ezeket a képzőket, ilyen például az *elgonoszul* ige (ebben tehát a *gonosz* tő nem jelenik meg). Ezért végeztük egy lemma-alapú keresést is, amelyben csak azt figyeltük, hogy az adott ige lemmája az *-ul*, *-ül*, *-odik*, *-edik*, *-ödik* betűsorok valamelyikére végződik-e. Ezeket a lemmákat külön is elemeztük az emMorph középmagyar nyelvállapotra finomhangolt verziójával. Az elemzés eredményét minden – összesen 5839 darab – lemma esetében átnéztük, és a hamis pozitív találatokat kiszűrtük. Ilyen volt például a *kutyul*, amelynek esetében nem arról van szó, hogy valami kutyává válik. Természetesen előfordulnak olyan találatok, amelyek erősen kontextusfüggők, ilyen például a *vadul* szó. Ez lehet hamis pozitív találat, amennyiben határozói funkcióban, 'vad módon' jelentésben áll (például *vadul forgatja a szemét*), és lehet számunkra releváns találat is, '(lassanként) vaddá válik' jelentésben (például *nehogy állatokat táplálván*

⁴ A példában látható emMorph címkék feloldása: [/Adj|col] = színt jelölő melléknév; [_AdjVbz_Ntr/V] = olyan igeképző, amely melléknévből intranszitiv igét képez; [Prs.NDef.3Sg] = tárgyatlan ragozás, egyes szám harmadik személy

és nevelvén maguk is állatokká vaduljanak). Minden egyes előfordulás kézi átnézésére nem vállalkoztunk, mivel 6,2 millió mondatból álló anyaggal volt dolgunk; az ilyen kétértelmű eseteket egységesen megőriztük. Ezáltal hibás adatok is bekerülnek a vizsgálatainkba, de ez ekkora mennyiségű szöveg automatikus feldolgozása során elkerülhetetlen.

Ezután kinyertük a mondatokból minden egyes *-(s)Odik* vagy *-(s)Ul* képzős ige összes olyan morfológiai és szintaktikai jellemzőjét, amelyet lényegesnek tartottunk a kutatásunk szempontjából. Ezek a következők: lemma, tehát az inflexiók toldalékok nélküli szóalak; igekötő;⁵ információ arról, hogy az igekötő prefixumként fordul elő vagy elválik az igéjétől; igelemma (igekötő nélkül); az igeképzés bázisául szolgáló névszó; igeképző; vonzatok és bővítmények. Ami az igekötőket illeti, a prefixumként megjelenőket a részletes morfológiai elemzés segítségével tudtuk azonosítani, az elváltakat pedig a függőségi elemzés *compound:preverb*⁶ relációja alapján találtuk meg. A vonzatokat és bővítményeket úgy állapítottuk meg, hogy az ige dependenseit kerestük, ezek lemmáját és esetragját nyertük ki. A névutókat is esetragként kezeltük. A (2) példában összefoglalva láthatók az eddig bemutatott jellemzők, azokkal a változónevekkel, amelyeket később az adattáblában is használunk.

(2) Legerősebben két kép *elevenedett meg* az emlékeimben.

→ form:	elevenedett
→ lemma:	meg+elevenedik
→ prev:	meg
→ prevtype:	SEP (azaz: <i>separated</i> 'igétől elvált')
→ verb:	elevenedik
→ root:	eleven
→ suffix:	-(s)Odik
→ argframe_cases:	Ess Ine Nom
→ argframe_long:	erős[Ess] kép[Nom] emlék[Ine]

⁵ Egymás alakvariánsainak tekintettük és azonos alakra normalizáltuk a következő igekötőket: *belé* → *bele*, *bé* → *be*, *föl* → *fel*, *föül* → *felül*, *fönn* → *fenn*, *odább* → *odébb*, *reá* → *rá*, *te* → *tele*.

⁶ A magyar igekötőre nehéz megfelelő angol terminust találni, a nemzetközi szakirodalomban (*verbal*) *prefix*, *preverb*, *coverb* vagy *verbal particle* néven hivatkoznak rá. A UD formalizmus a *preverb* ill. *prev* címkéket használja, ezért mi is ezeknél maradtunk az adattábla készítése során. Ez az elnevezés arra utal, hogy az igekötő leggyakoribb, tipikus pozíciója az ige előtt van, de természetesen más pozíciókban is előfordulnak igekötők.

Az igeképzés bemenetétől szolgáló névszókat megvizsgáltuk word2vec szóbeágyazáson alapuló módszerrel is (Siklósi–Novák 2016).⁷ Az eljárás lényege az, hogy a szavakat egy valós vektortér pontjaiként képezzük le, és ebben a térben a hasonló szavak egymáshoz közel, az eltérőek egymástól távol rendeződnek el. A hasonlóság a választott modelltől függően szintaktikai vagy szemantikai jellegű lehet. Mi az utóbbira fókuszáltunk: arra voltunk kíváncsiak, hogy kirajzolódnak-e szemantikailag hasonló csoportok az igeképzések bemenetétől szolgáló szavak között. Minden találat esetében megjelenítjük annak a klaszternek – csoportnak – a számkódját, amelybe az adott névszó a szóbeágyazás során került. A (3) példásor a 3-as klaszter tagjait illusztrálja. Amint látjuk, ide különféle nemzetiségnevekből képzett szavak kerültek.

- (3) a. Nagypját még Virágnak hívták, az ír körülmények között ez *angolosodott* Bloomra.
 b. A bányászok kezdetben németek voltak, akik később *elhorvátosodtak*.
 c. [...] a második világháború után részben újra *visszaszlovákosodtak*.
 d. E város régi német lakosai a 16-ik században *tótosodtak el* egészen.
 → w2v_cluster: 3

Az adatgyűjtés során megőriztük a *-(s)Odik* és *-(s)Ul* képzős igék mondat-szintű kontextusát is, valamint a dokumentumszintű metaadatokat: a publikálás évét, a stílusregisztert és a fájlnevet. Ezeket mutatja be a (4) példa.

- (4) Jónak nevezte az előterjesztésben, hogy *egységesülnek* a közteheralapok.
 → doc_year: 2008
 → doc_style: press
 → doc_id: 2.press_hu_origo_uzletinegyed_007_2008.vert

⁷ Indig Balázs jelzi a lektori hozzászólásában, hogy a választott szóbeágyazási módszer nem járul hozzá a tanulmányunkban bemutatott eredményekhez. Bár végül valóban nem használtuk fel az eredmények értelmezésében, egy korábbi, hasonló szellemben készült publikáció (Kalivoda 2021) meggyőzően mutatja be a módszer hasznosságát, hozzánk hasonlóan csak a bázisszavakra alkalmazva. Az adattáblában meghagytuk ezt az információt, mert biztosak vagyunk benne, hogy jövőbeli kutatások még hasznát tudják venni.

- left_context: Jónak nevezte az előterjesztésben, hogy
- kwic:⁸ egységesülnek
- right_context: a közteheralapok.

3.3. Módszer az egyesített konstrukciók azonosítására

Miután a korpuszban szereplő összes, *-(s)Odik* és *-(s)Ul* képzős igéhez előálltak a fentebb bemutatott adatok, lehetővé vált annak a meghatározása, hogy ezek milyen valószínűséggel képviselnek egyesített konstrukciót. Arra számítottunk, hogy ha az igekötős alak jóval gyakoribb, mint az igekötő nélküli, az az egyesített konstrukció indikátora lehet. Ezen a ponton felmerül az a fontos módszertani kérdés, hogy pontosan mit értünk „jóval gyakoribb” alatt. Ennek a megválaszolására saját módszert dolgoztunk ki a korpuszadatok ismeretében, amely három számértéken alapszik:

1. Az igelemma tokengyakorisága, az igekötős és igekötő nélküli eseteket egybeszámolva (freqsum). Erre azért van szükség, hogy az adatrítkaságból adódó problémákat elkerüljük. Például az adattáblában a *metálosodik* ige nem szerepel önmagában, csak a *meg* igekötővel együtt (*megmetálosodott*), de mivel ez az egyetlen előfordulása a korpuszban, nem indokolt az egyesített konstrukció jellemző példányaként figyelembe vennünk.
2. Az igelemma igekötős előfordulásainak száma osztva az összes előfordulásának számával (prev_vs_all). Ez az érték azt mutatja meg, hogy az adott ige jellemzőbben áll-e igekötővel, mint igekötő nélkül, azonban arról, hogy egy- vagy többféle igekötővel áll-e, még nem teszünk pontos állítást. Például a korpuszban gyakran, 13747-szer előforduló *bizonyosodik* mindig valamilyen igekötővel jelenik meg, de változó, hogy melyik ez az igekötő (például *be-*, *meg-*, *rá-*, ... *-bizonyosodik*).
3. Az adott igekötő+igelemma előfordulásainak száma osztva az igelemma összes igekötős előfordulásának számával (actprev_vs_allprev). Ezen az értéken keresztül azt láthatjuk, hogy az igelemma mennyire „preferál” egy bizonyos igekötőt. Például a *bizonyosodik* igelemma az esetek 79%-ában a *be* igekötővel áll.

⁸ A korpusznyelvészeten használt *kwic* feloldása *keyword-in-context*, azaz a keresett szó vagy kifejezés.

Az olyan igék esetében van a leginkább okunk egyesített konstrukciót feltételezni, amelyeknél mindhárom felsorolt érték magas: az ige kellően gyakori ahhoz, hogy megalapozott állítást lehessen tenni róla, nagyon gyakran – vagy kizárólagosan – igekötővel áll, ezen belül pedig főként egy konkrét igekötővel szokott előfordulni. Ilyen igéket mutatunk be részletesen a 4. fejezetben.

3.4. Az elkészült adattábla

Az adatgyűjtési folyamat eredményeként egy olyan adattáblát kaptunk, amely a *-(s)Odik* és *-(s)Ul* képzős finit igék és infinitívusok összes, MNSZ2-UD korpuszban szereplő előfordulását tartalmazza. Ez összesen több mint 1,1 millió találatot jelent. Az adattábla formátuma TSV (tab separated values): egy egyszerű táblázat, amelyben az első sor a fejléc, a további sorok az egyes korpuszbeli előfordulásoknak felelnek meg, az oszlopok pedig a vizsgált jellemzőknek. Az oszlopok részletes ismertetése a II. számú mellékletben található. Az erőforrás nyílt hozzáféréssé, a <https://zenodo.org/record/7607661> linken érhető el.

4. Az egyesített konstrukciók jellemző példányai

Az 1. táblázatban szereplő igekötős igék esetében a legvalószínűbb, hogy az egyesített konstrukciót képviselik, azaz esetükben az igekötőről és a képzőről való tudás nem választható szét, hanem feltételezi egymást. A táblázat azokat az igéket tartalmazza – igekötő, bázisszó, képző bontásban –, amelyek legalább 500-szor fordultak elő, az esetek legalább 90%-ában az ige igekötővel áll, illetve az esetek legalább 90%-ában egy meghatározott igekötővel fordul elő. Ladányi Mária javaslatára először ‘adott igekötő – adott képző’ párok szerint rendeztük az adatsort, hogy a tipikus mintázatok azonnal kivehetők legyenek, majd ezen belül az ‘akt. IK %’ értékek csökkenő sorrendjében.

Az általánosságban elmondható, hogy ezen igék a folyamatot a végpont felől figyeltetik meg, minthogy az igekötő általában végpontot helyez a folyamatra (Tolcsvai Nagy 2013). A táblázatban szereplő igék tanúsága szerint ennek számos szemantikai motívuma lehet.

1. táblázat. Azok az igekötős igék, amelyek a legnagyobb valószínűséggel képviselnek egyesített konstrukciót

igekötő	bázisszó	képző	ált. IK %	akt. IK %	gyakoriság
el	külön	-(s)Ul	1,00	1,00	1823
el	híres	-(s)Ul	0,99	1,00	657
el	törpe	-(s)Ul	0,94	0,99	1953
el	homályos	-(s)Ul	0,95	0,98	533
el	hideg	-(s)Ul	0,97	0,97	675
el	sötét	-(s)Ul	0,91	0,96	1511
el	bizonytalan	-(s)Odik	0,99	1,00	2772
el	néptelen	-(s)Odik	0,99	1,00	711
el	értéktelen	-(s)Odik	0,92	1,00	589
el	adós	-(s)Odik	0,99	0,99	2170
el	hatalmas	-(s)Odik	0,96	0,99	1091
el	komor	-(s)Odik	0,93	0,98	552
el	mérges	-(s)Odik	0,94	0,95	1541
meg	valós	-(s)Ul	0,99	1,00	42 513
meg	sebes	-(s)Ul	0,99	1,00	11 892
meg	semmis	-(s)Ul	0,99	1,00	6397
meg	testes	-(s)Ul	0,96	0,99	1450
meg	özvegy	-(s)Ul	0,99	0,94	688
meg	hibás	-(s)Odik	1,00	1,00	1919
meg	honos	-(s)Odik	0,98	0,99	1031
be	népes	-(s)Ul	0,97	0,97	1181
be	lázás	-(s)Odik	0,99	1,00	558
ki	egész	-(s)Ul	1,00	1,00	7918
fel	bősz	-(s)Ul	0,92	0,97	652

A jellemző példányok közül néhány pontszerűen (vagyis nem skalárisan) feldolgozott jelentésszerkezetű eseményt jelöl. A *meghibásodik*, *megsebesül*, *megözvegyül* igék alapvetően két, egymástól lényegileg különböző állapot pontszerű egymásba váltását jelölik (például a *meghal* igéhez hasonlóan), vagyis a meghibásodás, megsebesülés és megözvegyülés feldolgozása

magával vonja a végpont (valami/valaki hibás, sebesült, özvegy stb.) megfigyeltetését. Előbbi igékkel a skalaritás bizonyos mértékű (fogalmi) hiányának szempontjából rokoníthatók a *kiegészül, elkülönül, megsemmisül, megvalósul, megtestesül, meghonosodik* igék (a középfokú melléknemek létrehozásának kérdése ssége a skalaritás fogalmának hiányára utal, vö. *?egészebb, ?különebb, ?semmisebb, ?valósabb* 'létezőbb' értelemben, *?testesebb* 'testté váltabb' értelemben, *?honosabb*). A *megsemmisül* és *megvalósul* párhoz hasonlóan antonímiát látunk továbbá az eseményszerkezetüket tekintve hosszabb időintervallumot, nem pontszerű eseményt jelölő *benépesül* és *elnéptelenedik* igéknél. A szemantikai motívumok kapcsán érdemes hangsúlyoznunk, hogy a jellemző példányok között nem egyszerűen olyan igéket találunk, amelyeknek a fogalmi tartalma motiválja az esemény végpont felőli megfigyeltetését, hanem az egyesített konstrukciót képviselő igék a fogalmi tartalmat – esetleg más alternatíváktól eltérően – a végpont felől figyeltetik meg. Jó példa erre az *elsötétül* ige, amelynek ismeretes a hasonló fogalmi tartalmú, de imperfektív jelentésszerkezetet (is) jelölő, alapvetően skalárisabb természetű, nagyobb időintervallumot átfogni képes *(be)sötétedik* alternatívája is. A jellemző példányok lexikalizálódásának magasabb fokát mutatja továbbá az *elhidegül* ige, amely alternatíváival szemben *(hideg)edik, (hideg)szik* ma már kifejezetten kiterjesztett jelentésben használatos *(el)hidegülnek egymástól*.

5. Produktivitás és hapaxok

Noha a hapaxok (lásd a már említett *megmetálosodik* igét) nem tekinthetők az egyesített konstrukció tipikus példányainak, nyilvánvaló indikátorai az egyesített konstrukció produktivitásának (Baayen 1989). Ezért érdemesnek tartottuk néhány kvantitatív produktivitás-vizsgálat elvégzését is a korpuszból kinyert adattáblán. A továbbiakban rövid áttekintést adunk a produktivitás számszerűsítésének a lehetőségeiről, majd rátérünk az eredményeinkre. Ezek közül elsőként az igekötős és igekötő nélküli, *-(s)Odik* illetve *-(s)Ul* képzős igékre vonatkozó vizsgálat, majd a konkrét igekötők szerint elvégzett vizsgálat eredményeit ismertetjük.

5.1. *Áttekintő a kvantitatív produktivitásról*

Egy produktív szóalkotási minta alapján elvileg tetszőleges számú, szemantikaiilag transzparens szó hozható létre egy adott szemantikai tartományban (Kiefer–Ladányi 2000: 149). A nyelvi rendszer szintjén a produktivitás egy lehetőség: a minden további nélkül létrehozható szavak és szóalakok nem biztos, hogy ténylegesen létrejönnek, és még kevésbé valószínű, hogy benne lesznek egy korpuszban. Emiatt több elméleti morfológus (például Dressler 2003) nem tartja helyesnek a kompetenciaszintű lehetőség performanciaszintű valószínűségeen alapuló vizsgálatát.⁹ Az is külön problémát jelent, hogy a korpuszokból szemantikai információt csak közvetetten tudunk kinyerni, a produktivitás pedig szorosan összefügg a szemantikával. Így a produktivitás mérésére kidolgozott, kvantitatív módszerek szemantikai információt expliciten nem vesznek figyelembe; ezt leginkább különféle gyakoriságokon keresztül próbálják megragadni. Mi azonban úgy gondoljuk, hogy az utóbbi irány, a korpuszadaton alapuló, kvantitatív vizsgálat olyan tendenciákat mutathat ki, amelyek a kompetenciát kutatóknak is relevánsak lehetnek, és amelyekre introspekción alapú módszerrel nem lehetne szert tenni. A használatalapú megközelítések szerint ugyanis a kompetenciaszintű lehetőséget alapvetően a performancia, a nyelvi tapasztalatok határozzák meg, vagyis nincs köztük dichotomikus különbség.

A morfológiai produktivitás kvantitatív meghatározása Baayen (1989) nevéhez köthető. Baayen több típust és ennek megfelelően több mérési módszert definiál, amelyek a produktitásnak különféle aspektusait igyekeznek megragadni.¹⁰ Ezeket az elmúlt több mint harminc évben számos korpusznyelvészeti kutatásban alkalmazták, így az évek során sokat változtak, finomodtak az eredeti módszerek. Napjainkban a legszélesebb körben alkalmazott módszer a **globális produktivitás** (Global Productivity, G) meghatározása (Baayen–Lieber 1991). Ezt négy érték ismeretében állapíthatjuk meg. Az alábbiakban röviden ismertetjük ezeket, zárójelben megadva a leggyakrabban használt angol elnevezésüket, valamint a nemzetközileg sztenderd jelöléseket, amelyekkel a cikkünkben hivatkozni fogunk rájuk:

⁹ A Dressler nevéhez köthető természetes morfológia kvalitatív elemzése a produktivitást illetően azt vizsgálja, hogy egy szóképzési séma milyen akadályok leküzdésére alkalmas (például jövevényszavakon, betűszavakon való alkalmazás), és szintén alkalmas a produktivitás fokozatiságának megragadására. E kvalitatív produktitásfogalomról részletesebben lásd Ladányi (2007: 29–40, 47–50), valamint Ladányi Mária cikkünkről írt bírálatát.

¹⁰ Ezekről bővebben lásd Kalivoda (2019) és Modrián–Horváth–Kalivoda–Tanos (ebben a kötetben). Az utóbbi tanulmány részletesen tárgyal olyan általános módszertani megfontolásokat is, amelyekre a kvantitatív fókuszú kutatásokban tekintettel kell lennünk.

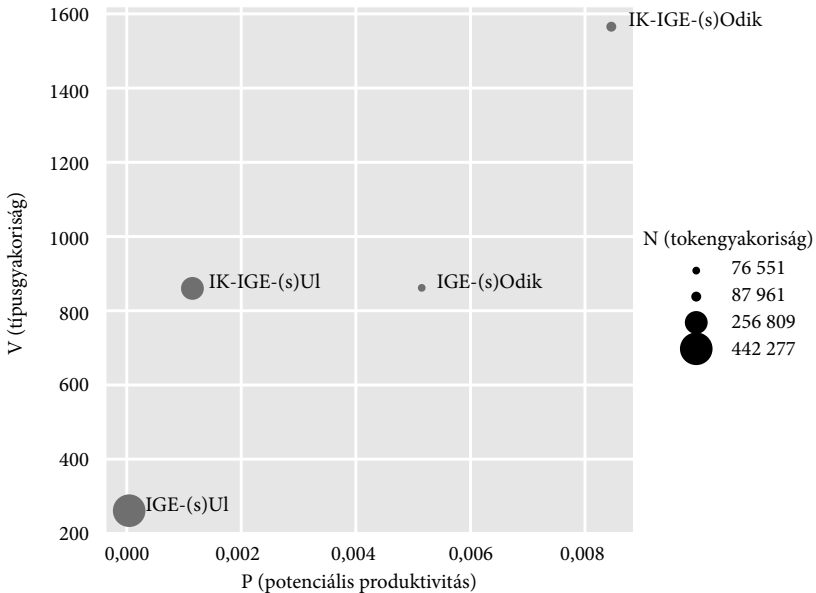
1. **Tokengyakoriság** (token frequency, N): azt mutatja meg, hogy az adott affixum hány példányban adatható a korpuszban.
2. **Típusgyakoriság** (type frequency, V): a szótípusok – esetünkben igelemmák – száma, amelyekben az affixum előfordul.
3. **Hapaxok száma** (hapax legomena, n1): a korpuszban egyszer előforduló szótípusok száma, amelyek az affixumot tartalmazzák. A hapaxok jellemzően produktív szóalkotás eredményei, így a morfológiai szerkezetük és a jelentésük is kompozicionális – legalábbis nagyobb eséllyel az, mint a gyakori szavak esetében.
4. **Potenciális produktivitás** (Narrow Productivity, P): egy származtatott érték, amelyet úgy kapunk, hogy a tokengyakoriságot elosztjuk a hapaxok számával. Arról ad jóságot, hogy az affixumnak várhatóan mekkora szerepe lesz a szóalkotásban a közeljövőben.

A globális produktivitás sajátossága az, hogy nem egyetlen számértékkel fejezhető ki, hanem vizuálisan megjelenítve nyer értelmet. Úgy kapjuk meg, hogy a potenciális produktivitást és a típusgyakoriságot egy-egy dimenzióknak tekintjük – az előbbit jellemzően a koordináta-rendszer X-tengelyén, az utóbbit az Y-tengelyén vesszük fel –, és ebben a két dimenziós térben helyezzük el a vizsgált affixumokat a két számértéküknek megfelelően. Minél magasabb értékei vannak egy affixumnak az egyes tengelyeken, annál produktívabbnak tekinthető. Még szemléletesebb képet kaphatunk a produktív értékekről akkor, ha az affixumok tokengyakoriságát az adatpontok nagyságával szemléltetjük.

5.2. Igekötő nélküli, valamint igekötős -(s)Odik és -(s)Ul szerkezetek

Elsőként azt vizsgáltuk meg, hogy hogyan alakulnak a globális produktív értékek a -(s)Odik és -(s)Ul képzős igék esetében akkor, ha igekötőkkel, illetve ha igekötők nélkül állnak. A mérésünkhöz csak a 2000-ben vagy annál később publikált szövegeket vettük figyelembe.¹¹ Az eredményt a 2. ábrán mutatjuk be, az egyes adatpontok részletes értékei pedig a III/A. mellékletben láthatók.

¹¹ Mindkét bírálónk jelzi, hogy ezt a döntést szükséges lenne indokolni. A döntést megelőző gondolatmenetünk az volt, hogy kvantitatív produktivitás meghatározásánál célszerű minél rövidebb időintervallumot választani, mivel ez egy szinkron korpuszvizsgálatra kidolgozott módszer. Az általunk használt korpusz 1937 és 2014 közötti szövegeket tartalmaz, 77 év összevonása azonban sokat torzíthatott volna az eredményen. Azzal, hogy – önkényesen, de nem ok nélkül – a 2000-es évet jelöltük ki határvonalnak, a vizsgálatot a 21. század elejére szűkítettük le.



2. ábra. Globális produktivitas értékek az igekötő nélküli illetve igekötős, -(s)Odik, illetve -(s)Ul képzős igék esetében

A 2. ábráról leolvasható, és a korábbi szakirodalom alapján (lásd Kiefer–Ladányi 2000: 154–155; Ladányi 2007: 245–247; Ladányi 2017: 608) számítani is lehetett rá, hogy jelentős a különbség a -(s)Ul és -(s)Odik képzők produktivitasában: a -(s)Ul képző önmagában – igekötő nélkül – szinte egyáltalán nem produktív, a -(s)Odik viszont igen. Ugyanakkor megfigyelhetjük azt is, hogy a -(s)Ul tokengyakorisága sokszorososa a -(s)Odik képzőének. Ez főként olyan igéknek köszönhető, amelyek jelentése nem vagy alig transzparens, a képzőt a beszélők többsége feltételezhetően nem is érzékeli ezekben a szavakban. Az öt leggyakoribb -(s)Ul képzős ige a *készül*, *épül*, *bizonyul*, *részesül* és *bövíl*. Az is látható, hogy a tokengyakoriság közel fordítottan arányos a produktivitással – ez alátámasztja Bybee (1985) tézisét, miszerint a tokengyakoriság az adott lexikai egységet erősíti, és nem (vagy legalábbis kevésbé) járul hozzá a mintázat produktivitasához. Az igekötőzés mindkét képző esetében a globális produktivitas jelentős növekedését eredményezi. A típusgyakoriságok azonos mértékben növekednek, a potenciális produktivitas pedig a -(s)Odik esetében jobban, a -(s)Ul esetében valamivel kevésbé. Noha a -(s)Ul képző

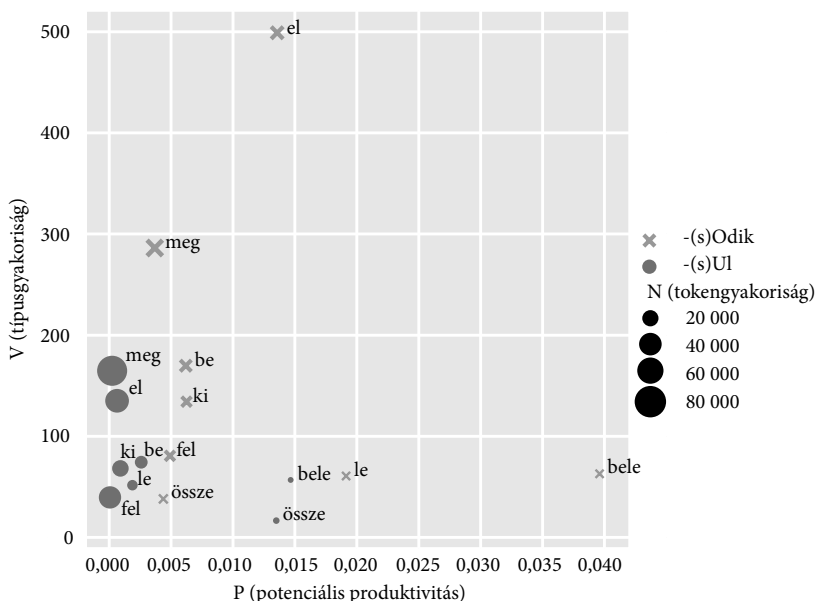
önmagában kevésbé produktív, a táblázatból az is látható, hogy viszonylag termékeny az igekötővel egyesített konstrukció komponenseként. Az (5a) igekötő nélküli, az (5b) igekötős igékre hoz néhány példát.

- (5) a. zöldeskékül, szállóigésül, iparosul, tartósul, férgesül
 b. felnemesül, hozzáaljasul, elmémesül, belerészegül, behomályosul, bedemagógul, elunalmasul, felelevenül, beletéeszésül, begyávuul, hozzáferdül, rálelkesül, szétpuhul

Már a bevezetőben említett, előzetesen megfigyelt példányok esetében is feltűnő volt az, hogy a 2.1 részben tárgyalt szótagszám az *-Ul* igeképzőt illetően kimenetalapú általánosításként nem működik. Az *elérzékenyül* igének például a szótagszáma is arra utal, hogy nem vezethető vissza egy (akár csak potenciális szóként felfogott) *érzékenyül* igére. Az *érzékenyül* ige ugyanis nem lehet potenciális szó sem, mivel nem felel meg az *-Ul* képzős séma szótagszáma vonatkozó általánosításának. A bázisszó tehát az egyesített konstrukció esetében két szótagosnál hosszabb is lehet (lásd még Kalivoda 2021: 60–61), a hapaxok közül ebbe a sorba illeszkedik (5b) adatai közül a *bedemagógul*, *elunalmasul* és a *felelevenül* (vö. *felelevenedik*) ige.

5.3. Igekötők közötti különbségek

Érdeemesnek tartottuk úgy is elvégezni a produktivitas-mérést, hogy az igekötőket külön-külön is vizsgáltuk. Arra számítottunk, hogy ezáltal látszódni fog, melyik igekötő milyen mértékben járul hozzá egy-egy szerkezet produktivitásához. Olyan igekötős, *-(s)Odik* illetve *-(s)Ul* képzős szerkezeteket vizsgáltunk, amelyekben az adott képzővel az adott igekötő legalább ötszázszor előfordult – tokengyakoriságot nézve –, így a vizsgált igekötők a *meg*, *el*, *fel*, *le*, *ki*, *be*, *bele* és *össze* lettek. Az előző méréshez hasonlóan itt is a 2000-ben vagy később publikált szövegeket vettük figyelembe. A 3. ábra mutatja be a mérés eredményét, az egyes adatpontok részletes értékei a III/B. mellékletben láthatók.



3. ábra. Globális produktívitas értékek különböző igekötőkkel ellátott, *-(s)Odik*, illetve *-(s)Ul* képzős igék esetében

A 3. ábra alapján a vizsgált minimálpárok globális produktívitasá úgy alakul, hogy a *-(s)Odik* képzővel együtt álló igekötő egyértelműen produktívabb, mint a *-(s)Ul* képzővel álló. Ez alól csak az *össze* igekötő kivétel: az *össze...-(s)Ul* potenciális produktívitasá magasabb, mint az *össze...-(s)Odik* szerkezeté, de típusgyakoriságát illetően kicsivel elmarad tőle. Ennek azonban pusztán az adatrítkaság az oka: az előbbi konstrukció kilenc példánnyal fordul elő a korpuszban, az utóbbi hússzal, és a két alacsony gyakoriság közötti kontraszt nagyobbak tűnik, mint a magas gyakoriságok közötti. Emellett megfigyelhetjük azt is, hogy az *el...-(s)Odik* és a *bele...-(s)Odik* konstrukciók globális produktívitasá kimagasló, viszont más-más tengelyen. Az előbbinek a típusgyakorisága magas, az utóbbinak pedig a potenciális produktívitasá. Mivel a típusgyakoriság egy olyan érték, amelyet közvetlenül a korpuszból ki tudunk nyerni – szemben a potenciális produktívitasá, amely származtatott érték –, megállapíthatjuk, hogy az *el...-(s)Odik* konstrukció a legproduktívabb a méréseink alapján. Ez a megállapítás összhangban van azzal, hogy a jellemző példányok

(lásd 4. rész, 1. táblázat) között az *el* igekötős igék abszolút többségben, az *el...-(s)Odik* mintázatot képviselő igék pedig relatív többségben vannak.

6. Összegzés

A tanulmányban egy egyedi példányokon alapuló megfigyelést, a szimultán igekötőzést és igeképzést vizsgáltuk kvantitatív alapon, adatvezérelten. A szimultán igekötőzést és igeképzést konstrukcióegyesítésként közelítettük meg, és azt feltételeztük, hogy ezt a folyamatot a különböző szóképzési konstrukciók hálózatos szerveződése teszi lehetővé. További elméleti háttérfeltevésünk, hogy a morfológiai konstrukciók produktivitásában és kidolgozásában a különböző általánosításoknak és a gyakori példányoknak, vagyis a sémáknak és az analógiának egyaránt szerepe van. A korpuszalapú kutatás az igekötőket és passzív állapotváltozást kifejező igeképzőket egyesítő konstrukciók jellemző példányainak és a konstrukciók számszerűsített, globális produktivitásának feltérképezésére irányult. A kutatás alapját az MNSZ2-UD korpuszban található összes *-(s)Odik* vagy *-(s)Ul* képzős finit ige és infinitívusz képezte.

Az egyesített konstrukció jellemző példányainak tekintettük azokat az igéket, amelyeknek az esetében az igekötős ige jóval gyakoribb, mint az igekötő nélküli, illetve az ige kifejezetten egy adott igekötővel fordul elő nagy arányban. A jellemző példányokat három számérték segítségével határoztuk meg. Az egyesített konstrukciót képviselő jellemző példányok esetében megfigyelhető az ige által jelölt állapotváltozás pontszerűsége, vagy legalábbis az állapotváltozás skalaritásának relatív hiánya, ami fogalmi magyarázatul szolgál arra, hogy miért tipikus az adott folyamat igekötővel való, vagyis végpont felőli megfigyeltetése. A jellemző példányok között továbbá több antonimát is találtunk, a pozitív és negatív irányú változásokat kódoló igekötős igék eszerint sok esetben implikálhatják egymást. A tanulmányban megvizsgáltuk az egyesített konstrukciók globális produktivását is. A vizsgálat egyik fő eredménye, hogy azt a szakirodalomból ismert megállapítást, hogy a *-(s)Ul* képző ma már nem produktív, az adatok ismeretében korrigálnunk kell: az igekötő nélküli igével álló *-(s)Ul* képző nagyon kevésbé, az igekötős igével álló viszont egyértelműen produktív. Összességében mind a *-(s)Ul*, mind a *-(s)Odik* képző produktívabb az igekötőkkel egyesített konstrukciók komponenseként, mint önmagukban.

Források

- MNSZ2=Magyar Nemzeti Szövegtár v2.0.5. <http://clara.nytud.hu/mnsz2-dev/index.html>. (Lásd Oravecz–Váradi–Sass 2014.)
- MNSZ2-UD = Magyar Nemzeti Szövegtár v2.0.5 UD treebank. (Lásd Kalivoda et al. 2024.)
- Verbal derivational suffixes in Hungarian: -(s)Odik and -(s)Ul. <https://zenodo.org/records/7607661>. (Lásd a jelen cikket; Kalivoda–Palágyi 2024.)

Irodalom

- Baayen, Harald 1989. *A Corpus-Based Approach to Morphological Productivity (Statistical Analysis and Psycholinguistic Interpretation)*. Doktori értekezés. Amsterdam: Centrum voor Wiskunde en Informatica.
- Baayen, Harald – Lieber, Rochelle 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29 (5): 801–844. <https://doi.org/10.1515/ling.1991.29.5.801>
- Booij, Geert 2010a. *Construction Morphology*. Oxford: Oxford University Press.
- Booij, Geert 2010b. Compound construction: schemas or analogy? A construction morphology perspective. In: Scalise, Sergio – Vogel, Irene (eds.): *Cross-disciplinary issues in compounding*. Amsterdam: John Benjamins. 93–108. <https://doi.org/10.1075/cilt.311.09boo>
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42: 531–573. <https://doi.org/10.1017/S0022226706004191>
- Bybee, Joan 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins. <https://doi.org/10.1075/tsl.9>
- Bybee, Joan 2001. *Phonology and language use*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511612886>
- Bybee, Joan 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511750526>
- de Marneffe, Marie-Catherine – Manning, Christopher D. – Nivre, Joakim – Zeman, Daniel 2021. Universal Dependencies. *Computational Linguistics* 47 (2): 255–308. https://doi.org/10.1162/coli_a_00402
- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Rivista di Linguistica (Italian Journal of Linguistics)* 15 (1): 31–62.

- É. Kiss Katalin 2004. Egy igekötőelmélet vázlata. *Magyar Nyelv* 100: 15–43.
- Haspelmath, Martin – Sims, Andrea D. 2010. *Understanding morphology*. London: Hodder Education.
- Kalivoda Ágnes 2019. Véges erőforrás végtelen sok igekötős igére. In: Berend Gábor – Gosztolya Gábor – Vincze Veronika (szerk.): *XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet. 331–344.
- Kalivoda Ágnes 2021. Az igekötők produktív kapcsolódási mintái. *Argumentum* 17: 56–82. <https://doi.org/10.34103/ARGUMENTUM/2021/4>
- Kalivoda, Ágnes – Malouf, Robert – Ackerman, Farrell – Sass, Bálint 2024. *Building a dependency treebank from the Hungarian Gigaword Corpus*. Kézirat, megjelenés előtt.
- Keszler Borbála 2000. A szóalkotás módjai. In: Keszler Borbála (szerk.): *Magyar grammatika*. Budapest: Nemzeti Tankönyvkiadó.
- Kiefer Ferenc – Ladányi Mária 2000. A szóképzés. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 3. Morfológia*. Budapest: Akadémiai Kiadó. 137–164.
- Ladányi Mária 2007. *Produktivitás és analógia a szóképzésben: elvek és esetek*. Budapest: Tinta Könyvkiadó.
- Ladányi Mária 2017. Alaktan. In: Tolcsvai Nagy Gábor (szerk.): *Nyelvtan*. Budapest: Osiris Kiadó. 503–660.
- Langacker, Ronald W. 1987. *Foundation of Cognitive Grammar (Vol. 1). Theoretical prerequisites*. Stanford: Stanford University Press.
- Modrián-Horváth Bernadett – Kalivoda Ágnes – Tanos Bálint 2024. A magmondatot szerepű, passzív jelentésű sémák produktivitásának kvantitatív vizsgálata. In: Kugler Nóra – Modrián-Horváth Bernadett – Simon Gábor (szerk.): *Korpusznyelvészeti módszerek a kognitív nyelvészeti kutatásban*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 153–203.
- Novák, Attila – Siklósi, Borbála – Oravecz, Csaba 2016. A New Integrated Open-source Morphological Analyzer for Hungarian. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA). 1315–1322.
- Oravecz, Csaba – Váradi, Tamás – Sass, Bálint 2014. The Hungarian Gigaword Corpus. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA). 1719–1723.

- Palágyi László 2022. A metafora, az analógia és a szemantikai megfelelések szerepe a *le* igekötős konstrukciók jelentésében. In: Tátrai Szilárd – Tolcsvai Nagy Gábor (szerk.): *A magyar mondat és kontextuális környezete*. Budapest: ELTE BTK – ELTE Eötvös Kiadó. 15–48.
- Siklósi Borbála – Novák Attila 2016. Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra. In: *XII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged: Szegedi Tudományegyetem, TTIK, Informatikai Intézet. 3–14.
- Szilágyi N. Sándor 1996. *Hogyan teremtsünk világot?* Kolozsvár: Erdélyi Tankönyvtanács.
- Szilágyi N. Sándor 2013. A mi (rend)szertelen nyelvünk. In: Kádár Edit – Szilágyi N. Sándor (szerk.): *Analógia és modern nyelvelírás*. Kolozsvár: Erdélyi Múzeum-Egyesület. 235–263
- Szili Katalin 2005a. A *be* igekötő jelentésváltozásai I. *Magyar Nyelvőr* 129: 151–164.
- Szili Katalin 2005b. A *be* igekötő jelentésváltozásai II. *Magyar Nyelvőr* 129: 282–299.
- Szili Katalin 2009. A *fel*, *le* és egyéb igekötős igék formai-szemantikai viszonyának kérdéséhez. *Magyar Nyelv* 105: 175–188.
- Tolcsvai Nagy Gábor 2013. Az igekötő + ige szerkezet szemantikája. *Nyelvtudományi Közlemények* 109: 187–226.



Simultaneous use of preverbs and verb formation suffixes – construction unification in Hungarian

This paper presents a data-driven approach to explore the simultaneous application of Hungarian preverbs and verb formation suffixes. We consider this phenomenon construction unification and assume that the process is made possible by the network organization of different word formation constructions. Our further assumption is that schemas and analogy play a role in the productivity and development of morphological constructions. The text source used in our research is the HGC-UD corpus, from which we extract all relevant derivatives, their contexts, metadata, and various formal features. We convert the result of this process (more than 1.1 million instances) into a freely accessible dataset. Our research aims to extract the

typical exemplars of the constructions unifying preverbs and verb derivational suffixes expressing a passive change of state, as well as to determine and analyze the global productivity of these constructions.

Keywords: construction unification, preverbs, verb formation suffixes, Hungarian grammar, data-driven approach

Kalivoda Ágnes

HUN-REN Nyelvtudományi Kutatóközpont

kalivoda.agnes@nytud.hun-ren.hu

ORCID: 0000-0003-2520-5523

Palágyi László

Eötvös Loránd Tudományegyetem

palagyilaszlo@btk.elte.hu

ORCID: 0000-0002-1437-4746

I. számú melléklet: Rövidítésjegyzék

Fonológiai jelölések:

- C mássalhangzó
C: hosszú mássalhangzó
V o, ö, e vagy a
O o, ö vagy e
U u vagy ü

Szófaji jelölések:

- N főnév
ADJ melléknév
V ige

A cikkben használt emMorph címkék:

- [/Adj|col] szint jelölő melléknév
[_AdjVbz_Ntr/V] olyan igeképző, amely melléknévből intranszitiv igét képez
[Prs.NDef.3Sg] tárgyatlan ragozás, egyes szám harmadik személy

A produktivitásmérésben használt nemzetközi jelölések:

- G globális produktivitás
N tokengyakoriság
V típusgyakoriság
n1 hapaxok száma
P potenciális produktivitás

Egyéb jelölések:

- ? a szerzők intuíciója alapján kérdéses grammatikalitású szó
* a szerzők intuíciója alapján agrammatikus szó
(...) nem mindig megjelenő elem
[...] séma
\$ szótaghatár

Amint látható, a V és az N jelölés túlterhelt, mert mindig igyekeztünk az adott szakterületen (fonológiában, morfológiában illetve korpusznyelvészetben) sztemderddé vált jelölést alkalmazni. Abban bízunk, hogy a kontextus egyértelművé teszi ennek a két rövidítésnek a feloldását a konkrét esetekben.

II. számú melléklet: Az adattábla oszlopai

Oszlopnév	Leírás
id	A korpusztalálat egyedi azonosítója.
form	A szóalak kisbetűsítve. Ez azt jelenti, hogy ha az adott szó a korpuszban mondatkezdő helyzetben, nagybetűvel állt, itt akkor is kisbetűvel szerepel. Az eredeti írásmód a <i>kwic</i> cellában lesz látható.
lemma	A lemma (inflexiós morfémák nélküli szóalak) kisbetűsítve. Igekötős igék esetében az igekötő és az ige határán + jel áll. A lemmában akkor is feltüntettük az igekötőt, ha az elvált volt.
prev	Az igehez tartozó igekötő kisbetűsítve.
prevtype	A cella értéke PFX, ha az igekötő tapad az igehez, és SEP, ha az igekötő elvált.
verb	Az igelemma kisbetűsítve, minden esetben igekötő nélkül.
root	Az igeképzés bázisául szolgáló névszó kisbetűsítve.
suffix	A képző: <i>-ul/ül/sul/sül</i> esetében <i>-(s)Ul</i> , <i>-odik/edik/ödik/sodik/sedik/södik</i> esetében <i>-(s)Odik</i> .
w2v_cluster	Annak a klaszternek az azonosító száma, amelybe a báziszó – a <i>root</i> mező tartalma – került a szóbeágyazás során.
argframe_cases	Az igenek az adott mondatban megjelenő vonzatai, amelyeket esetragokkal jelölünk.
argframe_long	Az igenek az adott mondatban megjelenő vonzatai, amelyeket lemma+esetrag kombinációkkal jelölünk.
doc_year	A dokumentum keletkezésének vagy publikálásának az éve. Amennyiben ismeretlen az évszám, a cella értéke 0.
doc_style	A dokumentum stílusrétege.
doc_id	A dokumentum egyedi azonosítója.
left_context	A korpusztalálatot megelőző mondatzszakasz.
kwic	A korpusztalálat.
right_context	A korpusztalálatot követő mondatzszakasz.
freqsum	Az igelemma tokenygyakorisága, az igekötős és igekötő nélküli eseteket egybeszámolva.

Oszlopnév	Leírás
prev_vs_all	Az igelemma igekötős előfordulásainak száma osztva az összes előfordulásának számával.
actprev_vs_allprev	Az adott igekötő+igelemma előfordulásainak száma osztva az igelemma összes igekötős előfordulásának számával.

A hiányzó cellaértékek jelölésére alulvonást (_) használunk.

III. számú melléklet: A produktivitásmérésekhez kapcsolódó adatok

A. Igekötő nélküli ill. igekötős -(s)Odik/-(s)Ul:

szerkezet	N	V	n1	P
IGE-(s)Odik	76.551	867	398	0,0052
IGE-(s)Ul	442.277	265	60	0,0001
IK-IGE-(s)Odik	87.961	1.570	751	0,0085
IK-IGE-(s)Ul	256.809	864	307	0,0012

B. Igekötők közötti különbségek:

igekötő	képző	N	V	n1	P
be	-(s)Odik	12.251	169	77	0,0063
be	-(s)Ul	9.210	76	25	0,0027
bele	-(s)Odik	780	63	31	0,0397
bele	-(s)Ul	1.558	57	23	0,0148
el	-(s)Odik	18.160	496	248	0,0137
el	-(s)Ul	66.099	136	44	0,0007
fel	-(s)Odik	6.603	81	34	0,0051
fel	-(s)Ul	48.460	40	11	0,0002
ki	-(s)Odik	8.669	133	55	0,0063
ki	-(s)Ul	21.212	67	22	0,0001
le	-(s)Odik	1.670	61	32	0,0192
le	-(s)Ul	7.505	52	14	0,0019
meg	-(s)Odik	30.587	285	117	0,0038
meg	-(s)Ul	97.539	164	37	0,0004
össze	-(s)Odik	4.496	38	20	0,0044
össze	-(s)Ul	663	17	9	0,0136