

# KORPUSZNYELVÉSZETI ESZKÖZ A MAGYAR IGÉK BŐVÍTMÉNYSZERKEZETÉNEK VIZSGÁLATÁRA

SASS BÁLINT

## 1. Bevezetés: korpuszok a nyelvtudományban

A XX. század második felében a nyelvészet fő irányvonalát a generatív nyelvészet jelentette, de mindig jelen volt ezzel párhuzamosan a korpuszalapú, adatközpontú megközelítés is. Firth (1957) szerint „You can know a word by *the company it keeps*”, azaz egy szót legjobban a környezete alapján ismerhetünk meg. A korpusz-nyelvészet hagyománya a generatív nyelvtan igen/nem grammatikalitási döntéseivel szemben a gyakorisági szempontok fontosságát emeli ki, illetve a valós, élő nyelvhasználat vizsgálatát tartja elsődlegesnek az introspekcióval és a konstruált példamondatok alkalmazásával szemben. Mára a korpuszok alapvető, széles körben használt eszközzé váltak a nyelvtudomány valamennyi területén, azok az állítások az igazán hitelt érdemlőek, melyeket korpuszból származó nyelvi adatokkal támasztanak alá, a korpuszkutatás a nyelvészet egyik kulcsterületévé vált. (Teubert 2005).

Jól elkülöníthető a korpuszok felhasználásának alábbi két módja. A korpuszalapú (corpus-based) felfogásban a korpusz segéd-eszköz, mely empirikus adataival támogatja az intuíciót, mérhetővé teszi a nyelvi jelenségeket, meglévő elméleteket bizonyít/cáfol. A radikálisabb korpuszvezérelt (corpus-driven) megközelítésben viszont a korpusz maga szolgáltatja az „elméletet”, a nyelvész előzetes feltevések és elvárások nélkül fordul az adatokhoz. Minden következtetést kizárólag korpuszmegfigyelésekből von le, minden állítás az empirikus alapokon, korpusz-megfigyeléseken nyugszik (Tognini-Bonelli 2001).

A korpuszok használata a különböző elméleti és alkalmazott nyelvészeti vizsgálódások során nem új ötlet: a lényegi változás az, hogy a mai informatikai környezet lehetővé teszi, hogy nagy méretű

korpuszokat építsünk és hatékonyan kezeljünk nagy mennyiségű nyelvi adatot. Ma viszonylag kis ráfordítással vizsgálhatók nagy méretű szövegek, ami korábban elképzelhetetlen volt. Már Simonyi Zsigmondnál tisztán megjelenik az adatközpontú felfogás a nyelvtanítás kapcsán. „Simonyi új grammatikai módszert akar behozni, könyve inductive halad, azaz a példákból kiindulva tanítja a szabályt, nem pedig dogmatica. A grammaticát tehát valami olvasmány alapján akarja előadni, úgy hogy a szabályokat a tanár tanítványai közreműködésével vonhatja le ésszerű következtetések útján. Ilyenképp tehát ezen módszer véget vet a lelketlen magolásnak, és azt észfejlesztő inductióval pótolja. Eszerint a szabályok is mélyebben vésődnek be a gyermek emlékezetébe, mert amit magunk találunk, azt jobban tudjuk, mint amit más mond vagy más tanultat velünk” (Riedl 1882).

A korpuszhasználat tehát az utóbbi időben a nyelvészet szinte minden területén hódít, mégis tapasztalható ellenérzés, amit általában úgy fogalmazznak meg, hogy a korpusz „csal” vagy „félrevezet”. Jellemző a két alábbi mondat, mindkettő magyar nyelvész szájából hangzott el: Komlósy András szerint „A korpusznál jobban semmi nem vezetheti félre az embert.”<sup>1</sup> míg Kornai András úgy véli, hogy „Manapság már a tisztán introspektív nyelvészre nézünk furán.”<sup>2</sup> Fontos megjegyezni, hogy nem állja meg a helyét az a széles körben elterjedt vélekedés, miszerint egy jelenség korpuszbeli hiánya nem értékelhető negatív evidenciaként. Bizonyos esetekben statisztikailag biztosra vehető egy adott jelenség nem-létezése (Stefanowitsch 2006). Természetesen egy korpusz mindig véges, és nem képes az elvben interpretálható megnyilatkozások sokaságát tükrözni, csak a valóban megjelenő, jellemző nyelvi formákról tud képet adni. Ez a kép azonban a korpuszméret növelésével egyre pontosabb a viszonylag ritkább jelenségek tekintetében is.

A korszerű korpuszokból nyert nagy mennyiségű adatot a legtöbb esetben lehetetlen manuálisan kezelni, „átolvasni”. Olyan kutatóeszközre van szükség, amely valamiképpen összegzi a korpu-

---

<sup>1</sup> Elhangzott az *Elméleti nyelvészeti fiesztán*, 2006. november 29-én az MTA Nyelvtudományi Intézetében.

<sup>2</sup> Elhangzott 2006 őszén az egyik *Matematikai nyelvészet* órán az MTA Nyelvtudományi Intézetében.

szokból leszűrhető tényeket, és ezt olyan formában adja a kutatók elé, hogy hatékonyan felhasználhassák adott nyelvészeti kérdések vizsgálatához, elméleti megfontolások alátámasztására, illetve cáfolatára. Az egyik első ilyen eszköz az ún. *Sketch Engine* (Kilgarriff et al. 2004). Ez a rendszer egy tömör táblázatban foglalja össze a lekérdezett szó statisztikailag lényeges kollokációit, grammatikai viszony szerint csoportosítva. Jelen dolgozatban az ehhez hasonló *Mazsola* fantázianevelő kutatóeszközt mutatom be, és felvetek néhány elméleti nyelvészeti kutatási kérdést, ahol haszonnal alkalmazhatónak vélem.

## 2. Modell és lényegesség

A Mazsola a magyar igék bővítményszerkezetének korpusz-alapú vagy korpuszvezérelt kvantitatív vizsgálatára alkalmas számítógépes kutatóeszköz. Az általa vizsgálható nyelvi alapegység a tagmondat, pontosabban az egy igét és a hozzá tartozó bővítményeket tartalmazó egység. Egy ilyen egységen belül (lényegében egy magyar egyszerű mondatról van szó) a szórend szabad, azaz az ige, illetve az egyes bővítményeket képviselő szavak, szerkezetek szinte bármilyen sorrendben elhelyezkedhetnek. Egy tagmondatot tehát felfoghatunk egy halmaznak, amiben egy ige és valahány bővítmény van. A rendszer kizárólag a névszói szerkezetként megjelenő bővítményeket tekinti, ezeket reprezentálhatjuk a szerkezet fejét adó szótóvel és a fej esetragjával, illetve névutójával. (Az esetragokat és a névutókat teljesen egyenrangúaknak tekintjük.) A fentieknek megfelelően tehát a magyar egyszerű mondat, illetve egy bővítménykeret reprezentációja a következő lehet:

ige + bővítmények(fej+eset)

Ebben a modellben egy bővítménykeret egy „pozícióján” nem fizikai, sorrendi pozíciót értünk, hanem morfoszintaktikailag-funkcionálisan meghatározott, például alanyi, tárgyi, datívuszi, *-ban/-ben-i* stb. pozíciót. „Szabad” a pozíció, ha a hozzá tartozó névszói fej széles szóosztályból szabadon választható, és „kötött”, ha a pozíción előforduló névszói fej meg van határozva. Egy egyszerű példán bemutatva:

(1) A lány vállat vont.

Az (1) példamondatban az ige mellett két pozíció van, egy alanyi és egy tárgyi, a mondat a *vki vállat von* bővítménykeret egy megvalósulása, melyben a tárgyi pozíció kötött, az alanyi pozíció pedig szabad. Ez a modell egy sokkal általánosabb modell egy esetének tekinthető, ahol az entitásokat olyan speciális jegyekkel írjuk le, melyek az egyes entitások esetén vagy megjelennek, vagy nem, és ha megjelennek, akkor többféle értékük lehet.

Ige „bővítményszerkezetén” az igének azt a tulajdonságát értem, hogy milyen bővítményekkel, névszói szerkezetekkel szokott általában együtt előfordulni. Ez implicit módon tartalmazza a korpusznyelvészet gyakorisági szempontját is: a gyakoribb bővítménykeret a fontosabb. A *von* ige (egyszerű tárgyas kereten kívüli) négy legfontosabb bővítménykerete a következő (zárójelben a Mazsola által szolgáltatott előfordulási számok szerepelnek): *kétségbe von vmit* (2435 db), *felelősségre von vkit* (1267 db), *maga után von vmit* (1193 db), *vállat von* (931 db). A *kétségbe von vmit*, *kockán forog vmi*, *górcső alá vesz vmit* típusú szerkezeteket „összetett igének” nevezem. Ezek általában véve azok a szerkezetek, amikor egy vagy több névszói szerkezet szervesen hozzátartozik az ige jelentéséhez, illetve ahol az ige jelentése kiüresedett (pl.: *moziba megy*, *egyetemre jár*); ahol a vonzat jelentése kiüresedett (pl.: *útnak indul*, *ott marad*); valamint azok is, ahol az inkorporáció jelenségével találkozunk (pl.: *fogat mos*, *fára mászik*) (Kálmán 2006). Az összetett igék önálló „igességét” nem csak az mutatja, hogy az alapigétől eltérő jelentéssel bírnak, hanem az is, hogy az alapigétől független új vonzatkerettel rendelkezhetnek: a *részt vesz* mellett megjelenő *-ban/-ben* vonzat vagy a *hírt ad* melletti *-ról/-ről* az alapige (*vesz* illetve *ad*) mellett nem szerepelt.

A modellből következik, hogy a számítógépes előfeldolgozás során milyen lépéseket volt szükséges megtenni, hogy a nyers szövegből egy ilyen módon hozzáférhető adatbázis álljon elő. A 187 millió szónyi magyar szöveget tartalmazó Magyar Nemzeti Szövegtár (MNSZ, Váradi 2002) teljes anyagát először heurisztikus eljárással egy bővítménykeretet tartalmazó egységekre bontottam.

Részleges szintaktikai elemzéssel azonosítottam a kerethordozó igét és a mellette álló névszói szerkezeteket. Az elemzőnek része az igekötő- és igető-azonosítás, az elváló igekötőket az igetőhöz kapcsolom, a gyakori *-hat/-het* képzőt levágtam, mivel nem befolyásolja a vonzatkeretet. Főnévi igenév jelenléte esetén a főnévi igenév tövét tekintetem a mondat igéjének. A határozószókat figyelmen kívül hagytam, csakis a névszói szerkezet által képviselt bővítményeket vettem tekintetbe. A rendszer tartalmazza azt az egyszerűsítést, hogy a bővítménykeretek minden pozíciójában csak egy darab névszói szerkezetet enged meg, ha egy mondatban több azonos esetű névszói szerkezet is szerepelt, akkor azok közül csak a legutolsót veszi tekintetbe.

A bővítmények két alapvető osztályának, a vonzatoknak és szabad határozóknak az automatikus elkülönítésére, azaz a valódi vonzatkeretek megragadására nincs megbízható automatikus eszköz. A lényeges kollokációkat ezzel szemben viszonylag egyszerűen meg tudjuk fogni automatikusan: mégpedig a lényegesség, jellegzetesség (bevett angol terminológiával ún. *saliency*) mérték segítségével (Kilgarriff 2001). Ezt a két elem együttes előfordulásának vizsgálatára kidolgozott mértéket alkalmazzuk itt azzal a lényegi trükkel, hogy az egyik elem a vizsgálandó bővítménykeret egyik (kiválasztott) pozíciója lesz, a másik elem pedig az ige az esetlegesen mellette megkövetelt egyéb bővítményekkel együtt. Így valójában az adott bővítménynek az igei keret többi részéhez viszonyított lényegességét tudjuk mérni. A tipikus kutatási kérdés tehát, amit vizsgálni tudunk: Melyek a lényeges szótövek, melyek egy adott keret egy adott pozícióját betölthetik? A kérdésben megadhatunk egy igetővet és valahány névszói csoportot (függetlenül attól a névszói csoportnak adott esetben mi a szerepe), és megnézhetjük, hogy egy további pozíción milyen szavak jelennek meg. Kiderül például, hogy a *hány vmire* szerkezet legtöbbször (644-ből 288 esetben) *szemére hány vkinek vmit* vonzatkeretként jelenik meg, vagy hogy a *megköszörül* igének szinte kizárólag (147-ből 134 esetben) a *torkát* lehet a tárgya, az ige szó szerinti jelentése (a várt *kést, ollót* stb. tárggyal) kivételesen ritka. Vizsgálhatóvá válik az összetett igék önálló bővítményszerkezete, valamint összevethető egy alapige és egy összetett ige bővítményszerkezete is.



jelenlétét, kizárhatunk bizonyos szótöveket. A sor végén az 'Eloszlás' alatt lehet megjelölni, hogy melyik az a pozíció, amelyet vizsgálni akarunk, azaz hogy melyik bővítményi pozícióban megjelenő jellegzetes szótövek listáját kérjük. Az eloszlás vonatkozhat az igére is, azonban tudni kell, hogy az ilyen „ige nélküli” lekérdezések futási ideje néhány perc is lehet. A legalsó sorban, egy szintén tagadható mezőben szabadszavas kereséssel szűkíthető a vizsgálat. Itt szóközzel elválasztva több szót is megadhatunk. A honlapon található példákra kattintva azonnal világossá válik az egyes beviteli mezők szerepe.

Az 1. ábrán példaként látható lekérdezésben a következő kérdésre keressük a választ: „*A kér vmit vkitől* keretben mik a jellemző szavak a tárgyi pozíción?”, más szóval „Mi mindent szoktunk általában kérni?” A válaszban a lekérdezési űrlap alatt a kívánt pozíciót betöltő szótövek listáját látjuk (esetünkben: *elnézés, bocsánat, segítség* stb.), a jellegzetesebb szó nagyobb betűmérettel jelenik meg, szögletes zárójelben tájékoztatásképpen az előfordulási szám szerepel. A szavakra kattintva, a hozzájuk tartozó releváns korpuszpéldákhoz jutunk. Az MNSZ-ben meglévő morfológiai elemzésnek és az előfeldolgozásnak köszönhetően természetesen a lekérdezések során megadott igének az összes alakját megkapjuk, az elváló igekötő is a helyére kerül. A speciális 'NULL' szótó két dolgot jelöl: egyrészt a tárgyas ragozású, de konkrét tárgyi bővítményt nem tartalmazó mondatok tárgyát, másrészt a névszói állítmányú mondatokban a jelen idő harmadik személyben elmaradó létigét.

Ha alaposabban megnézzük az egyes korpuszpéldákat, több helyen hibákat fedezhetünk fel. Egyetlen automatikus számítógépes nyelvészeti program sem tud tökéletes eredményt szolgáltatni, így tartalmaz hibákat az MNSZ morfoszintaktikai elemzése, a tagmondatrabontás és a részleges szintaktikai elemzés is. Bár emiatt az egyes mondatról sok esetben hibás specifikus megállapítást tesz a rendszer, ettől még igaz az, hogy a bővítmények lényegességéről szóló általános állítások megfogalmazásához biztos alapot ad. A statisztikai alapú általános állítások igazságára az alkalmazott eljárásban előforduló ritka hibák nincsenek számottevő hatással (Teubert 2005, Kilgarriff et al. 2004).

#### **4. Gyakorlati alkalmazás**

A fent ismertetett eszköz a gyakorlati munka során több területen is hasznosnak bizonyult. Jelen dolgozat elsődleges célja, hogy felhívja az eszközre az elméleti nyelvészek figyelmét, mégis fontosnak tartom, hogy a gyakorlati alkalmazásokat is röviden bemutassam. A magyar igei WordNet adatbázis (Kuti et al. 2007) építése során két szempontból volt hasznos a Mazsola. Egyrészt az igeik egyes jelentéseinek elkülönítésében segítenek a bővítménykeretek, ugyanis a különböző bővítménykeretek sok esetben az ige különböző jelentéseivel járnak együtt; másrészt a WordNet-be összetett igeik is belekerültek, ezeket az eszköz közvetlenül szolgáltatja. A <http://www.webforditas.hu> címen elérhető magyar-angol gépi fordítórendszer lexikai adatbázisának építésekor fontos kérdés volt, hogy melyek azok a szótövek, melyek adott ige melletti pozícióban megjelenve az igei szerkezet speciális fordítását követelik meg. Ezek legtöbbször éppen a Mazsola által megadott lényeges szavak. A Mazsola alkalmas lehet bármilyen magyar nyelvet érintő lexikográfiai munka támogatására, például egy igei bővítménykeretgyakorisági szótár összeállítására; valamint a magyar mint idegen nyelv oktatása során, a különböző gyakori igei szerkezetek megismertetésekor is hasznos segédeszköz lehet.

#### **5. Elméleti nyelvészeti relevancia**

Az említett gyakorlati szempontok mellett helye van egy ilyen eszköznek az elméleti nyelvészeti kutatásban is. Nyilvánvaló, hogy a szigorú igen/nem grammatikalitási döntések sok esetben vita tárgyát képezik, az anyanyelvi beszélők véleménye itt sokszor nagy mértékben eltér. Sampson (2007) alapján elhibázottnak tartja a grammatikus vs. nem grammatikus elkülönítést, a nyelvet, a megnyilatkozások rendszerét egy nyílt mezőn kialakuló úthálózathoz hasonlítja, szerinte a nyelvészetnek lényegében azt kellene leírnia, hogy melyik „útvonal” mennyire szokásos, azaz gyakorisági állításokat kellene megfogalmaznia. Felhívja a figyelmet (az általa Dunster-konstrukciónak nevezett) furcsa szerkezetekre, amik bár szabálytalanok, vagy érthetetlenek is tűnnek, de nem mondhatjuk ki róluk, hogy nem grammatikusak, ha valaki használta, és beszélőpartnere pedig megértette őket.



Ezzel egybecsengő felvetés szerint a nyelvtan feladata nem kizárólag az, hogy a grammatikus és a nem grammatikus mondatokat elkülönítse egymástól, hanem arra kell magyarázatot találnia, hogy bizonyos – adott esetben nem grammatikus – megnyilatkozások miért jelennek meg, és bizonyos adott esetben grammatikusak miért nem (Stefanowitsch 2006). Eszerint tehát nem a grammatikalitás, hanem a megjelenés, illetve meg-nem-jelenés az elsődleges. Pontosán ennek – a megjelenésnek és a meg-nem-jelenésnek – a kvantitatív vizsgálatára alkalmas a Mazsola eszköz.

## 6. Kutatási irányok – felhasználási lehetőségek

A dolgozat záró fejezetében néhány olyan lehetőséget, kutatási kérdést vetek fel, aminek a további vizsgálata elméleti relevanciával bírhat, és amelyekben haszonnal alkalmazható a Mazsola eszköz.

### 6.1. Stílusrétegek összehasonlítása

Az MNSZ közzétett részkorpuszai lehetőséget adnak a különféle stílusrétegű szövegek bővítményszerkezetének összehasonlítására. Alább a Magyar Nemzet és az Index fórum korpuszból nyert, *ad vmit* keretre vonatkozó adatokat elemzem. A tárgyi pozícióban közös szavak: *hang, lehetőség, válasz*, ezek azok a szavak, melyek a stílusrétegtől függetlenül lényeges szerkezeteket adják. A tárgyi pozícióban különböző szavak: *otthon, hír*; illetve *igaz, tipp*. Valóban, az *otthont ad vminek* és a *hírt ad vmiről* satjónyelvbe illő keretek, szemben az *igazat ad* és a *tippet ad* kollokvialis, hétköznapi, beszélt nyelvi jellegével. Hasonlóan felfedezhetjük a stílusrétegekre karakterisztikus kereteket a *hoz vmit (határozatot hoz vs. hozza a formáját)* vagy a *jut vmire (egyezésre jut vs. hitre jut)* esetében.

### 6.2. Szólások variációi

A szólások számtalan különféle formában megjelenhetnek. Sok szólást megragadhatunk egy igei bővítménykerettel (pl.: *Egy követ fűjnek. Szöveget ütött vkinek a fejébe.*) A korpuszból képet kapunk arról, hogy az adott szólás mennyire variábilis, illetve, hogy mely alakjai milyen arányban szerepelnek. A *Nincs új a nap alatt*. szólás például az összes 100 előfordulásából 70-szer fordul elő az idézett formában (ezen belül 8-szor a *sincs* tagadószóval, illetve néha

kiegészítő bővítményekkel: *számukra* vagy *a két ország kapcsolatában*). Tíz esetben egy közbevetett szóval jelenik meg (*Nincs semmi új a nap alatt.*), a további húsz előfordulásban pedig sokszor ellentétes értelemben (*Van még új a nap alatt!, Csak új van a nap alatt.*), különféle szórendi variációkban, valamint jelzővel kiegészítve találkozunk vele (*Mennyi minden új van a színházi nap alatt.*).

### 6.3. Igei szinonimák vizsgálata

Nagyon kevés olyan szinonim pár van, amelyek valóban ugyanazt jelentik. Egyes nyelvészek szerint lényegében nincsenek szinonimák, ha két szó jelentése nagyon közeli, akkor a nyelv úgy változik, hogy mindegyik megtalálja a saját önálló használati területét (Tognini-Bonelli, 2001:34). A *kever* vs. *kavar* pár tárgyi pozíciójának összetevésekor például markáns különbséget tapasztalunk. A *kever* leginkább az alapjelentésében szerepel, tárgya sokféle szó lehet, jellemzőek az élelmiszerek nevei (*tojássárgája, cukor, vaj, koktél* stb.). A *kavar* mellett tárgyként előforduló néhány jellegzetes szó két szemantikai csoportba osztható: dinamikus természeti jelenségek (*vihar, porfelhő* stb.), illetve konfliktust kifejező szavak (*botrány, vita, indulat* stb.) jelennek meg itt.

### 6.4. Bővítményi pozíció - szemantikai osztály

Bizonyos pozíciókban szinte kizárólag valamilyen értelemben hasonló jelentésű szavak jelenhetnek meg, más szóval olyan szavak, melyek közös jelentéssel rendelkeznek, melyek egy szemantikai osztály elemei. Például: *fest vmire* – színek; *veszélyben forog vmi* – pozitív konnotációjú dolgok (*élet, győzelem, lét, továbbjutás, épség, érdek* stb.); *költséget becsül vmire* – pénznekem; *fogad vmivel vkit* – érzések, lelkiállapotok (*öröm, fenntartás, mgdöbbenés, ováció, lelkesedés* stb.).

### 6.5. *óta* vs. *múlva*

Egyszerűen az által, hogy megvizsgáljuk, hogy milyen speciális névutókkal szokott állni egy adott ige, következtetést lehet levonni arról, hogy milyen fajta eseményszerűséget képvisel. Azok az igék amelyek jellemzően *múlva* névutós bővítménnyel fordulnak elő és *óta* névutós bővítménnyel nem, általában pontszerű eseményt

jelentenek: *indul, kezdődik, kiderül, megérkezik, meghal, megjelenik, találkozik, visszajön, visszatér.*

## 6.6. Igei konstrukciós idiómák

Jackendoff (2002:172) említi az igei szerkezeteknek egy speciális fajtáját (az ún. konstrukciós idiómákat), melyekben az az érdekes, hogy az igei pozíció szabad. Másként fogalmazva nem az ige határozza meg a bővítménykeretet, hanem a konstrukció, a konkrét ige pedig széles igeosztályból szabadon választható. A Jackendoff által említett *V <időtartam> away* (eltölti az időtartamot az adott tevékenységgel) és *V one's head off* (túlzásba viszi az adott tevékenységet) konstrukciókhoz hasonlóak az alábbi magyar példák.

*Az átmulatta az éjszakát* szerkezetben az igeő viszonylag szabadon változtatható, kötött elem az *át* igealkötő és tárgyi pozícióban egy időtartam. A Mazsolában közvetlenül nem tudunk igealkötőre keresni (ennek megvalósítása a tervek között szerepel), ún. reguláris kifejezéseket viszont használhatunk az 'Igeő' mezőben, így az *át* igealkötős igeek osztályát az *át*.\* megadásával közelíthetjük. Ha a tárgyi pozícióban időtartamot jelentő szót (*perc, hét, nap, pillanat, tél* stb.) adunk meg, és az 'Eloszlás'-t igeő szerint kérjük, az eredményül kapott igeek között megtaláljuk például a következőket: *átvöröszt, átalszik.*

A *kibeszéli magát* konstrukcióban az igeő szintén viszonylag szabadon változtatható, a konstrukció az ún. szaturatív akcióminőséget fejezi ki, hogy ti. hosszú ideig csinál valamit, és végül megelégedvén befejezi. A fentihez hasonló módon keresve (igeő: *ki*.\*, tárgy: *maga*, eloszlás: *igeő*) a gyakoribb tárgyas igeek mellett (*kifejez, kivon* stb.) a konstrukció számos példáját találjuk meg: *kípiheni, kibeszéli, kidühöngi, kibeszélgeti, kipanaszkodja, kibőgi, kidolgozza, kiveszekezi, kigyönyörködí, kiforogja, kidumálja, kicsodálkozza, kiörüli magát.* Ezek az igeek igealkötős formában *magát* nélkül a legtöbb esetben nem is léteznek, ez is mutatja a konstrukció elsődlegességét.

## 6.7. Vonzatos igealkötők

A *bele-* igealkötős igeek sokszor *-ba/-be* vonzattal, a *hozzá-* igealkötősek pedig *-hoz/-hez/-höz* vonzattal járnak. (Adott igealkötőjü

igékre reguláris kifejezéssel kereshetünk rá, l. 6.6 pont.) Érdekes kérdés, hogy az *össze-* igekötőnek van-e ilyenfajta vonzata. Intuición alapján azt gondolhatjuk, hogy általában *-val/-vel* vonzattal jár. Ha a Mazsola segítségével összehasonlítjuk a *-val/-vel* vonzattal járó és az azt nélkülöző *össze-* igekötős igéket, két szemantikailag is körvonalazható osztályt kapunk. A vonzatosak (*összefügg, összeköt, összehasonlít, összekever, összevet* stb.) mindig két dolog (alany és *-val/-vel*, illetve tárgy és *-val/-vel*) valamiféle szembeállítását fogalmazzák meg; a vonzat nélküliek (*összegyűlik, összeállít, összehív, összeszed, összefoglal* stb.) pedig több dolog egy egységbe vonásáról szólnak, ez a fajta *össze-* sokszor helyettesíthető az *egybe-* igekötővel. Úgy tűnik, hogy a szintaktikai különbség szemantikai különbséget is jelent itt, ezek alapján strukturálható lenne például az *össze-* ÉKszben szereplő nyolc jelentése.

### 6.8. Melléknevek és főnevek vonzatossága

A felületen hozzáférhető egy speciális korpusz (a nem túl szerencsés elnevezésű „MNSZ: melléknevek” korpusz), mely az MNSZ-ből csak azokat a tagmondatokat tartalmazza, melyekben az ige vagy létige, vagy jelen idő harmadik személyben elmarad (ezt 'NULL' igező jelöli). Ez a korpusz a vonzatos melléknevek (pl.: *büszke vmire, hű vmihez, egyenlő vmivel*) és főnevek (pl.: *felelősség vmiért*) vizsgálatára szolgál. A keresés úgy adható meg, hogy a keresett melléknevet vagy főnevet a szabadszavas kereséshez ('Szó' mező) írjuk, az 'Igező' mezőt üresen hagyjuk, de eszerint a mező szerint kérjük az eloszlást. Ha megvizsgáljuk például a *felelősség vmiért* vonzatos főnevet, akkor jőpár olyan igei szerkezetet kapunk, melyek az igével szemben valóban a főnév vonzatosságát támasztják alá: *felelősséget vállal vmiért, felelősség terheli vmiért, felelősséggel tartozik vmiért, felelősséget visel vmiért* stb.

### 7. Befejezés

A dolgozatban bemutattam egy korpusznyelvészeti kutatóeszközt, mellyel a magyar igék bővítményszerkezete vizsgálható. Alkalmas arra, hogy nyelvészeti állításokat empirikus adatokkal alátámasszon, illetve cáfoljon. A felület regisztráció után ingyenesen hozzáférhető a <http://corpus.nytud.hu/mazsola> címen. A Mazsola,

illetve a módszertan, melyre ez az eszköz épül, két értelemben is kiterjeszhető a jövőben. Egyrészt a magyartól különböző bármely nyelvre hasonló kutatóeszköz készíthető, ha a modellnek megfelelő formátumú korpusz rendelkezésre áll. Másrészt az igei bővítményke-  
reték mellett egyéb szerkezetek vizsgálatára is hasonló módon alkalmazható, amennyiben ezek a szerkezetek reprezentálhatók úgy, hogy egy nagyobb egységen (esetünkben a tagmondatban) belül bizonyos pozíciókat töltenek ki kisebb elemek (esetünkben a névszói csoportok).

## HIVATKOZÁSOK

- Firth, John Rupert 1957: A synopsis of linguistic theory 1930-1955, *Studies in linguistic analysis*, The Philological Society, Oxford, 1–32.
- Jackendoff, Ray 2002: *Foundations of Language: Brain, Meaning, Grammar, Evolution*, Oxford University Press.
- Kálmán László 2006: Miért nem vonzanak a régensek? in Kálmán L. szerk.: *KB 120: A titkos kötet*, MTA Nyelvtudományi Intézet– Tinta Könyvkiadó, Budapest, 229–246.
- Kilgarriff, Adam–David Tugwell 2001: Word Sketch: Extraction and display of significant collocations for lexicography, *Proceedings of the 39th Meeting of the Association for Computational Linguistics, workshop on COLLOCATION: Computational Extraction, Analysis and Exploitation*, Toulouse, 32–38.
- Kilgarriff, Adam–Pavel Rychly–Pavel Smrz–David Tugwell 2004: The Sketch Engine, *Proceedings of EURALEX, Lorient, France*, 105–116.
- Kuti Judit–Varasdi Károly–Gyarmati Ágnes–Vajda Péter 2007: Hungarian WordNet and representation of verbal event structure, *Acta Cybernetica* 18/2, 315–328.
- Riedl Frigyes 1882: Simonyi kis nyelvtana, *Egyetemes Philológiai Közöny* 573–590.
- Sampson, Geoffrey R. 2007: Grammar without grammaticality, *Corpus Linguistics and Linguistic Theory* 3/1, 1–32.

- Stefanowitsch, Anatol 2006: Negative evidence and the raw frequency fallacy, *Corpus Linguistics and Linguistic Theory* 2/1, 61–77.
- Teubert, Wolfgang 2005: My version of corpus linguistics, *International Journal of Corpus Linguistics* 10/1, 1–13.
- Tognini-Bonelli, Elena 2001: *Corpus Linguistics at Work*, John Benjamins.
- Váradi Tamás 2002: The Hungarian National Corpus, *Proceedings of LREC 2002, Las Palmas, Spain*, 385–389.