

A szótárról

1. Mi ez?

A szótár, amit az olvasó a kezében tart, a leggyakoribb magyar igei szerkezeteket tartalmazza. Egynyelvű szótár explicit szótári értelmezések nélkül; a szerkezeteket, azok jelentését autentikus, korpuszból származó példák illusztrálják. Egyrészt *vonzatkeret*-szótár és *kollokációs* szótár egyszerre: az igék legjellemzőbb vonzatkereteit és legjellemzőbb névszói szókapcsolatait is tartalmazza, illetve azokat a szerkezeteket is, melyekben e két aspektus kombinációja jelenik meg. Másrészt *gyakorisági* szótár: kvantitatív információt szolgáltat a szerkezetek gyakorisági viszonyairól. Harmadrészt *összehasonlító* szótár: lehetőséget ad a magyar igei szerkezetek többszempontú összevetésére, a közöttük lévő kapcsolatok feltárására.

Igei szerkezet alatt az egy központi igéből és a hozzá tartozó (nulla vagy több) névszói csoport bővítmenyből álló szerkezeteket értjük. A bővítmenyek két típusát különböztetjük meg. A *lexikálisan szabad bővítmenyek* (LSzB) esetén a bővítmenyként megjelenő szó számos lehetőség közül választható (pl.: ‘*néz vmit*’ – ‘*képet*’, ‘*adást*’, ‘*lányt*’, ‘*mennyezetet*’, ‘*vizet*’, ‘ *eget*’ stb.); a *lexikálisan kötött bővítmenyek* (LKB) tartalmaz szava viszont konkrét, kötött szó (pl.: ‘*vállat von*’ vagy ‘*pontot tesz a végére*’). Az első esetben a bővítmenyt az esetragja (vagy névutója) képviseli, a második esetben ezen felül még az adott esetraggal (névutóval) szereplő konkrét szó is. (Szótárunkban az esetragokat és a névutókat egységesen, egy kategóriaként kezeljük, mivel a magyarban szerepük azonos: névutók ugyanúgy képviselhetik egy ige vonzatát, mint az esetragok (pl.: ‘*vki elé tár vmit*’)). Sok szerkezetben – ezek az ún. *komplex igék* – mindkét bővítmenytípus jelen van, a ‘*kilátásba helyez vmit*’ esetén például a ‘*vmit*’ LSzB, a ‘*kilátásba*’ pedig LKB. A komplex igéket valóban tekinthetjük (több szóból álló) önálló igéknek. Egyrészt az alapigétől eltérő új jelentéssel bírnak; másrészt sok esetben az alapigétől független új vonzatkerettel rendelkeznek, ahogy ezt a ‘*részt vesz*’ mellett megjelenő ‘*-ban/-ben*’, vagy a ‘*hírt ad*’ melletti ‘*-ról/-ről*’ vonzat példázza; harmadrészt pedig sokszor egybe is írjuk őket, ilyenkor az LKB igekötőként viselkedik.

Szótárunk alapegységei nem szavak, hanem az említett igei szerkezetek. A szótári gyűjtés a lehetséges szerkezet típusok teljes spektrumát átfogja, az igei vonzatkeretektől a kollokációkon, intézményesült, idiomatikus kifejezéseken, komplex igéken át a szólásokig. A lexicográfiai hagyománnyal szemben az összes típussal egyetlen szótárban, egységes keretben kívánunk foglalkozni: ez lehetőséget ad az átmeneti esetek és a kombinációk bemutatására is. A szótárban minden kellően gyakori, jellemző szerkezet helyet kap. Amennyiben egy ige tipikus szerkezetei között csak vonzatkeret (pl.: ‘*bízik vmiben*’) illetve csak szókapcsolat

lat (pl.: *'csóválja a fejét'*) található, akkor ezeket közöljük. A bővítményeket egységesen kezeljük, a szótárba minden jellegzetes szerkezet bekerül, függetlenül attól, hogy a benne szereplő bővítmény vonzat (pl.: *'hisz vmiben'*) vagy szabad határozó (pl.: *'történik vmiben'*). Nem csak idiomatikus, hanem tipikus kompozicionális szerkezeteket is közlünk. Említettük, hogy ha egy bővítményi helyzetben sokféle szó szerepelhet, akkor az LSzB-ként jelenik meg (pl.: *'végez vmit'*), ugyanakkor ha ezek közül a szavak közül bizonyosak gyakoriak, jellegzetesek, akkor ezek LKB-ként, önálló szerkezet formájában *külön is* megjelennek (pl.: *'munkát végez'*). Mivel megnyilatkozásaink túlnyomó része igék köré épül, az igei szerkezetek révén a magyar nyelv egészéről kapunk átfogó képet.

Önállóság, összevethetőség. Minden szerkezet „azonos jogon”, önállóan szerepel a szótárban, így az egyes szerkezetek különböző tulajdonságai – típusuk, bonyolultságuk, vonzat- ill. bővítményszerkezetük vagy gyakoriságuk – közvetlenül összehasonlíthatókká válnak. Szerkezetileg és gyakoriság szempontjából eltér például a *'titkol vmit'* (egyszerű ige, gyakoribb; a 169. oldalon) és a *'véka alá rejt vmit'* (komplex ige, szólás, ritkább; a 151. oldalon) szerkezet, jelentésük viszont hasonló, és tárgyat megkövetelő vonzatkeretük is azonos.

A gyakorisági elv. Az anyaggyűjtés – a magyar lexikográfiában újszerű módon – *automatikusan* történt, vezérlő elve a *korpuszgyakoriság*. Az igei szerkezeteket és gyakorisági adataikat reprezentatív szövegkorpuszból, a 187 millió szavas Magyar Nemzeti Szövegtárból nyertük. A szótáraknak nyilván a tipikust, a jellegzetest kell bemutatnia. Jelen esetben szigorúan az objektív gyakorisági adatokból indultunk ki: azt, hogy mely szerkezet került be a szótárba kizárólag a szerkezet korpuszbeli gyakorisága határozta meg. A szótárba intuitív alapon nem vettünk fel szerkezeteket, csak azért mert „odaillestek” volna, vagy mert a vélt „alapjelentést” képviselik. A szerkezetgyűjtő algoritmus (ld. a 15. oldalon a szótárkészítés lépéseinél) segítségével a szerkezetekhez közvetlenül rendeltünk *gyakorisági mérőszámot*. Az erre alkalmazott egységes, 250-es gyakorisági küszöb fölött az összes igei szerkezet bekerült, az ennél ritkébbak pedig kimaradtak. A szerkezetek szintjén érvényesül tehát a gyakorisági elv, azaz a szótár nem a leggyakoribb igék összes szerkezetét, hanem a korpuszban meglévő összes ige leggyakoribb szerkezeit tartalmazza. Ebből következően adott igenél, melynek egyéb szerkezetei ritkák, esetleg csak egyetlen összetett szerkezet szerepel, mint ahogy ezt a *'bevált' – 'beváltja a (hozzá fűzött) reményeket'* (31. oldal), a *'fér' – '(nem) fér kétség vmihez'* (64. oldal) vagy a *'varr' – 'nyakába varr vmit'* (188. oldal) esetén láthatjuk is. A szótárban 6266 önálló igei szerkezet található, mely a 19. oldaltól kezdődő betűrendes szótári részben 2226 igei címszóhoz sorolódik be. A gyakorisági elv egy megnyilvánulása az is, hogy az igei szócikken belül a szerkezetek gyakorisági sorrendben szerepelnek: a szótárhasználó vélhetően elsősorban a gyakran előforduló szerkezetekre lesz kíváncsi.

Igekötő mint önálló elem. Szótárunkban az *igekötőket* önálló elemnek tekintjük. A klaszikus szótári gyakorlat szerint az igekötős igéket (pl.: *'összevon'*) külön egységként, külön lexémaként kezelik, a komplex igéket (pl.: *'kétségbe von'*) pedig az alapige (*'von'*) alatt tárgyalják. (Így az előbbi az *ö*, az utóbbi pedig a *v* betűhöz kerül a betűrendben.) Attól eltekintve, hogy az előbbit egybeírjuk, az utóbbit pedig külön, ezeknek a szerkezeteknek a felépítése valójában nagyon hasonló. Hasznos tehát, ha ezeket a szerkezeteket egy helyen, együtt láthatjuk. Ezért – amellett, hogy a szótári részben megtartottuk a hagyományos betűrendet – létrehoztuk az ún. *alapige szerinti* mutatót (a 421. oldaltól), mely a javasolt csoportosítást valósítja meg: az egy alapigéhez tartozó összes szerkezetet az alapigénél tünteti fel. Így szótárunkban mindkét módon megtalálhatjuk a keresett szerkezeteket.

2. Hogyan használjuk?

A szótár hagyományos betűrendes szótári részből, valamint öt különböző mutatóból áll. A mutatók nem kiegészítő funkciót látnak el, hanem szerves részét képezik a szótárnak. Minden mutató a saját szempontja szerint rendezve, csoportosítva mutatja be a teljes anyagot, lehetővé téve a szerkezetek e szempont szerinti összevetését.

Az igei szerkezetek megjelenítése. Maguk az igei szerkezetek a teljes szótárban egységes módon jelennek meg. A szerkezetben mindig elsőként az igét (igetövet) tüntetjük fel: az esetleges igekötőt ']' jel választja el az alapigétől (pl.: 'ki|vesz'). Az ige után *eset/névutó szerinti betűrendben* következnek a bővítmények, legvégül az alany. Az esetragokban a hangrendi váltakozási pontot nagybetűvel jelöljük, azaz a '-ban/-ben' rag jele: '-bAn', a '-hoz/-hez/-höz' jele pedig '-hOz'. A névutókat egy eléjük írt pont jelöli: '·nélkül'. Ha LKB szerepel a szerkezetben, akkor annak a szótóvét jelenítjük meg, kiskapitális szedéssel, közvetlenül hozzákapcsolva a ragot/névutót. Az '*asztalhoz ül*' szerkezet formája tehát: 'ül ASZTAL-hOz', a '*véka alá rejt*'-é pedig: 'rejt VÉKA-alá'. Az alanyi üres esetragot külön nem jelöljük, az alany utolsóként jelenik meg a bővítmények sorában (pl.: 'kerül -rA SOR'). Fontos látni a különbséget az 'áll ·mögött ASZTAL' ('*egy asztal áll vmi mögött*') és az 'áll ASZTAL·mögött' ('*vmi áll (egy) asztal mögött*') szerkezet között. Azokat a szerkezeteket is önálló szerkezetként kezeljük, melyek csak a birtokos személyrag meglétében térnek el, a birtokos személyrag jele: '-A'. A '*felhívja vmire (vkinek) a figyelmét*' szerkezet megjelenése ezek alapján a következő: 'fel|hív -rA FIGYELEM-A-t'. A továbbiakban az igei szerkezeteket – illetve részeit – a most ismertetett formában, az egyéb nyelvi példákat továbbra is aposztrófok között kurzíván szedve közöljük.

1. Szótári rész: a szócikk felépítése. A 19. oldaltól kezdődő klasszikus betűrendes szótári részben az egyes igék a címszavak, a szerkezetek az igék alá sorolva szerepelnek. A címszó után következik kerek zárójelben az adott ige előfordulási száma a korpuszban. A címszó alatt sorakoznak az adott igéhez tartozó igei szerkezetek a gyakorisági mérőszám szerinti sorrendben. A szerkezet után következik szögletes zárójelben a hozzá tartozó gyakorisági mérőszám, majd pedig egy példamondat. A gyakoriság szerinti felsorolás egy helyen bomlik meg: ez a szerkezetek közötti tájékozódást hivatott elősegíteni. Behúzással jelölve, a felső szintű általános szerkezetek „alá” vannak sorolva azok a specifikusabb szerkezetek, melyeknek a gyakorisága kisebb mint az általános szerkezeté, és egy bővítményi pozícióban konkrét szóval és/vagy konkrét alannyal bővebbek. Így a komplex igék – gyakoriságuk révén – általában a felső szinten maradnak (pl.: 'fel|tesz KÉRDÉS-t', 'helyez KILÁTÁS-bA -t', 'játszik -bAn SZEREP-t', 'jön LÉT-rA'); azok a szerkezetek pedig, melyekben a jellegzetes kötött szó nem jár külön speciális jelentéssel, általában az alsó szintre sorolódnak (pl.: 'fel|emel KÉZ-A-t', 'fizet DÍJ-t', 'iszik SÖR-t'). Ez a besorolás nem befolyásolja azt, hogy a szerkezetek a gyakorisági mérőszámaikkal együtt *önállóak*: minden szerkezetnek a saját jogán van hozzárendelt gyakorisági mérőszáma, és ez az esetleges általánosabb szerkezetek gyakorisági mérőszámával *nem fed át*, abba nem számít bele. A szótári rész segítségével összevethetjük az egy igéhez tartozó bővítménykeretek gyakorisági viszonyait. Látjuk, hogy az 'óv' (145. oldal) és a 'tanul' (163. oldal) esetén is '-t', '-t -tÓl', '-tÓl' gyakorisági sorrendben szerepel ez a három keret. Ez utalhat arra, hogy ezen igék mellett tárgyi és egy opcionális '-től/-től' -ragos vonzat szerepel, és nem elhanyagolható gyakoriságúak azok a mondatok, ahol a tárgyat (eliptikusan) elhagyjuk.

2. Gyakoriság szerinti mutató. A szótári részt követő gyakoriság szerinti mutatóban (a 197. oldaltól) a gyakorisági mérőszám szerint rendezve listázzuk a szerkezeteket. Láthatjuk például, hogy a leggyakoribb LKB-t tartalmazó szerkezet a ‘van -rÓI SZÓ’, vagy hogy eltérő bonyolultságuk ellenére nagyjából azonos gyakoriságúak a ‘tart FONTOS-nAk -t’, ‘be|számol -rÓI’, ‘tervez -t’ és ‘él’ szerkezetek (199. oldal). Ez a lista akkor hasznos, ha egy adott szerkezettel nagyjából azonos gyakoriságú (vagy ritkább, gyakoribb stb.) egyéb szerkezeteket keresünk.

3. Keretek szerinti mutató. *Keret* vagy *bővítménykeret* alatt az igei szerkezetek igétől különböző részét, azaz a bővítmények összességét értjük. A keretek szerinti mutatóban (a 267. oldaltól) a szerkezeteket az igétől függetlenül a keret szerint rendezve látjuk (a rendezés leírását ld. alább). Azonos bővítménykereten belül ige szerint rendezve sorakoznak a szerkezetek. Segítségével azt vizsgálhatjuk, hogy milyen különféle igék társulnak egy adott kerettel (pl.: ‘-bÓI -t’ (296. oldal) vagy ‘MAGA-bAn -t’ (294. oldal)).

4. A kötött szavak szerinti mutatóban (a 355. oldaltól) az LKB-ként megjelenő kötött szavak szerint csoportosítva látjuk a szerkezeteket. Adott kötött szón belül gyakoriság szerinti a sorrend, így a legjellemzőbb szerkezetet látjuk legelől (pl.: BÍRÓSÁG (357. oldal), IDŐ (367. oldal)). A gyakoriságot itt fel is tüntettük a szerkezetek mellett. A több LKB-t tartalmazó szerkezet (pl.: ‘hajt MALOM-A-rA VÍZ-t’) többször, az összes benne szereplő szónál külön-külön szerepel. Szótárunk az igék, igei szerkezetek feltérképezésére vállalkozik, ennek a mutatónak a segítségével viszont éppen az LKB-ként megjelenő *névszók* viselkedéséről kaphatunk (vázlatos) képet, köszönhetően annak a döntésünknek, hogy minden tipikus szerkezetet szerepeltettünk a szótárban, idiomatikusakat és kompozicionálisakat egyaránt. E mutató segítségével a névszó felől közelíthetjük meg az igei szerkezeteket. Kiderül például, hogy a ‘*szereződés*’ szóval legjellemzőbben együtt járó ige a ‘*köt*’, az ‘*alátír*’ és a ‘*megköt*’ (381. oldal).

5. Igekötős keretek szerinti mutató. Az igekötők sok esetben önálló elemként, az igétől függetlenül kapcsolatban állnak a bővítményekkel. Ezeket a viszonyokat ragadja meg ez a mutató a 389. oldaltól kezdődően. Abban különbözik a keretek szerinti mutatótól, hogy itt az igekötőt is hozzávesszük a bővítménykerethez. A szerkezetek igekötővel kiegészített keretek szerinti sorrendben szerepelnek. Tudjuk, hogy a magyarban bizonyos igekötők együtt járnak bizonyos ragokkal, másképp fogalmazva az igekötő és a rag együtt egy szerkezetet alkot (pl.: ‘bele -bA’ (392. oldal), ‘fel -rA’ (400. oldal), ‘ki -bÓI’ (403. oldal) stb.) Ennek a mutatónak a segítségével az ilyenfajta szerkezeteket tanulmányozhatjuk.

6. Az alapige szerinti mutató (a 421. oldaltól) a másik olyan rész, melynek segítségével az igekötők viselkedése elemezhető. A szerkezeteket itt az igekötő leválasztása után kapott ige (az ún. *alapige*) szerint rendezzük, illetve ezen belül keretek szerint. Azt vizsgálhatjuk, hogy milyen igekötőkkel jár egy alapige, és hogyan viszonyul ez a bővítményekhez (pl.: ‘át|csap -bA’, ‘le|csap -rA’, ‘be|csap -t’ (428. oldal)). Megváltoztatja-e az igekötő bővítménykeretet (pl.: ‘ad -hOz -t’ vs. ‘hozzá|ad -hOz -t’ (421. oldal))? Mely igék járnak szinte mindig igekötő nélkül (pl.: ‘*aggódik*’ (422. oldal)), illetve szinte mindig igekötővel (pl.: ‘*ki/megfejt*’ (436. oldal), ‘*be/el/lehuny*’ (446. oldal))?

A keretek szerinti rendezés, melyet több mutató esetében is rendező elvként alkalmazunk, az alábbiak szerint történik. Amint említettük, az igei szerkezeteken *belül* az eseteket betűrend szerint rendezzük, az alanyt a végére sorolva. Magukat a kereteket először az elsőként szereplő eset szerint rendezzük, ezen belül a 2. eset szerint ..., végül az alany szerint; aztán

ezen belül az 1. esetben megjelenő esetleges kötött szó szerint, ezen belül a 2. esetben megjelenő esetleges kötött szó szerint . . . , és legvégül az alanyként megjelenő esetleges kötött szó szerint. E módszer eredményeképpen egy áttekinthető, eset szerint könnyen kereshető rendszert kapunk, melyben a ‘-bŐI’ után következnek a ‘-bŐI’ mellett valamilyen kötött alanyal bíró keretek, majd a ‘-bŐI’-ragos kötött szót tartalmazó keretek; ezt követően a különféle további bővítményekkel bíró keretek: ‘-bŐI -rA’ . . . ‘-bŐI -t’, rendre utánuk az őket kötött szavakkal kiegészítő keretek, ahogy ez a 295-297. oldalon látható.

Oda-vissza. Látjuk, hogy minden mutatóban minden ott releváns szerkezet megjelenik csak más szempont szerint csoportosítva. Mindent mindenhol megtalálunk, a fent ismertetett különböző szótári részek jelentősége abban rejlik, hogy *mit mi mellett* azaz egy adott szerkezetet mely szempontból hasonló szerkezetek között találunk meg. Az LKB-t tartalmazó szerkezetek például a keretek szerinti és a kötött szavak szerinti mutatóban is megjelennek. Előbbi arra szolgál, hogy felfedezzük, hogy bizonyos kötött szavas keretek több igével is (és hogy konkrétan melyekkel) alkotnak szerkezetet (pl.: a ‘-rŐI KÉP-t’ a 321. oldalon), utóbbi pedig arra, hogy milyen különféle szerkezetekben szerepel az adott kötött névszó, hogyan használhatjuk ezt a szót. Ez az „oda-vissza” kereshetőség a szótár számos pontján tetten érhető.

Mindent mindenhol. Akkor jó egy szótár, ha a többelemű egységeket bármely elemükből kiindulva könnyen meg lehet találni benne. Ezt általában kereszthivatkozásokkal és/vagy az elemek többszöri feltüntetésével szokták megoldani. Szótárunk a szerkezeteket minden részben külön feltünteti, ezek a bejegyzések felfoghatók a szótári rész megfelelő címszavára utaló kereszthivatkozásként is. A plusz információt mindig a kikeresett szerkezet *környékén* lévő egyéb szerkezetek hordozzák. A ‘le|von -bŐI KÖVETKEZTETÉS-t’ szerkezet elemeiről, részeiről például a következők szerint tudhatunk meg további információt: a ‘le’ igekötős szerkezeteket a szótári részben (a 106. oldaltól) és az igekötős keretek szerinti mutatóban (407. oldal); a ‘von’ alapige szerkezeteit az alapige szerinti mutatóban (504. oldal); a ‘-bŐI’, ‘-t’, ‘-bŐI -t’ kereteket a keretek szerinti mutatóban (rendre: 295., 322., és 296. oldal); a ‘KÖVETKEZTETÉS’ szót tartalmazó szerkezeteket a kötött szavak szerinti mutatóban (371. oldal); a ‘le|von’ ige szerkezeteit a szótári részben (114. oldal); az egyetlen ‘le -bŐI -t’ igekötős kerettel bíró szerkezetet az igekötős keretek szerinti mutatóban (407. oldal); a hasonló (923-as) gyakoriságú szerkezeteket pedig a gyakoriság szerinti mutatóban (211. oldal) találjuk meg.

3. Kinek szól és mire jó?

Elsősorban a nyelvész szakmának szánjuk szótárunkat. Korpuszalapú *elméleti nyelvészeti* kutatásban nyelvi adatok hiteles forrásaként, illetve a szerkezetek többszempontú összevetése révén hasznosítható. Gyakorisági adatokat szolgáltat nyelvi adatokra épülő (pl. *pszicholingvisztikai*) kísérletekhez. Lexikális erőforrásként jelenhet meg a *nyelvtechnológia* számos területén az információ visszakereséstől a gépi fordításig; valamint hasznos segédeszköz lehet más *lexikográfiai* munkák készítésekor: korpuszból nyert autentikus adatokat foglal össze, manuálisan ellenőrizve és javítva, alkalmas korpuszpéldákkal kiegészítve.

A nyelvünk összefüggéseire kíváncsi, anyanyelv iránt érdeklődő nagyközönséget is megszólítjuk. Szótárunk bemutatja a legtipikusabb igei szókapcsolatokat: könnyen azonosít

hatjuk azokat a névszókat, melyek adott igével szerkezetet, komplex igét alkotnak. A 355. oldalon kezdődő kötött szavak szerinti mutató segítségével pedig a *fordított irányú* kollokációs kapcsolatokat is számba vehetjük, kideríthetjük, hogy az adott névszó mely igékkel milyen kifejezéseket alkot. Választ kapunk tehát a nyelvhasználat során felmerülő mindkét alábbi kérdésre: mi adott igéhez a szokásos bővítmény ill. mi adott bővítményhez a szokásos ige.

A fordítói munka során a különféle ige-bővítmény szókapcsolatok feltérképezésére kollokációs szótárként használható szótárunk. Egyrészt a többszavas kifejezések megfelelő fordításakor, másrészt pedig az egyes névszókhoz legjobban illeszkedő igék, illetve az egyes igékhez legjobban illeszkedő névszók megtalálásakor. A jó szótár *ötletet ad* arra, hogy ténylegesen hogyan mondják az adott dolgot az adott nyelven. Ennek megvalósításához a jelen szótár készítésekor alkalmazott korpuszvezérelt módszertan vihet legközelebb.

Magyar nyelvtanítási célú tananyagot készítő nyelvtanároknak, kutatóknak szintén hasznos lehet a szótár, magyar mint idegen nyelv oktatása illetve anyanyelvi nevelés terén egyaránt. Itt szintén nyilvánvalóan fontos szempont a gyakoriság a szókinccsfeljesztésben, a törzsszókinccs kiválasztásában. Az igei szerkezetek a nyelv teljességét átfogják, megnyilatkozásaink jelentős részének egy ige a központi eleme.

A haladó magyarul tanulók egyfajta speciális tanulói szótárat kapnak a kezükbe, mely a legkülönbözőbb típusú gyakori igei szerkezetek bemutatása révén elősegíti az „idiomatikus”, a magyar nyelvre jellemző nyelvhasználatot, a nemcsak nyelvtanilag helyes, hanem magyarul *megszokott* kifejezésmódot. Képzeljük magunkat egy magyarul tanuló angol anyanyelvű helyébe. Hogyan is mondjuk magyarul, hogy *'meet the requirements'*? Tegyük fel, hogy tudjuk, hogy a *'követelmény'* szót kell használni, de mi a hozzá illeszkedő ige? A válasz: *'megfelel'* (ld. a 371. oldalon), és az is kiderül, hogy az angol tárgyias kifejezéstől eltérően a *'követelmény'* szót *'-nak/-nek'* raggal kell használnunk. Hasonlóan találhatjuk meg a *'make a contract'* kifejezés kapcsán a *'szerződés'* szóhoz a *'köt'* igét a 381. oldalon, kiegészítve azzal az információval, hogy a szerződéskötésben szereplő másik fél *'-val/-vel'* ragos bővítményként jelenhet meg a magyarban. Az *'ajándék'* szónál megtalálható *'kap AJÁNDÉK-BA -t'* szerkezetből pedig azt tudhatjuk meg, hogy itt a *'-ba/-be'* ragot kell használnunk eltérően az angol *'as a gift'* formától.

Mazsola. A szótárt kiegészítheti az azonos szövegtörzshöz épülő speciális korpuszlekérdező eszköz, a *Mazsola*, mely elérhető a <http://corpus.nyttud.hu/mazsola> internetes címen. Segítségével a magyar igék bővítményszerkezetét vizsgálhatjuk, egyes bővítményi pozíciókban megjelenő jellegzetes szavakra tudunk rákérdezni; fontos azonban látni, hogy ez az eszköz *nem* tartalmazza a jellegzetes igei szerkezeteket összegyűjtő lépést. Ha az a kérdés, hogy mely szerkezetek tipikusak, akkor a jelen szótárhoz kell fordulnia az olvasónak.

4. Mi a jelentősége?

Lexikográfiai újdonságok. Szótárunk lexikográfiai szempontból több újdonságot tartalmaz. Alapegységei nem szavak, hanem *szó szerkezetek*; az anyaggyűjtés *automatikusan* történik, a nyers szócikkek a lexikográfus nyelvi intuíciójától függetlenül automatikusan állnak elő; autentikus *korpuszpéldák* világítják meg a szerkezetek jelentését; valamint *gyakorisági* mérőszámot is rendel a szerkezetekhez. A fentiek jól illeszkednek a modern szótárkészítési trendekbe, melyek szerint a szavak helyett egyre inkább a többszavas kifejezéseket állítjuk

a középpontba, és különféle automatikus eljárásokkal próbáljuk csökkenteni a lexikográfia „rabszolgamunka” részét, így téve gyorsabbá és olcsóbbá a szótárkészítést.

A szótárkészítés lépéseit itt csak röviden mutatjuk be, a teljes körű ismertetés külön tanulmányban olvasható (Sass Bálint és Pajzs Júlia: Igei szerkezetek gyakorisági szótára – félautomatikus szótárkészítés nyelvtechnológiai eszközök segítségével. In: *Alkalmazott Nyelvtudomány*, 2010(1-2): 5-32.). Kiinduló szövegtörzsünk a morfológiailag elemzett, 187 millió szavas Magyar Nemzeti Szövegtár. A teljes szöveget automatikusan tagmondatokra bontottuk, majd automatikus részleges szintaktikai elemzés során azonosítottuk a bővítményeket és az igéket (az elváló igekötőket az igéhez kapcsoltuk). Ezt követte a legfontosabb lépés: az igei szerkezetek automatikus gyűjtése. Az alkalmazott eljárás lényege, hogy képes automatikusan megállapítani, hogy a szerkezetek bővítményi pozíciójában megjelenő konkrét szó kellően gyakori-e ahhoz, hogy LKB-ként, „saját jogán” feltüntessük. Ha igen, akkor az adott kötött szóval kiegészített szerkezetet *teljes jogú, önálló* szerkezetként kezeli. Például bár a ‘vet SZEM-A-rA -t’ és a ‘vet -rA PILLANTÁS-t’ szerkezet egyaránt egy ‘-rA’ ragos bővítményt és egy tárgyat tartalmaz, az algoritmus automatikusan állapítja meg, hogy az elsősben a ‘-rA’-ragos bővítmény LKB és a tárgy LSzB, a másodikban pedig fordítva. Az algoritmus dönti el, fedezi fel tehát, hogy *mik* a korpuszban meglévő szerkezetek. Kimenatként a korpuszban fellelhető igei szerkezetek listája áll elő, gyakorisági mérőszámmal kiegészítve. Ebből a listából az említett egységes 250-es gyakorisági küszöbérték fölötti szerkezeteket választottuk ki, ezeken végeztük el manuálisan a lexikográfusi munkát, ami mindössze a hibás szerkezetek elhagyásából, és alkalmas példamondat kiválasztásából állt.

A gyakorisági mérőszámok az adott igei szerkezetre illeszkedő korpuszmondatok összeszámlálásából adódnak. Az algoritmus minden szerkezethez egyértelműen hozzárendeli azokat a mondatokat, melyek egy-egy találattal gyarapítják gyakorisági mérőszámát, azaz minden szerkezet mérőszáma *más-más* mondatok összeszámlálásából adódik. Ha egy mondat több szerkezetre is illeszkedik, akkor az algoritmus *dönt*, hogy az adott mondatot melyik szerkezethez számítsa. Ez azt jelenti, hogy a ‘vesz -bA -t’ gyakorisági mérőszámába például *nem* számítanak bele a ‘-bA’-ragos LKB-t tartalmazó különféle szerkezetek (‘vesz FIGYELEM-bA -t’, ‘vesz IGÉNY-bA -t’, ‘vesz ÓRIZET-bA -t’, ‘vesz KÉZ-bA -t’, ‘vesz CÉL-bA -t’ stb.). A 189-191. oldalon láthatjuk, hogy e specifikus szerkezetek gyakorisági mérőszámainak összege jelentősen meg is haladja az általános szerkezetét. A ‘vesz FIGYELEM-bA -t’ és a ‘vesz -bA -t’ gyakorisági mérőszáma *nem fed át*, az előbbi 5063 db rá illeszkedő mondat összeszámlálásából adódik, az utóbbi pedig 524 db *az előbbiektől különböző* mondat összeszámlálásából, melyekben nem a ‘figyelem’ szó szerepel ‘-ba/-be’ raggal. Természetesen ugyanígy igaz ez minden specifikusabb-általánosabb viszonyban lévő szerkezetre. Úgy is mondhatjuk, hogy minden szerkezet rekurzívan „kihasítja” a maga részét a formailag az általánosabb szerkezethez tartozó mondatokból. A fentiek miatt válik közvetlenül összehasonlíthatóvá a különböző bonyolultságú szerkezetek gyakorisága, ahogy ezt a gyakoriság szerinti mutatóról szóló részben a 12. oldalon elővételteztük. Most elemzett példánk azt mutatja, hogy a ‘vesz -bA -t’ szerkezet így, ebben az általános formában nagyon ritka, szinte „nem is létezik”, viszont különféle specifikus szerkezetek gyakoriak.

A korpuszvezéreltség azt jelenti, hogy szótárunk csakis a felhasznált korpuszban meglévő szerkezeteket tartalmazza. Nem a lexikográfus intuíciója dönti el, hogy mi kerül a szótárba, hanem kizárólag a korpuszgyakoriság alapján dolgozó igei szerkezeteket gyűjtő algoritmus. Az automatikusan előállított szerkezetlistához a lexikográfusok *nem adtak hozzá* elemeket,

és törölni is csak az egyértelműen valamely elemző lépés nem tökéletes működéséből származó hibás szerkezeteket törölték. A Magyar Nemzeti Szövegtárból származik az összes példamondat (valójában példatagmondattal). Nem mindig egy az egyben átvett, hanem szükség esetén kis mértékben módosított *korpuszalapú* példákkal dolgozunk: a kötőszót, határozószót, a szerkezet illusztrálásához nem feltétlenül szükséges elemeket sok esetben elhagytuk, illetve törekedtünk arra, hogy a bővítmények lehetőleg ne névmásként, hanem tartalmas szóként jelenjenek meg. A példákat egységesen kisbetűsítettük, a tagmondat végén lévő írásjelek közül csak a kérdőjelet és a felkiáltójelet hagytuk meg. A példák helyesírását szükség esetén javítottuk. A szótár abban az értelemben *teljes*, hogy valóban tartalmazza az összes 250-nél nagyobb gyakorisági mérőszámmal bíró korpuszbeli igei szerkezetet. Az MNSZ 2002-ben készült el, ennek megfelelően a legújabb nyelvi fejlemények nem szerepelnek benne, így ezek szótárunkban sem jelenhetnek meg.

Módszerünk korlátait e helyen foglaljuk össze. A szerkezeteknél az igét, a bővítményi eseteket és a konkrét bővítményi szavakat vesszük számba, nem vesszük tekintetbe a szerkezetek esetleges *egyéb tulajdonságait*, mint például a kötött szó jellemző jelzőjét, a jellemző igeidőt, határozószókat, vagy a ‘van’ jelen idejű, egyes szám harmadik személyű alakjának viselkedését. A gyakoribb (személyes, mutató és vonatkozó) *névmások* az előfeldolgozás során – esetragjuk megtartása mellett – elhagyjuk, mivel ezeknek LKB-ként általában nincs szerepük. A többi névmást megtartjuk, így megkapjuk azokat a szerkezeteket is, ahol ezeknek jellegzetes szerepe van, mint például a ‘von -t MAGA-után’ vagy a ‘el|választ -t EGYMÁS-TÓI’ esetében. Az igei címszóknál megjelenő *gyakorisági értékek* az elemzett korpuszból származnak. Ha az automatikus morfológiai elemző nem tudja eldönteni, hogy mi a tényleges igező (pl.: ‘ér’ vagy ‘érik’), akkor az adott mondat mindkét ige gyakoriságába beleszámít. A morfológiai elemző a ‘van’ és a ‘lesz’ igét *két különálló igeként* elemzi. Mivel vannak különböző szerkezeteik (pl.: ‘lesz -vÁ’ vagy ‘lesz -n ŰR-vÁ’), ezeket utólagosan nem vonjuk össze. A gyakorisági elvből következően egyébként a szerkezetek nem jelentéktelen része (8,8 %-a) a leggyakoribb igével, a ‘van’-nal jár. A korpuszban található névszói-igei állítmányok kapcsán a szótárban megjelennek a leggyakoribb vonzatos és vonzat nélküli *melléknevek* is. A melléknevek a ‘van’, ‘lesz’, ‘marad’ igét tartalmazó szerkezetekben az alanyhoz hasonlóan utolsó tagként jelennek meg (pl.: ‘van -bAn BIZTOS’, ‘van -rA BÜSZKE’ vagy ‘marad VÁLTÓZATLAN’). A szintaktikai elemzés során, ha egy tagmondatban adott esetragos bővítményből több szerepel, akkor csak az *utolsóként megjelenőt* vesszük figyelembe, emiatt előfordul, hogy bizonyos konkrét mondatokban helytelen vagy hiányos bővítménykeretet regisztrál a szerkezeteket gyűjtő algoritmus, de ez a szótárban megjelenő összesített adatokat lényegében nem befolyásolja. A bővítményeket *pusztán felszíni jegyek* (esetragok/névutók) alapján kapcsoljuk az igékhez, azaz nem végeztünk semmiféle szemantikai elemzést. Emiatt fordul elő, hogy a ‘lakik VHOL’ szerkezet helyett a (gyakoribb) ‘lakik -bAn’ és a (sokkal ritkább) ‘lakik -n’ jelenik meg, illetve a fordított eset, mikor egy esetrag szempontjából egységes bővítmény számos különböző jelentést fed le, pl. ‘nyer -vAl’: ‘pontozással’, ‘lelkesezséssel’, ‘kisgazdákkal’. Az igei szerkezeteket olyan mértékig különíti el szótárunk, amennyire a felszíni jegyekre épülő megközelítés ezt lehetővé teszi.

A szerkezetek önállóságáról még egyszer. Bizonyos bővítmények vonzatok (pl.: ‘hisz -bAn’) vagy komplex igék kötött elemei (pl.: ‘von VÁLL-t’); mások szabad határozók (pl.: ‘történik -bAn’) vagy önálló jelentést nem hordozó *pusztán* gyakori szavak (pl.: ‘iszik SÖR-t’). Tiszta esetben az előbbieket *idiomatikus*, az utóbbiakat pedig *kompozicionális* bővítmények; va-

lójában azonban számos átmeneti eset létezik, ilyenkor nehéz besorolni a bővítményeket e két kategória valamelyikébe. A ‘von VÁLL-t’ típusú szerkezetet önállóan, „idiómaként” szokás kezelni, az ‘iszik SÖR-t’ típusút pedig az ‘iszik -t’ *alá* szokás besorolni. Automatikus módszer híján nem vállaltuk fel annak ódiomát, hogy manuálisan, intuitív alapon döntsünk a fenti két kategória tekintetében, inkább azt a megoldást választottuk, hogy *minden szerkezetet önállóan* kezelünk. Ez abban jelentkezik, hogy minden szerkezet külön példamondattal bír, és főként abban, hogy minden szerkezetnek a saját jogán van gyakorisági mérőszáma, ahogy ezt korábban említettük. Elmondható, hogy az algoritmus főként idiomatikus szerkezeteket szolgáltat, és a többi szerkezet is elég jellegzetes ahhoz, hogy egy szótárban önállóan szerepeljen. A fő indok, ami miatt ezt a megoldást választottuk az, hogy nem tartjuk szerencsésnek, ha az önálló jelentéssel bíró idiomatikus szerkezetek *formai* alapon más szerkezetek alá sorolódnak. A ‘vesz RÉSZ-t -n’-t külön akarjuk kezelni a ‘vesz -t -n’-től, és a ‘varr NYAK-A-bA -t’ szerkezetet is a ‘varr -bA -t’-től. További érv, hogy éppen így, az elkülönítés révén derülhet ki, hogy bizonyos esetekben (ilyen a két imént idézett is), a formailag általánosabb keret jelentősen ritkább, mint a specifikusabb idiomatikus szerkezet, az előbbi szinte mesterséges csomópontként jelenik meg a szótárban.

Ez a megoldás független attól, hogy szótárunk egyes részeiben hogyan *prezentáljuk* a szerkezeteket, ti. a betűrendes részben igék köré rendezve, valamint bizonyos szerkezeteket beljebb szedéssel jelölve. Jelen munkában szótározandó alapegységnek (*lexémának*) az igei szerkezeteket tekintjük, és ettől elkülöníthetjük a szótári rész igei *szócikkeit*, melyek ilyen lexémák rendezett csoportjaiból állnak. Ezzel a terminológiával élve elmondhatjuk, hogy szótárunkban a ‘*csóvál*’ szócikkhez (36. oldal) egyetlen lexéma (a ‘csóvál FEJ-A-t’) tartozik, és ‘csóvál’ lexémát a szótár nem tartalmaz.

Önálló mozgathatóság. Az, hogy az egyes igei szerkezeteket önállóan kezeljük, azzal az előnnyel is jár, hogy a szerkezetek önállóan *mozgathatókká* válnak. A többi szerkezet-től függetlenül lehet őket ide-oda lehet tenni, csoportosítani a szótárban. Ez ad lehetőséget arra, hogy az adott szerkezethez tartozó szótári anyagot (minden „szócikk-részlet”) szigorúan *csak egyszer* írjunk meg, és azt használjuk fel a szótár különféle pontjain. Azaz, hogy ezekből a részletekből utólag (automatikusan) szerkesszük össze a szótárt. Ez is hozzájárul a szótár *egységességéhez*, nincsenek egymásnak ellentmondó előfordulások, megszűnnek az abból eredő problémák, hogy a szótár különböző részein más-más lexikográfus dolgozik. A mozgathatóságból következik a szótár könnyű *kiterjeszhetősége* is: a szótár kiegészíthető a ritkább szerkezetekkel úgy, hogy a mostani szótári anyagot teljes egészében, egy az egyben felhasználjuk, tőle függetlenül dolgozzuk fel az új szerkezeteket, majd automatikusan fésüljük össze a kettőt.

Budapest, 2010. október

Sass Bálint