

Sass Bálint–Váradi Tamás–Pajzs Júlia– Kiss Margit

## MAGYAR IGEI SZERKEZETEK

*A leggyakoribb vonzatok és szókapcsolatok szótára*

Budapest, Tinta Könyvkiadó, 2011, 508 oldal

A magyart mint idegen nyelvet tanítók gyakran szembesülnek azzal a ténnyel, hogy kevés a „kezüik alá dolgozó” könyv, irodalom. Nincsenek vagy alig akadnak – az angollal vagy a némettel ellentétben – kezdő nyelvtanulók számára írt vagy átírt, egyszerű szókincsű mesék, rövid történetek, a nemzet klasszikusai egyszerűen vagy feladatgyűjtemények, melyekhez az órára készüléskor nyúlhatnak.

Most végre egy olyan művet vehetnek a kezünkbe a gyakorló nyelvtanárok, amely összeállításakor a célcsoport tagjai között szerepeltek ők is. Ugyan nem a tanítás során közvetlenül használható anyagról van szó, de ez a kiadvány segítségükre lehet a mindennapi munkában. A négy szerző által jegyzett gyakorisági szótár a magyar igei szerkezeteket állítja középpontba. Vonzatszótár több nyelven létezik, de olyan összeállítás, amely az igét tartalmazó szerkezeteket ilyen formában gyűjti egybe, nem látott még napvilágot. Ahogy az *Előszóban* (7–8) a munkálatokat összefogó Váradi Tamás tollából értesülünk róla, a kötet jelentőségét nagyrészt az adja, hogy az alapul szolgáló anyag gyűjtése, többszöri szűrése és válogatása „emberi kéz érintése nélkül”, számítógéppel történt. Világszerte elterjedt eljárás már, hogy egy szótár szerkesztői szövegtárak, adatbázisok segítségével, az ezekből kinyert példák alapján állítják össze művüket. Arra viszont, hogy az anyaggyűjtés teljesen automatikusan történjék, eddig még nem volt példa a szótáriródalomban. A megjelenést megelőző munka a MTA Nyelvtudományi Intézetében csupán néhány évvel ezelőtt létrejött (bár azóta az intézet egyik legjelentősebb részlegévé lett) Nyelvtechnológiai Kutatócsoportban folyt.

S hogy mi alapján készült? A *Magyar Nemzeti Szövegtár* (MNSz) nevű korpuszt a folyóirat olvasóinak bizonyára nem kell részletesen bemutatni. Mivel ebből az elektronikus szövegyűjteményből dolgoztak a szerzők, egy rövid leírás mégsem kerülhető el:

A több mint 187 millió szövegszót tartalmazó korpusz a mai magyar írott köznyelvből vett, annak általános célú és reprezentatív mintája. Öt regionális nyelvváltozathoz (a magyarországi mellett *szlovákiai, kárpátaljai, erdélyi, vajdasági*) és öt stílusrétegből (sajtó, szépirodalmi, tudományos, hivatalos, személyes<sup>1</sup>) válogatták a szövegeket. A MNSz-ben tehát mind stílus-, mind pedig területi változat szerint kereshetünk. Ezek mellett lehetőségünk van minden szóról megtudni, hogy mi a töve, milyen szófaji kategóriába tartozik és milyen morfológiai elemzés adható róla. A különböző szempontok szerinti eredményeket nemcsak adatok egyszerű felsorolásaként kapjuk meg, hanem

---

<sup>1</sup> Ez az alkorpusz (a teljes anyag mintegy 10%-a) internetes fórumok (az index.hu és több kárpátaljai fórum) beszélgetéseit tartalmazza. Ez a nyelvi változat azért fontos, mert ez áll a legközelebb a spontán nyelvi megnyilvánulásokhoz, bizonyos esetekben nagyon hasonlít a beszélt, élő kommunikációhoz.

színes kördiagramok is bemutatják. A korpusz nagyságát tekintve könnyen kitalálható, hogy ezek az elemzések csak automatikusan, számítógépes programokkal voltak elvégezhetőek. Ebből adódik a csekély, mintegy 2,5 %-os hibalehetőség, melyre az alkotók a bevezetőben felhívják a figyelmet.

Az ingyenesen használható online adatbázis bárki számára hozzáférhető, aki kitölti a regisztrációs űrlapot, s elfogadja a szinte egyetlen feltételt, miszerint csak oktatási, kutatási vagy magáncélra használhatók az eredmények.

Amit tehát a kezünkben tartunk, az egy egynyelvű szótár, mely a magyar igéket veszi lajstromba. A kulcsszó már a címlapon feltűnik: **leggyakoribb** – és valóban ez a megszületett munka egyik fő erénye. A kötetbe – különféle automatikus szűrés és válogatás után – bekerült **minden** olyan igét tartalmazó szerkezet, amely legalább 250-szer előfordul a korpuszban. Ennek a feltételnek 6266 szerkezet felelt meg, s ezeket összesen 2226 címszóba rendezték a szerzők. A szócikkeken belül is a gyakoriság a rendezőelv. Értelmezéseket nem találunk – azaz a nem anyanyelvi beszélők csak egy bizonyos tudásszint elsajátítása után forgathatják haszonnal –, de ha megfelelően használjuk az egyes mutatókat, a kötet vonzatkeret-, kollokációs, gyakorisági és összehasonlító szótár is egyben. Ezek a mutatók nem pusztán az anyagban való eligazodást segítik, hanem önálló kutatási eredményeket rejtenek magukban, s megfelelően használva a lehetőségek széles tárházát nyitják meg előttünk. Minden egyes mutató a válogatás szempontjának megfelelően tartalmazza a szótár teljes anyagát.

A kötet két alapvető struktúrával operál, a **lexikálisan kötött** és a **lexikálisan szabad bővítmenyekkel** (LKB és LSzB). Lexikálisan kötött az a bővítmeny, amely tartalmazza szava konkrét, kötött szó, azaz nem helyettesíthető mással. A lexikálisan szabad bővítmenyek esetében a bővítmenyként megjelenő szó sok lehetőség közül választható – ezekben az esetekben a bővítmenyt esetragja vagy névutója képviseli (ez utóbbi két kategóriát a szótár szempontjából azonosnak tekintik a szerzők). A komplex igék egyszerre LKB-t és LSzB-t is tartalmaznak. A *derít* –*rA FÉNY-t* (37) szerkezetben például a *FÉNY-t* a lexikálisan kötött, míg a 'valamire' vonzat a lexikálisan szabad bővítmeny, azaz a *fényt derít valamire* (vonzatos) komplex ige.

Mivel nem egy hagyományos szótárt tartunk a kezünkben, a bevezető fejezetet érdemes és szükséges figyelmesen áttanulmányozni. A kötet koncepcióját kidolgozó és a számítógépes munkálatok oroszlánrészét elvégző Sass Bálint jegyzi ezt a részt (*A szótárról* 9–17). A bevezető fejezetcímei (*1. Mi ez?*; *2. Hogyan használjuk?*; *Kinek szól és mire jó?*) is tükrözik írójuk legfőbb a célját, a gyakorlati megközelítést. Elsődleges szempont lehetett megírásakor, hogy aki kezébe veszi ezt a könyvet, gyors és egyszerű eligazítást kapjon. Szerzője a kitűzött célt eléri, hiszen nem is gondolnánk, hogy számítógépes szakember tollából született a szöveg, annyira felhasználóbarát: szakszerű, de ezzel egyidőben egyszerű és könnyen érthető is. A használatot segítő, az anyagot jól átláthatóan közvetítő tipográfia nemcsak a bevezető tanulmányra, hanem a teljes kötetre jellemző. Több esetben jelentős információhordozó a betűtípus vagy egyéb tipográfiai megoldás. Ha ezeket megjegyezzük, könnyen használható a szótár. Például a hangrendi váltakozási pontot nagybetű jelöli a toldalékokban: *költ* –*rA* [költ valamire]; a névutókat egy eljűk kitett pont jelöli: *fordul* -*ellen* [valami ellen fordul].

A legterjedelmesebb a *Betűrendes szótári rész* (19–265), itt a vastagon szedett címszó alatt sorakoznak az előfordulások. A címszóban jelölve van, ha igekötős az ige (virgula választja el egymástól a két részt), utána kerek zárójelben a gyakorisági mutató látható, majd az egyes ide tartozó szerkezetek szintén gyakorisági sorrendben; az utánuk következő szám szögletes zárójelben áll, s szintén a korpuszbeli gyakoriságra utal. A kötet arra törekszik, hogy a szerkesztői intuíciót elkerülje, s ez meg is valósul minden olyan esetben, amikor a számítógépes technika ezt lehetővé teszi. Van azonban egy pont, ahol nem lehetett mellőzni a „humán erőforrást”. A betűrendes rész szócikkeinek minden tételénél szerepel egy példa. Ezeket a szerkezeteket a rendelkezésre álló mintából Kiss Margit és Pajzs Júlia válogatta.

A *Gyakoriság szerinti mutatóban* (197–266) a 136 354 előfordulással az első helyen álló *mond* -t-től csökkenő sorrendben van rendezve az anyag egészen a 251 előfordulást magáénak mondható 29 szerkezetig (ezek között találjuk például a következőket: *fel|oszt* -t, *meg|tisztel* -t -vAl és a *lesz IDEGES*). Ha ezt a részt kezdjük vizsgálni, szembesülhetünk azzal, hogy a kezdő nyelvtanulóknak elsőként tanított igék mennyire gyakoriak a korpuszban, s elgondolkozhatunk azon, hogy ha az eltérés jelentős, annak mi az oka, s hogy továbbra is így kell-e maradnia. (Megnyugtatóként egy személyes megjegyzés: az első 50 tételnek mindössze a 20%-át ítélem olyanoknak, ami feltételezhetően nem kerül elő a nyelvtanulás korai szakaszában.)

Ami jó a korpuszalapú építkezésben, abból fakad némi hátrány is, azaz a gépi feldolgozás oltárán fel kellett áldozni néhány olyan dolgot, ami egy csak emberi eszközökkel készült szótár esetében nem jelentett volna problémát. Ilyen például, hogy a *van* mellett a *lesz* külön igeként van feltüntetve.

Ha tanításhoz keresünk ötleteket, gyakorlatokat akarunk összeállítani, a *Keretek szerinti mutató* (267–353) jól használható az egyes ragok, a differenciáltabb névutóhasználat vagy a párhuzamos szerkezetek szemléltetésére; a lexikai feladatok esetében a *Kötött szavak szerinti mutató* (355–387) lehet segítségünkre, mely a lexikálisan kötött bővítmények fogalomszavai szerint csoportosítja az anyagot. Az *Igekötős keretek szerinti mutató* (389–419) és az *Alapige szerinti mutató* (421–504) címek magukért beszélnek, nem igényelnek bővebb magyarázatot.

A kötet nemcsak nyelvészeti, pszicholingvisztikai kutatások előkészítő anyagaként, fordítók, az általános és középiskolák magyartanárai számára hasznos, hanem igencsak haszonnal forgathatják a magyart idegen nyelvként tanulók és elsősorban az azt tanítók. Segítségével a tananyagok, nyelvtani gyakorlófeladatok összeállításánál nemcsak a már unásig ismételt példamondatokat használhatjuk, hanem friss és autentikus forrásból származó anyaggal készíthetjük el, ill. egészíthetjük ki munkánkat.

Azon elgondolkozhatunk, hogy egy számítógéppel született mű miért (csak) nyomtatott formában jelenik meg. A válasz egyszerű: egyrészt mind a mai napig szívesebben veszünk kezünkbe egy könyvet, mint kezdünk keresgélni az interneten. Az órai szemléltetést is egyszerűbb megoldani egy könyv, mint a nem mindenhol könnyen hozzáférhető technikai háttér használatával. Másrészt létezik online forma is, azaz az alapanyag így is elérhető: nemcsak az MNSz áll rendelkezésünkre, hanem a Mazsola<sup>2</sup> is. Ez egy

<sup>2</sup> Elérhető a <http://corpus.nyud.hu/mazsola> címen.

olyan – szintén Sass Bálint nevéhez kötődő – lekérdező felület, mely a korpuszban a vonzatos szerkezetek keresését teszi lehetővé (használati feltételei megegyeznek a szövegtárával). Persze a teljes szövegtár áttekintése és az eredmények gyakoriság szerinti rendezése meghaladja egy általános felhasználó képességeit, de épp ezt nyújtja ez a szótár.

Bizton állíthatom, hogy a Tinta Könykiadótól már megszokott, igényes kivitelezésű kötetet, mely *A magyar nyelv kézikönyvei* sorozat 21. tagjaként jelent meg, nemcsak a tanárok, hanem a tanulók is élvezettel és haszonnal fogják forgatni.

*Oszkó Beatrix*