

## *Szita Szilvia*

### **A MagyarOK nyílt korpusz használatáról**

#### **1. Bevezetés**

A tanulmányban bemutatott korpusz a *MagyarOK A1–B2* magyart mint idegen nyelvet tanító tankönyvcsaládhoz készült. Pedagógiai korpuszról van tehát szó, amelynek elsődleges célja a nyelvtanítás és a nyelvtanulás hatékonyabbá tétele (Cavalla 2019; Chambers 2019; Poole 2018). A pedagógiai célokra épített korpusz fogalma nem új a szakirodalomban: a kétezres évek elejétől egyre több ilyen gyűjtemény készül – igaz, leginkább az angol és más nagyobb nyelvek, például a francia tanításához (pl. André 2020; Braun 2007, 2010; Chambers 2019; Flowerdew 2009). Ezek a korpuszok általában autentikus vagy félautentikus írott szövegeket és/vagy szóbeli megnyilvánulásokat tartalmaznak, amelyek lehetővé teszik a felhasználó számára, hogy az anyanyelvi beszélők egy-egy kommunikációs helyzetekben jellemző nyelvhasználatát tanulmányozhassa. Másik funkciójuk, hogy olyan nyelvi modelleket szolgáltatnak a tanuló számára, amelyeket az saját nyelvi produktumaiba is beépíthet (Kennedy és Miceli 2010, 2017; Szita 2014; Szita és Pelcz 2017; Szita 2020).

Írásomban először röviden a pedagógiai korpuszok építésének általános kritériumait, majd a *MagyarOK* nyílt korpusz tartalmát ismertetem. A cikk második részében a korpusz hatékony felhasználására adok javaslatokat.

#### **2. Néhány gondolat a nyelvészeti és a pedagógiai céllal épített korpuszok különbségéről**

A kutatási céllal épített nyelvi korpuszok bőséges szakirodalmában (áttekintéshez ld. Stefanowitsch 2020) több alapos leírást is találunk arra vonatkozóan, hogy mely kritériumoknak kell megfelelniük az ilyen céllal készülő gyűjteményeknek. A három legfontosabb követelmény azt mondja ki, hogy a nagy nyelvi korpusznak autentikusnak, kiegyensúlyozottnak és reprezentatívnak kell lenniük. Az autenticitás annyit jelent, hogy a korpusz csak valóban elhangzott vagy leírt, nem nyelvészeti elemzés céljából megalkotott megnyilvánulásokat tartalmazhat, a kiegyensúlyozottság és a reprezentativitás értelmében pedig minden szövegfajtából elegendő mennyiségűt találunk benne, hogy megbízható, statisztikailag szignifikáns elemzési eredményeket kapjunk (pl. Brezina 2018; Stefanowitsch 2020)<sup>1</sup>. A pedagógiai korpuszok építésével kapcsolatos szakirodalom lényegesen kisebb, különösen az alacsonyabb nyelvtudási szintekre szánt gyűjteményekre vonatkozó kritériumok tekintetében. Ennek az az oka, hogy kezdetben a kutatók nem tettek különbséget a kétféle (nyelvészeti és nyelvtanítási céllal létrehozott) korpusz között. A korpuszhasználatot is alapvetően magasabb nyelvi szintek és egyetemista nyelvtanulók számára javasolták, akikről feltételezték, hogy képesek autentikus megnyilvánulások megfigyelésére és elemzésére. A gondolat, hogy a korpuszhasználat alacsonyabb nyelvi szinteken is hasznos lehet, és ennek érdekében az eredeti korpusz módosítására lehet szükség, később merült fel. Ma azonban egyre inkább az az uralkodó álláspont, hogy – mivel a pedagógiai korpuszok céljai csak részben egyeznek meg a nyelvészeti-kutatási céllal összeállított korpuszokéival – az

<sup>1</sup> A kisebb, egy kiválasztott szövegtípust tartalmazó korpuszokra (pl. egy olyan korpuszra, amely a *Covid*dal kapcsolatos fórumbejegyzéseket tartalmazna) a kiegyensúlyozottság elve értelemszerűen nem vonatkozik.

összeállításukra vonatkozó kritériumok sem lehetnek ugyanazok (pl. Aston 2001; André 2020; Braun 2007, 2010; Chambers 2019). Kissé leegyszerűsítve azt mondhatnánk, hogy a nyelvészeti célokkal összeállított korpuszokba kizárólag autentikus nyelvi megnyilvánulások kerülnek, amelyekben a gyűjtő semmiféle módosítást nem végzett; a pedagógiai korpuszok készítőinek célja ezzel szemben az, hogy olyan szövegeket kínáljanak a nyelvtanuló számára, amelyeket az adott nyelvi szinten be tud fogadni és amelyek elemeit képes beépíteni saját nyelvhasználatába. A korpusz tehát egyrészt a tanuló számára nehezen értelmezhető nyelvi jelenségeket annak tudásszintjén befogadható példákon hivatott megvilágítani, másrészt a korpusz anyagát a tanuló modellként is használhatja saját szövegeinek megírásakor. A pedagógiai korpuszok építésének az alábbi négy alapelvét azonosíthatjuk (Aston 2001; Braun 2007, 2010; Kennedy és Miceli 2010, 2017 alapján). Ezeket az alapelveket tartottunk szem előtt a *MagyarOK* nyílt korpuszának építése során is (ld. Szita 2020):

- (1) A pedagógiai korpusz a tanuló nyelvtudási szintjén befogadható szövegek gyűjteménye. Ennek érdekében – különösebb alacsonyabb nyelvtudási szinteken – az autenticitás kérdésében bizonyos kompromisszumok elfogadhatóak. Így a pedagógiai korpusz tartalma magasabb nyelvi szinteken (B2–C2) legtöbbször semmiféle módosítást nem igényel, alacsonyabb szinteken a **természetes nyelvhasználatot tükröző, ám esetenként adaptált és rövidített szövegek** gyűjteménye is lehet. Vagyis a korpusz tartalmának autentikus, valódi közléshelyzetekben elhangzott/leírt nyelvi produktumokon kell alapulnia, de ezek módosítása – az alacsonyabb nyelvi szintek kiszolgálása érdekében – megengedett<sup>2</sup>.
- (2) A korpusz tartalmának **relevánsnak kell lennie**, azaz a témáknak, szövegtípusoknak hasznosnak kell lenniük a tanulók számára, így például meg kell felelniük a Közös Európai Referenciakeret (KER) elvárásainak. Ezért a pedagógiai korpuszok építésénél fontos követelmény az intertextuális koherencia. Ez azt mondja ki, hogy hatékonyabban használható a pedagógiai korpusz, ha szövegei tematikusan összefüggenek (pl. egy-egy alkörpuszban ugyanazzal a témával kapcsolatos szövegek találhatók), mert a tanuló így behatárolt és bizonyos szintig ismétlődő szókinccsel dolgozhat.
- (3) **A pedagógiai korpusz nyelvezetének összhangban kell állnia a tanuló nyelvi szintjével**, hiszen ez mindenekelőtt akkor használható hatékonyan, ha a tanuló külső (tanári) segítség nélkül is értelmezni tudja a benne szereplő nyelvi megnyilvánulásokat.
- (4) A pedagógiai korpusz ideális méretéről ugyan megoszlanak a vélemények, de abban a kutatók többsége egyetért, hogy **a korpusznak megfelelő számú megnyilvánulást kell tartalmaznia** ahhoz, hogy az adott témához és/vagy nyelvi szinthez rendelhető kulcsjelenségek megfigyelhetővé váljanak benne.

Amint a fentiekből talán kiderül, pedagógiai korpuszt építeni nem könnyű feladat. Időigényes elfoglaltságról van szó, amely ráadásul számos kompromisszummal jár: a korpuszépítőnek folyamatosan olyan döntéseket kell hoznia, amelyek kihatnak a korpusz tartalmára, jellegére. A hasznosság, használhatóság és hatékonyság érdekében a szövegeket előzetesen ki kell választania, ellenőriznie és javítania kell. Esetenként rövidítenie is kell rajtuk, és egy részüket át kell írnia, hogy a nyelvtanuló számára érthetőek, értelmezhetőek legyenek. Ezek a módosítások bizonyos fokig aláássák a korpusz tartalmának autentikus voltát, illetve a

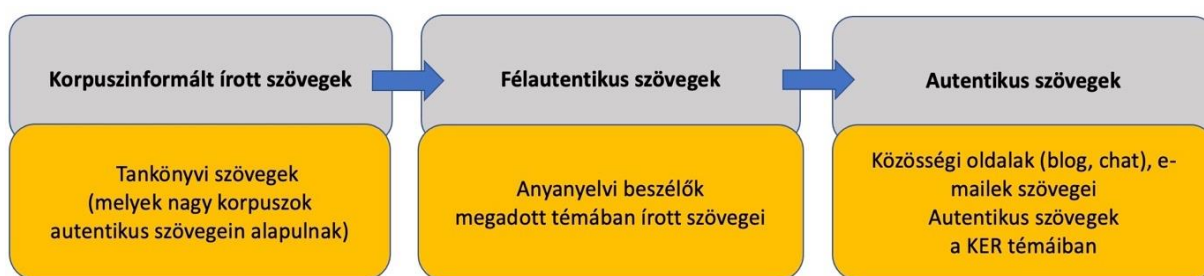
<sup>2</sup> Megjegyezzük, hogy egyes kutatók az adaptált szövegek autenticitását nem feltétlenül kérdőjelezzik meg, különösen, ha a korpuszépítő a szövegeket elsősorban rövidítette, de szókinccsüket csak kivételes esetekben módosította (pl. Meunier–Reppen 2015).

tartalmának reprodukálhatóságát. Ugyanakkor arról sem szabad megfeledkeznünk, hogy ha a pedagógiai korpusz építője minden esetben megőrizné a szövegek eredeti formáját – vagyis maximálisan ügyelne az autenticitás elvének betartására –, az a korpusz használhatóságát korlátozná<sup>3</sup>.

### 3. A MagyarOK nyílt korpusz rövid bemutatása

A MagyarOK tankönyvcsaládhoz készült pedagógiai korpusz végleges állapotában legalább másfélmillió tokenből fog állni, amelyet lehetőség szerint a későbbiekben tovább bővítünk. A korpusz 2021 őszén mintegy 485 000 tokent tartalmaz (az anyagok egy része még feldolgozás alatt áll), és a [MagyarOK nyílt korpusz](#) már most is bárki számára regisztráció nélkül elérhető a *Sketch Engine* felületéről (a pontos elérési címet ld. a szakirodalmi jegyzékben). A korpusz négy szintre tagolódik (A1, A2, B1 és B2), amelyek közül itt az A1–B1-es szintekkel foglalkozunk. Ezeken a szinteken a KER leírásai szerint a nyelvtanuló még nem tekinthető önálló nyelvhasználónak. Ahhoz, hogy a korpusz az A1–B1-es szintű nyelvtanuló számára használható anyagot tartalmazzon, minden nyelvi szinten egy háromlépcsős rendszerben jutunk el a tananyagban szereplő, vagyis természetes nyelvhasználatot tükröző, adaptált szövegektől az autentikus szövegekig.

- A korpusz első része a MagyarOK A1–B1 tankönyvek szövegeit, azaz körülbelül 150 000 tokent tartalmaz. Lényeges megjegyeznünk, hogy a tananyagban található szövegek legnagyobb része adaptált (rövidített, némileg leegyszerűsített és helyenként átírt), ám autentikus nyelvi inputon alapuló anyag.
- A második alkorpuszban anyanyelvi beszélők félautentikus szövegeit találjuk. Ez az alkorpusz jelenleg az A1-es és az A2-es szinten érhető el és összesen 61 000 tokent tartalmaz.
- A harmadik, mintegy 500 000 tokent tartalmazó alkorpusz pedig különböző műfajú autentikus szövegekből (újságcikkekből, blog- és fórumbejegyzésekből) és személyes szövegekből, például e-mailekből és a közösségi oldalakon folytatott interakciókból fog állni (jelenleg fejlesztés alatt, várhatólag 2022 végétől elérhető). A korpusznak ez a része terveink szerint folyamatosan bővülni fog.



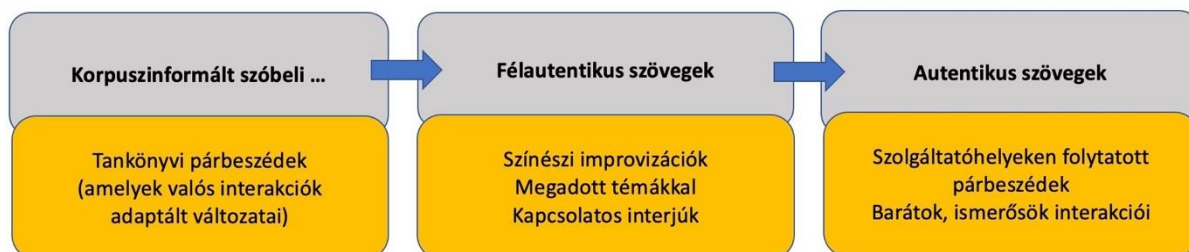
1. ábra: Az A1–B1-es szintre készült írott korpusz tartalma

<sup>3</sup> Ezen a helyen meg kell említenünk, hogy egyes kutatók – így például Johns (1991), Chambers et al. (2011) vagy Tyne (2012) – a nyelvészeti céllal készült korpuszok használatát ajánlják, és kitartanak amellett, hogy minden körülmények között autentikus nyelvi példákkal kell dolgozni. A magasabb nyelvi szinten álló nyelvtanulók számára ez valóban járható út lehet, ám a teljes mértékig autentikus korpuszok csak nehezen használhatók alacsonyabb nyelvtudási szinteken. Ezek szövegei ugyanis komplex szókincset tartalmaznak, amely megnehezíti a nyelvi megnyilvánulások rekontextualizálását (Widdowson 2003). További gondot jelent, hogy az autentikus nyelvi produktumok olykor nyelvhelyességi hibákat, stilisztikai szempontból problematikus megnyilvánulásokat és nem vállalható, szélsőséges nézeteket is tükrözhetnek.

Az írott korpuszt a szóbeli interakciók korpusza egészíti ki. Az írott korpuszhoz hasonlóan itt is három lépcsőn keresztül jutunk el a természetes nyelvhasználatot tükröző interakcióktól a félautentikus megnyilvánulásokon keresztül az autentikus megnyilvánulások alkorpuszáig.

- Az első alkorpusz a tankönyvi párbeszédék átíratát tartalmazza (kb. 62 100 token). Ezek a korpusz által inspirált, de szerkesztett beszélgetések különböző interakciók megvalósításához kínálnak modelleket, és a szóbeli interakciók számos jellegzetes vonását mutatják.
- A második alkorpusz kétféle félautentikus interakció átíratát tartalmazza: (1) színészi improvizációk (jelenleg 30 700 token, egy része még fejlesztés alatt) és (2) rövid, kérdésekkel irányított interjúk (kb. 110 400 token).
- A harmadik alkorpusz autentikus szóbeli megnyilvánulások átírataiból áll, és szintén két részre oszlik: (1) szolgáltatókkal folytatott párbeszédék (kb. 35 880 token), valamint (2) félhivatalos és informális (barátok, munkatársak, családtagok közötti) beszélgetések (kb. 34 500 token). Ez az alkorpusz egyébként számos olyan felvétel leíratát is tartalmazza, amely a tankönyvekben szereplő párbeszédéknek is alapjául szolgált. Az átírás az *Alrite* szoftverrel készült, a szövegeket manuálisan javítottuk. A korpusznak ez a része terveink szerint folyamatosan bővülni fog.

A teljes A1–B1-es szóbeli korpusz jelenleg mintegy 275 000 tokent tartalmaz (több mint 38 órányi felvétel, vagy körülbelül 115 token/perc). Ez jóval több anyag, mint amennyit bármely tankönyv nyújtani tudna. Az első és második alkorpusz alapjául szolgáló videók és hangfelvételek egy része elérhető a tankönyv honlapján és a *MagyarOK – Magyarul beszélni menő* YouTube-csatornán. A harmadik alkorpusz videó- és hanganyaga jelenleg feldolgozás alatt áll, amelynek célja a felvételek hangminőségének javítása, valamint az adatvédelmi problémák megoldása. (A korpuszok részletesebb bemutatásához ld. Szita 2020.)



2. ábra: Az A1–B1-es szintre készült szóbeli korpusz tartalma

A *Sketch Engine* honlapján a nyelvtanulóknak és nyelvtanároknak számos könnyen kezelhető eszköz áll rendelkezésükre, amelyek segítségével különböző szempontok szerint kereshetnek a korpuszban. Így a *Word Sketch* eszközzel több elemből álló nyelvi egységek (kollokációk) kérdezhetők le, a *Concordancer* konkordanciasorai számos példát szolgáltatnak a keresett nyelvi elem (morféma, szó vagy kifejezés) használatára, a *Wordlist* segítségével pedig szófaj szerinti gyakorisági listák állíthatók fel. Az eszközök használatát az alábbiakban egy-egy feladathoz kapcsolódóan röviden be is mutatjuk. Mindegyik eszközzel a teljes korpuszban vagy a választott alkorpuszban is lehet keresni.

#### 4. A korpuszalapú tanulás (*data-driven learning*)

A nyelvórai feladatok célja nyilvánvalóan az, hogy lépésről lépésre közelebb vigye a tanulókat végső céljukhoz, az idegen nyelv magas szintű elsajátításához. E cél elérését a korpuszalapú feladatok nagy mértékben segíthetik: egyrészt statisztikai alapon azonosíthatóvá teszik a gyakran használt nyelvi elemeket (általánosságban vagy nyelvi szint, kontextus és/vagy téma szerint), másrészt számtalan példával segítik bevézésüket. Ugyanakkor arra is ösztönzik a tanulót, hogy reflektáljon a célnyelv működésére, hiszen „a tudatos nyelvelemzés megkerülhetetlen a tanulási folyamat során”, és „a nyelvóra az a hely, ahol ezt lehet és kell is ösztönözni” (Salazar 2014: 158). Végül a korpuszhasználat arra is lehetőséget nyújt a tanulóknak, hogy a hasznosnak talált nyelvi elemeket beépítsék saját szövegeikbe (Kennedy és Miceli 2010, 2017; Szita 2020).

A korpuszhasználat módszertanát az úgynevezett „korpuszalapú tanulás” (*data-driven learning*) szakirodalma foglalja össze. A fogalom megalkotása és tartalmának kidolgozása elsősorban Johns (1991) nevéhez fűződik. A korpuszalapú tanulás lényege az anyanyelvi beszélők nyelvhasználatának megfigyelése célzott kérdések alapján, a tanár pedig e folyamat kísérése és a nyelvi jelenségek, használati tendenciák bevézését szolgáló, korpuszalapú feladatok összeállítása. Fontosnak tartjuk megemlíteni, hogy a „korpuszalapú tanulás” nem egy teljesen új tanításmódszertant, csupán egy új eszköz optimális használati lehetőségeit jelenti. Azon kérdések sorába, amelyek megválaszolásában a korpusz különösen nagy segítséget nyújt, tartoznak például a következők: Hogyan használják a „...” szót? Mit jelent pontosan az, hogy „...”? Mi a különbség „...” és „...” (két szó vagy kifejezés) között? Mondhatjuk-e azt, hogy „...”? (Johns 1991a, 1991b; Sinclair 2003) Amint látjuk, jellegetesen olyan kérdésekről van szó, amelyekre más eszközök – például nyelvtankönyvek, szótárak – segítségével, de még a nyelvész vagy a nyelvtanár intuícióját bevonva sem lehetséges teljes és kielégítő választ adni. A felsorolt eszközök korlátozott hatékonyságának oka az, hogy ezek csupán egy-két példával, egyszerű használati szabállyal és/vagy definícióval tudnak szolgálni, holott teljes választ általában nagyobb számú nyelvi megnyilvánulás megfigyelése – vagyis a korpuszelemzés – adhat. Az ismételt találkozásoknak köszönhetően „a tanuló a választott szót [vagy kifejezést] többféle kontextusban figyelheti meg, megismerheti tipikus kollokációs partnereit, jellemző szöveggörnyezetét és azokat a regisztereket, amelyekben előfordul. Ezáltal elmélyítheti a róla szerzett tudását” (Allan 2009:24).

Johns (1991:5) a korpusz lekérdezéséhez és az eredmények rendszerezéséhez három egymást követő lépést javasol, melyek közül az első a megfigyelés, a második az osztályozás, a harmadik pedig az általánosítás szakasza. A megfigyelés szakaszában a nyelvtanulók egy vagy több nyelvi (lexikai, nyelvtani, szemantikai, pragmatikai) mintázatokat (*linguistic patterns*) azonosítanak, az osztályozás fázisában csoportosítják megállapításaikat, végül megvizsgálják, hogy a megfigyelt mintákat lehet-e általánosítani más példákra (konkrét példákhoz ld. az 5.1. és az 5.2. fejezetrészt).

#### 5. Szóprofil készítése, nyelvi mintázatok (*linguistic patterns*) feltárása

Amikor arról beszélünk, hogy a tanuló a korpuszban megfigyelt példák segítségével nyelvi mintázatokat tárhat fel (Boulton 2010; Boulton és Thyne 2014; Hoey 2005; Hunston 2008, 2010, 2015; Sinclair 2003), fontos szem előtt tartani, hogy – amint a pedagógiai korpuszok nem elsősorban nyelvészeti-kutatási célokat szolgálnak, úgy – a tanuló (és a tanár) sem korpusznyelvész. Chambers (2019: 472–3) szerint a tanár feladata, hogy segítse a diákokat egy **elemzési módszert** és néhány olyan egyszerű **eszköz** elsajátításában, amelyek segítségével az adott **nyelvi szint anyagát** új, hatékony szempontok szerint **tanulmányozhatja** és

**rendszeresheti**, de ne próbálja meg korpusznyelvészeket faragni belőlük. A következő oldalakon ennek az alapelvnek a lehetséges megvalósítására mutatunk be néhány példát. A korpusz alapján megválaszolható nyelvi kérdéseket jellemzően két nagy csoportra oszthatjuk (elsősorban lexikai és elsősorban grammatikai jellegű kérdések), bár – amint az 5.1. és 5.2. pontban látni fogjuk – az elemzés során mindkét esetben feltárhatók lexikai, grammatikai, szemantikai és pragmatikai jellemzők is.

### 5.1. Korpuszalapú válasz lexikai kérdésre

A tanuló kérdése: Hogyan használhatom a *nagy* melléknevet? Mit jelent pontosan?

Eszköz: *Word Sketch*

A különösen nagy gyakorisággal előforduló szavak esetében a nyelvtanuló számára nem is annyira a kontextus nélkül álló szó jelentésének megértése okoz problémát (például a *nagy* szót a *'big'*, *'large'*, *'great'* szóval fordíthatjuk angolra, és szinte minden más nyelvben találunk – első látásra – pontos megfelelőt). A nyelvtanuló számára inkább az jelent gondot, hogy a gyakori szavaknak – kontextustól függően – számos (nemegyszer idiomatikus) jelentésük, ennél fogva igen sok kollokációs partnerük van (ld. pl. Nation 2013; Szudarski 2017). Ez alól a *nagy* melléknév sem kivétel.

Ahhoz, hogy a korpuszból feltárható jelentéseket és használati tendenciákat a tanuló – önállóan vagy tanári segítséggel – megfigyelhesse és rendszerbe állíthassa, a szakirodalmat követve négy jellemző feltérképezését javasoljuk: a lexikai, grammatikai, pragmatikai és szemantikai jellemzőkét. Ezt a négy jellemzőt mindenekelőtt Sinclair (1991, 2004), Hoey (2005) és Stubbs (2009) írták le korpusznyelvészeti szempontból, és többek között O’Keeffe et al. (2007), Flowerdew (2009), valamint Omidian és Siyanova-Chanturia (2020) adaptálták a nyelvtanuló szükségleteire. O’Keeffe et al. (2007: 14–17) a négy jellemző alapján úgynevezett „szóprofilok” (*word profiles*) felállítását javasolja. A korpusznyelvészeti eredményei alapján a jellemzők jelentőségét és a feltárásukat segítő munkamódszert következőképpen magyarázhatjuk el a nyelvtanulóknak:

- A legtöbb szó környezetében bizonyos szavak a véletlenszerűnél jóval nagyobb, statisztikai szempontból is szignifikáns gyakorisággal jelennek meg. Ezek a szó kollokációs partnerei, amelyekkel többszavas egységeket alkot. Figyelje meg és írja ki a kiválasztott elem környezetében rendszeresen előforduló szavakat! Használja a *Word Sketch* és a *Concordancer* eszközöket!
- A választott szót tartalmazó többszavas egységekhez gyakran társíthatók nyelvtani jellemzők (mintázatok) is. Ezek a többszavas egység kolligációi. Tanulmányozza a nyelvtani jellemzőket! Igék esetében például megvizsgálhatja a következő kérdéseket: Van-e olyan igeidő, igemód, szám és személy, főnévrag vagy más nyelvtani jellemző, amely rendszeresen kíséri a választott szót? Melléknevek esetében például a következő kérdésekre kereshet választ: Alap-, közép-, vagy felsőfokban jelenik-e meg túlnyomórészt a melléknév? Módhatározói vagy jelzői szerepet tölt-e be a mondatban? stb. A grammatikai jellemzők sorát mindig érdemes kiegészíteni a szórendre vonatkozó sajátosságok megfigyelésével is.
- Sok többszavas egységnek szemantikai jellemzői is vannak, ami a korpusznyelvészeti terminológiájában annyit jelent, hogy ezek az egységek gyakran vagy pozitív, vagy negatív töltetűek, és nemritkán jól körülhatárolható jelentésmező is társul hozzájuk. Tanulmányozza a többszavas egység jelentését! Pozitív vagy negatív töltetet hordoz-e az egység? Vannak-e egyéb, a jelentést pontosító jellemzők, például: elvont vagy konkrét

dolgokra vonatkozik-e inkább az egység? Minek a leírására használjuk tipikusan (emberek, természeti jelenségek stb.)?

- Sok többszavas egységhez pragmatikai jellemzők, vagyis tipikus kommunikációs helyzetek is társíthatók. Milyen szövegtípusban fordul elő? Formális vagy informális, írásbeli vagy szóbeli kommunikációs helyzetekben használjuk-e tipikusan a választott szóval alkotott többszavas egységet?

A konkrét munkafolyamat szemléltetésének első lépéseként a *nagy* melléknévvel alkotott melléknév–főnév szerkezetek listáját kérdezzük le a *Word Sketch* eszköz segítségével. A melléknév 282 alkalommal fordul elő az A1–B1-es korpuszban – ez egymillió tokennél 885 előfordulást jelent –, tehát gyakori szóval állunk szemben. Az alábbi táblázat a korpuszban előforduló tipikus kollokációs partnereket mutatja (gyakorisági mutatók alapján sorba rendezve):

nouns modified by "nagy"	szoba	...	fürdőszoba	...
	család	...	választék	...
		...	szereket	...
kert	konyha	...	előny	...
ház	káosz	...	Budapest	...
lakás	zaj	...	élmény	...
város	Dr	...	projekt	...
iroda	Dr	...	napilap	...
iroda	tehetség	...	iskola	...
rész	iPhone	...	múzeum	...
íróasztal	régió	...	épület	...
		...	ÁÁÁ-t	...
csend	megye	...	Caesar	...
erdő	nyugalom	...	hőség	...
asztal	különbség	...	kedvelő	...
tér	hal	...	rajongó	...
		...	sportember	...
		...	forgalom	...

3. ábra: A 'nagy' melléknévhez társuló főnevek a MagyarOK A1–B1-es szintű korpuszában

Mivel a pedagógiai korpusz a tanuló nyelvtudási szintjén, az ő igényeihez igazodó nyelvi inputot tartalmaz, mérete és tematikája egyaránt korlátozott (a MagyarOK nyílt korpusza egyértelműen a KER témáihoz illeszkedik), így számolnunk kell azzal, hogy nem tükrözi minden tekintetben hűen az anyanyelvi beszélők nyelvhasználatát. Például a lakás, lakóhely témája a KER A1, A2 és B1-es szintjén is hangsúlyosan szerepel, így nem meglepő, hogy korpuszunkban a *kert*, *ház*, *lakás*, *íróasztal* főnevek a *nagy* melléknév gyakori kollokációs partnerei. Érdekes azonban megjegyezni, hogy jól összeállított pedagógiai korpuszok esetén – legalább a gyakori szavak tekintetében – egy nagy méretű általános korpuszsal összehasonlítva igen hasonló eredményeket kapunk. A *huTenTen2012* korpuszban a következő gyakori kollokációs partnereket találjuk a *nagy* szóval: *rész*, *siker*, *probléma*, *baj*, *öröm*, *szerep*, *szükség*, *dolog*, *meglepetés*, *segítség*, *előny*, *terület*, *nap*, *változás*, *mérték*. A felsorolt kollokációs partnerek közül jó néhány szerepel a MagyarOK nyílt korpuszában.

Ugyanakkor fontos megjegyezni, hogy – bár a szavak többsége valóban a keresett szó kollokációs partnere – a lista tartalmazhat tévedéseket is. Így például a fenti listán a *dr.* szó a *Nagy szeretettel köszöntöm dr. Szikszai Dénest* mondatban szerepel, tehát a *nagy* melléknév valódi kollokációs partnere a *szeretettel* főnév. Ugyanígy az *iPhone* esetében sem *nagy*

*iPhone*-ról, hanem az *iPhone legnagyobb előnyéről* van szó, vagyis itt az *előnye* a nagy kollokációs partnere. A tévedések kiküszöbölése érdekében mindig érdemes lehívni néhány, az adott többszavas egységhez tartozó példamondatot, amit a kiválasztott kollokációs partner melletti három pontra kattintva tehetünk meg.

A példák fenti szempontok szerinti elemzése alapján a *nagy* melléknév profilja a következőképpen mutatható:

<b>Gyakori kollokációk: konkrét jelentésű főnevek</b>	város, ház, kert, múzeum
<b>Gyakori kollokációk: elvont jelentésű főnevek</b>	rész, siker, probléma, gond, öröm, szerep, szükség, dolog, meglepetés, segítség, előny, nap, változás, mérték, ember
<b>Kolligációk (grammatikai jellemzők)</b>	<b>Adj + N.</b> Leggyakrabban felsőfokban: <i>a legnagyobb rész/probléma</i> . Leggyakrabban középfokban: <i>nagyobb mértékben</i> . Leggyakrabban alacsonyfokban: <i>nagy meglepetés, nagy segítség, nagy siker, nagy gond, nagy nap, nagy ember</i> stb.
<b>Szemantikai jellemzők: konkrét jelentésű főnevek</b>	Egy adott dolog méretét, arányát jelöli ki.
<b>Szemantikai jellemzők: elvont jelentésű főnevek</b>	(1) Dolgok, események, emberek fontosságára utal (legtöbbször pozitív konnotáció). (2) Egy probléma, gond súlyos voltára utal (negatív konnotáció).
<b>Pragmatikai jellemzők</b>	Különbözőbb pragmatikai jellemzők nem társíthatók a szóhoz.

4. ábra: A 'nagy' melléknév profilja

A vizsgálódásnak természetesen nem kell a kétszavas egységek rendszerezésével véget érnie. A tanulók három- vagy négyzavas egységeket is megfigyelhetnek, így többek között lekérdezhetik a kiválasztott kételemű kollokációkhoz társuló gyakori igéket is. Például a *nagy siker* kifejezést tartalmazó többszavas egységek a következők: *nagy sikert arat, nagy sikernek számít, nagy sikere van*. Ezeket a kifejezéseket a tanulók szintén rögzíthetik a füzetükben.<sup>4</sup>

## 5.2. Korpuszalapú válasz nyelvtani kérdésekre

A tanuló kérdése: Miért van az, hogy a *pedig* kötőszó néha a mondat elejére kerül, néha más-hová? Felcserélhetők-e ezek a pozíciók?

Eszköz: *Concordancer*

A korpuszt használhatjuk elsődlegesen nyelvtani jellegű kérdések megválaszolására is. A munkafolyamatot a *pedig* kötőszó vizsgálatán fogjuk szemléltetni. A szó jelentése hol a *de*, hol az *és*, hol pedig a *viszont* kötőszóéhoz áll közelebb, így érthető, ha a nyelvtanulók nehezen

<sup>4</sup> Egy lekérdezés a *huTenTen2012* korpuszban azt mutatja, hogy a *nagy siker* többszavas egység igéi megegyeznek a pedagógiai korpusz szolgáltatása eredményekkel.



tudják helyre tenni a használatát. Ebben a korpuszból nyert konkrét példák elemzése segíthet, hiszen a példák láthatóvá tehetik a szó használati sajátosságait és tipikus jelentéseit.

A *pedig* kötőszó a *MagyarOK* A1–B1 korpuszában 236, az A1–A2 korpuszban 161, az A1-es korpuszban pedig 84 alkalommal fordul elő, vagyis még az egyetlen nyelvi szintet tartalmazó alkorpuszban is jóval több konkrét (természetesen hangzó, félaumentikus vagy autentikus) példát találunk a használatára, mint egy szótárban vagy nyelvtankönyvben. A következő táblázat az A1-es korpuszból mutat be tíz véletlenszerűen választott konkordanciasort. A kulcsszót (*keyword in context* vagy *KWIC*) piros szín jelöli.

jobb jegy, az egyes pedig a legrosszabb. Nálunk a legrosszabb jegy a hatos, a legjobb pedig az egyes. - Nálunk húsz jegy van. A húszas a legjobb, és tíz alatt minden jegy rossz. A	
ció. Egy családnál laktunk. Mi azért nem értettük őket, mert túl gyorsan beszéltek, őket pedig az akcentusunk zavarta. Szerencsére nagyon türelmesek voltak. Kétszer-háromszor is	
I vizsgáló? Tamás: Anatómiából. Rebeka: Hú! És hogy állsz? Tamás: Hát, nem túl jól, pedig tegnap is egész nap tanultam. Csak délben tartottam egy kis szünetet. Pizzát rendeltem	
és az óbudai egyetem. A pécsi egyetemet 1367-ben alapították, az óbudai egyetemet pedig 1395-ben. Mindkét egyetem az 1410-es évekig működött. Ma a legnagyobb egyetem	
a fizetted az ebédet? - Nem. Meghívott a barátom. Egyszer ő hív meg engem, egyszer pedig én őt. Utazás: 1. Várta már a nyarat? - Jaj, nagyon! Nagyon elfáradtam az első félév	
órára. Béci legtöbbször fáradtan érkezik. József biztosan nagyon jól érettségizik, Béci pedig sajnos rosszul. A kémia tanárom unalmasan magyaráz, mégis szeretem, mert igazságc	
beszélt velem, mert így gyorsan fejlődtem. 12 hónapig voltam ott. Nem utaztam sokat, pedig gyönyörű helyek vannak nem messze Firenzétől. Mégsem sajnálom, mert így sokat	
A fizikai munkát nem bírnám. Ráadásul kreatív lehetek. Ez is előny. Nem utolsósorban pedig sok pénzt kereshetek. Nem szeretnék fodrász lenni. - Én igen. Szerintem jó lehet egész	
munkám. Csak az a baj, hogy az emberek nem szeretnek hozzám járni. Félnék tőlem. Pedig aki szép, egészséges fogakat akar, az félévente egyszer eljön a rendelőmbe. Nagyon	
abban foglaljunk, amelyikhez golfpálya is tartozik és ahol van fitness terem. - Szerintem pedig abban, amelyikbe könnyen el lehet jutni és ahonnan szép a kilátás. 1. Olyan szállodát	
szott jelentős szerepet. Abaliget a cseppkőbarlangjairól nevezetes, Püspökszentlászló pedig az arborétumáról és gyönyörű természeti környezetéről. A2_MF7/1. 1. A hűgöm	

5. ábra: Példák a 'pedig' kötőszó használatára a MagyarOK A1-es korpuszában

A konkordanciasorokból kiindulva a tanulók megfigyelhetik a nyelvtani elem környezetét és a felmerülő szerkezeti ismétlődéseket. Ehhez legalább húsz konkordanciasort olvasnak el. Ha a szöveggörnyezet nem világos, a tanuló rákattinthat a kulcsszóra; a kattintás egy új szövegdobozt nyit meg, amely a tágabb szöveggörnyezetet – a kulcsszót tartalmazó mondat előtti és utáni néhány mondatot – mutatja. Az alábbi, célzott kérdésekkel segíthetjük a példák elemzését: Hol helyezkedik el a szó a mondatban? Mi előzi meg és mi követi? Vannak-e hasonlóságok az első és a második tagmondat szerkezetében? Van-e különbség a *pedig* kötőszó jelentésében a szó pozíciójától függően?

A következő lépésben a tanulók csoportosítják a példákat. A *pedig* esetében két fő csoportot különböztethetünk meg: (1) a kötőszó a tagmondat elején áll, mint a 3., 4. és a 7. mondatban, vagy (2) egy elem (szó vagy kifejezés) után második pozícióba kerül, mint az összes többi mondatban. A 3., 4. és 7. mondat alapján a tanulók arra a (helyes) következtetésre juthatnak, hogy amikor a *pedig* szó első helyen áll, akkor jelentése nagyjából a *mivel*, *bár* kötőszóéna felel meg. A többi mondatban a jelentése közelebb áll az *és*-hez, azzal a különbséggel, hogy a *pedig* nemcsak újabb elemmel egészít ki egy sort, hanem ezt az elemet az első tagmondattal szembe is állítja. Például: *oroszlóvoda – olaszgimnázium; 17 óráig/iskola – 17 óra után/otthon*. Az egyik szembeállított elem ilyenkor közvetlenül a kötőszó előtt áll, a másik utána.

Végezetül a tanulók a példamondatokból leszűrt következtetések érvényét további példákön vizsgálják meg, majd összefoglalják megfigyeléseiket. A 6. ábra két táblázata ezt mutatja.

Nagyon sokfajta nyelvtani kérdésre találunk hasonló módszerrel választ. Lekérdezhettük egy választott rag, például a *-ban/-ben* névszórag előfordulását, és feltérképezhetjük ennek használatát. Megfigyelhetjük, mely hely- és időhatározók kapják ezt a ragot, valamint azt is, hogy használatukban fellelhető-e bármiféle szabályszerűség. Ugyanígy rákereshetünk egy-egy nehézséget jelentő igekötőre (*meg-*, *el-*) vagy hasonló jelentésű igékre (ld. Szita 2017 és 2018), és ezek használati lehetőségeit is rendszerezhetjük a fenti szempontok alapján. Hasonlóképpen megfigyelhetjük a gyakori igék – például a *van* ige – különféle jelentéseit.

**pedig (1)**

<b>Kollokációk</b>	Nincs egyértelműen a szóhoz társítható kollokáció.
<b>Kolligációk (grammatikai jellemzők)</b>	<b>SZÓ/KIFEJEZÉS + pedig + tagmondat</b> A <i>pedig</i> a mondat második eleme. Az az elem előzi meg, amelyiket szembeállítjuk az első tagmondat valamely elemével.
<b>Szemantikai jellemzők</b>	Az első tagmondat egyik elemét szembeállítjuk a második tagmondat hasonló funkciójú elemével : <i>A régi város gyönyörű volt, a spanyolok pedig nagyon barátságosak. (a régi város gyönyörű – a spanyolok barátságosak)</i> <i>Az általános iskolában oroszul tanultam, a gimnáziumban pedig olaszul. (az általános iskolában oroszul – a gimnáziumban angolul)</i>
<b>Pragmatikai jellemzők</b>	Gyakran fordul elő elbeszélésekben, történetekben.

**pedig (2)**

<b>Kollokációk</b>	Nincs egyértelműen a szóhoz társítható kollokáció. Negatív vagy korlátozó értelmű szó jelenléte az első vagy a második tagmondatban, pl. <i>nem, ritkán</i>
<b>Kolligációk (Grammatikai jellemzők)</b>	<b>pedig + második tagmondat</b> A <i>pedig</i> szó a második tagmondat első eleme.
<b>Szemantikai jellemzők</b>	Az első tagmondat gyakran kudarcot, negatív eredményt fejez ki, amelyet a második tagmondat állításának nem várt következménye : <i>Soha nem tudtam szavakat magolni, pedig szeretem a nyelveket. Nem sikerült megtanulni angolul, pedig néztem sok Spongya Bobot.</i>
<b>Pragmatikai jellemzők</b>	A beszélőnek egy eseménnyel kapcsolatos – általában negatív – attitűdjét (frusztrációját) fejezi ki.

6. ábra: A 'pedig' kötőszó kétféle profilja

Az 5.1. és 5.2. pont összegzéseként elmondhatjuk, hogy – akár lexikai, akár nyelvtani kérdésből indulunk ki – a pedagógiai korpusz igen hasznos segítséget nyújthat az adott nyelvi elem jelentésének feltárásában. Különösen igaz ez a gyakori szavak esetében, amelyek használatára még egy kisebb korpuszban is számos példát ad, így ezek többféle szövegkörnyezetben megfigyelhetők. Ha a tanulók megfigyeléseik eredményét a fenti táblázathoz hasonló módon rögzítik, részletes áttekintést kapnak a kiválasztott elem viselkedéséről, használatáról. Nagy előny, hogy a legtöbb kérdés megválaszolásához ugyanazt a módszert alkalmazhatják, így idővel nagy jártasságot szerezhetnek az eszközök használatában.

Nagyon fontosnak tartjuk azonban megjegyezni, hogy a korpuszvizsgálatnak nem az efféle rendszerező táblázatok a legfontosabb hozadékai. Végeredményben ezek aligha adnak több információt, mint egy jó szótár. A korpuszelemzés lényege maga a korpusz vizsgálata, azaz **nagyszámú, természetes nyelvhasználatot tükröző példa megfigyelése, a választott nyelvi elem autentikus vagy félautentikus szövegekben való tanulmányozása**. A létrehozott szóprofil tulajdonképpen nem más, mint emlékeztető, amely a megfigyelés és a rendszerezés

fázisait zárja le. Ám az ezt megelőző lépések sokkal lényegesebbek: ezekben a nyelvtanuló számos mondatot figyel meg a választott szóval, így empirikus módon, konkrét példákon keresztül gyarapítja (intuitív és tudatos módon is) a többszavas egységekről és általában a célnyelvről való tudását. A példákából kirajzolódó ismétlődések pedig lehetővé teszik a különféle jellemzők feltérképezését.

## 6. A nyílt korpusz felhasználásával készíthető feladatok

Az 5.1. és 5.2. pontban arra láttunk példát, hogyan használhatjuk a nyílt korpuszt lexikai és grammatikai jellegű kérdések megválaszolására. Az alábbiakban olyan feladatokra mutatunk néhány példát, amelyek a korpuszban található mondatokra épülnek. A mondatokat a *Concordancer*ből tölthetjük le.<sup>5</sup> A javasolt feladatok a választott nyelvi elem és szövegkörnyezetének megfigyelésére, a többszavas egységek elmélyítésére szolgálnak. Maguk a feladattípusok nem újak, az újdonság inkább abban rejlik, hogy a feladatokban – akárcsak a konkordanciasorok esetében – a gyakorlandó elem többféle tipikus szövegkörnyezetben jelenik meg.

### 6.1. Egy szó, több mondat

Eszköz: *Concordancer*

Ez a feladat a fent bemutatott *pedig* kötőszó két jelentését és a jelentésekhez kapcsolódó szórendeket gyakoroltatja. A tanulóknak a *pedig* szót kell beírniuk a megfelelő helyre a mondatban (7. ábra). Efféle munkalapok percek alatt generálhatók, ha a mondatokat a korpuszból töltjük le. Mindössze a gyakorlandó szót vagy többszavas egységet kell belőle kitörölni.

Helyezze a *pedig* szót a mondatba! Ügyeljen a szórend és a jelentés kapcsolatára!

1. Este koncerten voltunk, napközben csak pihentünk.
2. Amikor én fiatal voltam, még nem lehetett utazgatni. De szerettem volna!
3. Mindig meglepődöm, mennyi ember elmegy Albert előadásaira. Unalmasan ad elő, ráadásul nem artikulál.
4. Mindjárt elalszom, még tanulnom kell.
5. Erdély és a Balkán irányába innen indultak a vonatok. A Nyugati pályaudvarról Bécs és Párizs felé.

7. ábra: Feladat a 'pedig' kötőszót tartalmazó mondatok szórendjének gyakorlására a nyílt korpuszból vett példákkal

Miután a tanulók kitöltötték és ellenőrizték a mondatokat, a következő lépésben kreatívan használhatják azok elemeit. Így megkérhetjük őket, hogy a második tagmondatot (a teljes mondatot vagy annak egy elemét) változtassák meg. Például: „Este koncerten voltunk,

<sup>5</sup> A *Concordancer*ben lekérdezett mondatokat a jobb felső sarokban sorakozó ikonok közül a kék nyílra kattintva tölthetjük le.

napközben pedig *sétálni*.” Vagy: „Este koncerten voltunk, *délután pedig egy múzeumban*.” Megadhatjuk, hány mondatot várunk, vagy megszabhatjuk a diákok rendelkezésére álló időt.

### 6.2. Keresse meg a hiányzó szót!

Eszköz: *Concordancer*

Ezt a feladattípust Johns (1991) javasolta, bizonyítandó, hogy a hagyományos feladatok hatékonysága növelhető, ha a példákat korpuszból vesszük, mivel ezek a természetes nyelvhasználatot tükrözik. A pedagógiai korpusz mondatai ugyan nem feltétlenül teljesen autentikusak, a természetes nyelvhasználat megőrzése azonban itt is elsődleges cél. A feladat mondataiból mindig ugyanaz a szó hiányzik, amelyet a diákok a korpusz segítségével (vagy enélkül) egészítenek ki (8. ábra).

Melyik szó hiányzik minden mondatból? A megoldáshoz használhatja a *MagyarOK A1* korpuszt.

1. A székem egy nagyon \_\_\_\_\_ fitneszlabda.
2. A szobám nem túl \_\_\_\_\_ : 10–12 négyzetméter.
3. A lányok mániákus ember, a szobájában is \_\_\_\_\_ a rend.
4. Kaposvár nem \_\_\_\_\_ város: hatvanezer ember él itt.
5. Gyalog járok az irodába. Ez Budapesten \_\_\_\_\_ luxus!

8. ábra: Feladat a 'nagy' melléknév bemutatására különböző szövegek környezetben, korpuszból vett példákkal (A mondatok száma tetszés szerint növelhető.)

Ilyen feladatokat a diákok is készíthetnek a csoportjuknak. Ehhez mindenki kiválaszt egy vagy több szót, amely nehézséget okoz számára, és a korpusz segítségével összeállít belőlük egy hasonló feladatlapot. A diákok a tanórán egymás feladatait oldják meg.

A saját feladatlap elkészítése és mások feladatlapjának kitöltése világossá teszi a szűkebb szövegek környezet és a mondatokból hiányzó lexikai elem közötti szoros kapcsolatot (Johns 1991a és b). Arra is megkérhetjük őket, jelöljék meg azokat a mondatokat, amelyek fordítása eltér, és azokat is, amelyeket ugyan értenek, de segítség nélkül nem tudnák megalkotni őket. Ez a lépés tudatosítja, hogy minden nyelvnek megvan a maga idiomaticitása, valamint, hogy a receptív és produktív készségek között eltérések lehetnek (a többszavas egységeket általában könnyebb megérteni, mint megalkotni; vö. Boers 2021).

### 6.3. Szókincsisméltés

Eszköz: *Wordlist*

Az alábbi feladat azt mutatja be, hogyan használható a pedagógiai korpusz szókincs átisméltésére. A választott alkorpusz(ok) vagy a teljes korpusz szókincsének megjelenítésére a *Wordlist* eszköz használható. Ha például a tanulók az A1-es alkorpusz leggyakoribb mellékneveit szeretnék átisméltetni, a *Wordlisttel* az alábbi listát generálhatják (9. ábra):

MagyarOK A1-B1 × adjective (2,212 items | 8,786 total frequency)

Lemma	Frequency ?	Lemma	Frequency ?	Lemma	Frequency ?	Lemma	Frequency ?	Lemma	Frequency ?
1 jó	335 ...	11 egy	91 ...	21 éves	61 ...	31 kevés	43 ...	41 helyi	35 ...
2 olyan	246 ...	12 fontos	86 ...	22 angol	60 ...	32 késő	43 ...	42 kényelmes	35 ...
3 magyar	239 ...	13 egyik	82 ...	23 magas	59 ...	33 igaz	41 ...	43 csendes	34 ...
4 amilyen	226 ...	14 régi	76 ...	24 másik	57 ...	34 drága	40 ...	44 rengeteg	34 ...
5 kicsi	196 ...	15 német	71 ...	25 nehéz	55 ...	35 szíves	39 ...	45 amerikai	31 ...
6 új	191 ...	16 érdekes	69 ...	26 gyönyörű	53 ...	36 szinonim	39 ...	46 modern	31 ...
7 nagy	183 ...	17 kedvenc	66 ...	27 fáradt	49 ...	37 hosszú	38 ...	47 olasz	31 ...
8 szép	123 ...	18 ilyen	65 ...	28 tilos	46 ...	38 francia	36 ...	48 biztos	30 ...
9 kis	102 ...	19 friss	63 ...	29 közösségi	44 ...	39 rossz	36 ...	49 európai	30 ...
10 kedves	95 ...	20 finom	62 ...	30 könnyű	43 ...	40 szuper	36 ...	50 türelmes	29 ...

9. ábra: Az ötven leggyakoribb melléknév a MagyarOK A1 korpuszban

A fenti fontos melléknévek még ebben a korlátozott méretű korpuszban is figyelemre méltó gyakorisággal fordulnak elő, ami számos lehetőséget kínál szövegtérületük megfigyelésére. A tanulók házi feladatként átismételhetik a melléknévek használatát, mindegyikkel átlagosan legalább tizenöt példát megfigyelve (a listán a három pontra kattintva a *Word Sketch* és a *Concordancer* eszköz is előhívható). A *Concordancer* előnye, a tananyagban egymástól távolabb eső előfordulásokat mutat meg egy helyen, vagyis azokat a példákat, amelyekkel a diák különféle szövegekben találkozott, egymás alatt vonultatja fel, lehetőséget adva a szerkezeti és szókincsbeli ismétlődések megfigyelésére. Mivel a szókincselsajátítás „kumulatív folyamat” (pl. Nation 2013: 119, 127), az ismétlés módszereinek a rendszerezett korpuszpéldák megfigyelésével való kiegészítése fokozottan segítheti a bevésést (Hoey 2005).

## 7. A korpuszban található szövegek mint nyelvi modellek

A szövegalkotás igen hatékony módja lehet a tudás rendszerezésének (Nation 2013; Boers 2021). Ezért a *MagyarOK* tankönyvek mindegyik fejezete egy olyan feladattal zárul, amelyben a tanulók a fejezetben tárgyalt témával kapcsolatban megírhatják saját szövegüket. Ennek a feladatnak a megoldásában is segíthet a nyílt korpusz, amely szövegeinek egy része nyelvi modellként is használható. Az A1-es és az A2-es tankönyvek zárófeladataihoz készített modellszövegek már részét képezik a nyílt korpusznak, a B1-es szövegek még feldolgozás alatt állnak (Szita 2020: 175–176). A modellszövegek azt mutatják meg, hogyan lehet természetesen hangzó, koherens szöveget alkotni, amely – az adott témával kapcsolatosan – személyes tartalmakat közvetít (Kennedy–Miceli 2010, 2017).

Mielőtt megalkotnák saját szövegeiket, a tanulók elolvashatják a korpuszban található modellszövegeket, vagy ezekben különböző elemek használatára kereshetnek rá.<sup>6</sup> A modellszövegekben megjelenő leggyakoribb melléknévek, igék vagy főnevek rendszerezésére használhatják az 5.3. pontban bemutatott *Wordlist* eszközt. A *Word Sketch* és a *Concordancer* segítségével pedig megfigyelhetik ezek használatát többszavas egységekben és teljes mondatokban, illetve szövegrészletekben. Ugyanakkor a szövegek egészét is tanulmányozhatják, hiszen ez teszi láthatóvá a szerkezeti és szórendi ismétlődéseket és variációkat, valamint a szövegkoherenciát biztosító elemek használatát. A modellszövegekből

<sup>6</sup> Ehhez először bármelyik eszközben a *File name* funkcióra kell rákattintani, majd kiválasztani a fejezethez tartozó mintaszövegeket, például az A1-es könyv negyedik fejezetéhez tartozó szövegeket a MagyarOK\_A1\_4fejezet\_mintaszovegek fájlban találjuk.

a tanuló annyi szót és kifejezést „hasznosít újra”, amennyit csak akar. Az egyetlen kritérium, hogy a végeredmény természetesen hangzó, logikusan felépített szöveg legyen. A diáknak tehát nem az a feladata, hogy „nulláról alkossa meg” a szövegét, hanem az, hogy a modellek segítségével jól olvasható, koherens szöveget hozzon létre, amelyben a tanult szókincs nagy részét természetes szövegek környezetben használja.

Álljon itt egy példa a „Lakóhelyem” (MagyarOK A1, 4. fejezet) témában írt zárószövegre, amelyet a diák a modellszövegek alapján készített el. A szövegekből átvett többszavas egységeket vastagon szedtük.

*A városom fantasztikus. Franciaországban, **Strasbourgban élek**. Az egyetemen kínaiul és magyarul tanulok. Strasbourgban **körülbelül négyszáz ezer ember él**. **Természetesen nem a nagyváros**.*

***Nagyon szép a város. Jó a közbiztonság, magas az életszínvonal és vannak kulturális és sportolási lehetőségek is.***

*Központban vannak **szép és régi épületek, modern könyvtárok, híres múzeumok, nagy mozik és színházak is**. A közlekedés jó, praktikus de egy kicsit drága. A városban **nem járok autóval. Általában villamossal és gyalog közlekedem**.*

*A kollégiumban lakom, a kollégium elég modern és világos. **Egy csendes utcában lakom, de sok diák lakik is itt. Szerintem Strasbourg nagyon érdekes és hangulatos, mert a városban sok olcsó és népszerű bár, kocsmá és étterem van.***

*Nagyon tetszik Strasbourgi kávézók. Itt lehet könyvet olvasni, kávé inni, pogácsát enni és e-mailt írni. A városban **kitűnő cukrászdák** vannak is. Strasbourg **nagyon biztonságos város, az emberek kedvesek, barátságosak és nyugodtak**.*

*Strasbourgban **nincsenek magas hegyek, tengerek, tók vagy nagy erdők**. Tél elég hideg is, de mégis **nagyon szeretem itt élni**.*

Ahhoz, hogy a feladat hatékonyan segítse a diákok nyelvi fejlődését (és ne egyszerű kivágásról és beillesztésről legyen szó), érdemes egy rutinszerű munkafolyamatot követni, amelynek során az egyes lépéseinek hasznáival a tanuló maga is tisztában van. Ez az alábbi lépésekből állhat:

- (1) A diákok elolvassák a mintaszövegeket, majd kijegyzetelik a hasznos kifejezéseket – például azokat, amelyek rájuk is igazak. A nyílt korpusz eszköztárának segítségével többszavas egységeket is gyűjtenek, és megfigyelik a választott szavak használatát.
- (2) A diákok a modellszövegek és többszavas egységek felhasználásával megalkotják saját szövegüket.
- (3) A tanár kijavítja a szöveget, javításait lehetőség szerint el is magyarázza.
- (4) A diák a javított szöveget kézzel lemásolja, hogy a helyes kifejezések kinetikus úton is rögzüljenek.
- (5) A következő órába beépítünk legalább egy olyan lépést, amelyben a szövegekkel „történik valami” (ld. alább).

Az első lépésben a tanuló tehát megfigyeli az anyanyelvi beszélők kifejezésmódját, és megkeresi azokat a kifejezéseket, amelyeket a saját szövegében felhasználhat. Nagyon fontosnak tartjuk, hogy ne ösztönözzük a diákokat A1–A2-es szinten arra, hogy „saját szavaikkal” fogalmazzák meg mondanivalójukat, hiszen kezdő nyelvtanulóként még nincsenek saját szavaik, azaz még nem áll rendelkezésükre elegendő nyelvi elem, amelyből tudatosan válogathatnának. Modell nélkül valószínűleg az anyanyelvükre fognak támaszkodni vagy egyes szavakat egymás után felsorakoztatva igyekeznek mondatokat alkotni, amelyek sem tipikusak, sem tökéletesen érthetők nem lesznek. Vagyis nyelvtanilag gyenge minőségű

szöveget hoznak létre, amelynek megalkotása során igen keveset tanultak, sőt félő, hogy a folyamatban még néhány hibás alakot is bevéstek (Boers 2021).

A modellszövegek használatának előnye, hogy az ezekben megtalálható többszavas egységeknek többségét a tanuló már ismeri; így amikor a szövegébe beépíti ezeket az elemeket, és koherens elbeszéléssé rendezi, átismétli és elmélyíti őket. Ennélfogva az önálló szövegalkotás fázisában többnyire helyes nyelvi megnyilvánulásokat hoz létre, vagyis ahelyett, hogy szabadon kísérletezne a nyelvvel, ám közben hibás fordulatokat is rögzítene, a természetes nyelvhasználatot gyakorolja. Ha a javított szöveget a tanuló kézírással le is másolja, annak nyelvi elemei még inkább bevéődnek. S ha a diákok szövegeivel a következő órán történik is valami (például ezek a szövegek adják egy beszélgetés, egy interjú vagy egy kiselőadás vázát), a szókincs és szerkezetek még inkább elmélyülnek, ugyanakkor az íráskészség után a beszéd-készséget és a szövegértést is fejlesztjük.<sup>7</sup>

## 8. Összegzés

Írásomban a *MagyarOK* nyílt korpusz néhány használati lehetőségét mutattam be. Kitértem annak ismertetésére, hogy a pedagógiai korpuszok nem követhetik mindenben a nyelvészeti korpuszok építésének kritériumait, hiszen ez a kétfajta korpusz nem ugyanazzal a céllal készül. A pedagógiai korpusz anyagának összehangban kell állnia a tanuló nyelvi szintjével és tematikájában relevánsnak kell lennie. Ahhoz, hogy a tanuló a korpuszt hatékonyan tudja használni, a szövegeknek a természetes nyelvhasználatot kell közvetíteniük, ám indokolt esetben nyelvezetük lehet némileg módosított.

A bemutatott feladatok azt mutatják, hogyan járulhat hozzá a korpuszhasználat a nyelvi ismeretek megszilárdításához és a nyelvi készségek fejlesztéséhez. A feladatok egy része a korpuszban egyrészt válaszokat kereshet a nyelvvel kapcsolatos kérdésekre. Ezek a feladatok tudatosítják a többszavas egységek fontosságát, megtanítják felismerni és rendszerezni a különféle nyelvi elemek jellemzőit. A feladatok másik része a korpusz mondatait használja. Ezek során a korpuszhasználat előnye a feladat mondatainak minőségében mutatkozik meg, hiszen a tanulók – a tanár által kifejezetten a gyakorlás céljára megalkotott mondatok helyett – számos, a természetes nyelvhasználatot tükröző példával dolgoznak. Végül a korpusz modellszövegek gyűjteményeként is felfogható, amelyek nyelvi elemeit a tanuló saját szövegeibe is beépítheti. A modellszövegek abban segítenek, hogy a tanuló koherens, jól olvasható, gazdag szókincset tartalmazó szövegeket alkosson, s a megfigyelésnek és az aktív használatnak köszönhetően minél több hasznos szerkezetet és fordulatot bevessen. A feladat hatékonyságának érdekében nagyon fontos a szövegekkel való aktív munka.

Összegzésül elmondhatjuk, hogy a *MagyarOK* nyílt pedagógiai korpusz használata az A1–B1-es szinten többféle módon hozzájárulhat a fontos nyelvi elemek megfigyeléséhez, elsajátításához és a nyelvi tudatosság növeléséhez.

## Irodalom

- Allan, Rachel 2009. Can a graded reader corpus provide ‘authentic’ input? *ELT Journal* 63/1: 23–32.  
 André, Virginie 2020. Corpus d’interactions et apprentissage du français langue étrangère. In: Benzitoun C. – Rebuschi M. (eds.): *Les corpus en sciences humaines et sociales*. Presses Universitaires de Nancy. 101–121.  
 Aston, Guy 2001. *Learning with Corpora*. Athelstan.  
 Boers, Frans 2021. *Evaluating second language vocabulary and grammar instruction*. Routledge.

<sup>7</sup> A nyílt korpusz a nyelvi ismétlődés és variáció megfigyelésére is kiválóan alkalmas, de ennek taglalására – terjedelmi okokból – ebben a cikkben nincs módunk.

- Boulton, Alex 2010. Consultation de corpus et styles d'apprentissage. *Cahiers de l'APLIUT* 29/1: 98–115.
- Boulton, Alex – Tyne, Henry 2014. *Des documents authentiques aux corpus. Collection langues et didactique.* Les Éditions Didier, Paris.
- Braun, Sabine 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL* 19/3: 307–328.
- Braun, Sabine 2010. Getting past Groundhog day. Spoken multimedia corpora for student-centred corpus exploration. In: Harris T. – Moleno Jaén, M. (eds.): *Corpus linguistics in language teaching.* Peter Lang, 75–98.
- Cavalla, Crystelle 2019. Corpus numériques: critères pour l'enseignement des langues. *PERL : Entre présence et distance. Enseigner et apprendre les langues à l'université à l'ère numérique.* Equipe PERL: Paris.
- Chambers, Angela 2019. Towards the corpus revolution? Bridging the research-practice gap. *Language Teaching*, 52/4: 460–475.
- Chambers, Angela – Farr, Fiona – O'Riordan, Stéphanie 2011. Language teachers with corpora in mind: From starting steps to walking tall. *Language Learning Journal* 39/1: 85–104.
- Flowerdew, John 2009. Corpora in language teaching. In: Long, H. – Doughty, C. J. (eds.): *The handbook of language teaching.* Blackwell. 327–351.
- Hoey, Michael 2005. *Lexical priming: A new theory of words and language.* Routledge.
- Hunston, Susan 2008. Starting with the small words. *International Journal of Corpus Linguistics* 13/3: 271–295.
- Hunston, Susan 2010. How can a corpus be used to explore patterns? In: O'Keeffe, A. – McCarthy, M. (eds.): *The Routledge handbook of corpus linguistics.* Routledge. 152–167.
- Hunston, Susan 2015. Lexical grammar. In: Biber, D. – Reppen, R. (eds.): *The Cambridge handbook of English corpus linguistics.* Cambridge University Press. 201–216.
- Johns, Tim 1991a. Should you be persuaded: Two samples of data-driven learning materials, *English Language Research Journal* 4: 1–16.
- Johns, Tim 1991b. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In: Odlin, T. (ed.): *Perspectives on pedagogical grammar.* Cambridge University Press. 293–313.
- Kennedy, Claire – Miceli, Tiziana 2010. Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology* 14/1: 28–44.
- Kennedy, Claire – Miceli, Tiziana 2017. Cultivating effective corpus use by language learners. *Computer Assisted Language Learning* 30/1–2: 1–24.
- MagyarOK open corpus, Sketch Engine:  
<https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok>
- Martinez, Ron – Schmitt, Norbert 2012. A phrasal expression list. *Applied Linguistics* 33/3: 299–320.
- Martinez, Ron – Schmitt, Norbert 2015. Vocabulary. In: Biber, D. – Reppen, R. (eds.): *The Cambridge handbook of English corpus linguistics.* Cambridge University Press. 439–460.
- Meunier, Fanny – Reppen, Randi 2015. Corpus versus non-corpus-informed pedagogical materials: Grammar as the focus. In: Biber, D. – Reppen, Randi (eds.): *The Cambridge handbook of English corpus linguistics.* Cambridge University Press. 498–514.
- Nation, Paul 2013. *Learning Vocabulary in Another Language.* Cambridge University Press.
- O'Keeffe, Anne – McCarthy, Michael – Carter, Ronald 2007. *From corpus to classroom.* Cambridge University Press.
- Omidian, Taha – Siyanova-Chanturia, Anna 2020. Semantic prosody revisited. Implications for language learning. *TESOL Quarterly* 54/2: 1–13.
- Poole, Robert 2018. *A guide to using corpora for English language learners.* Edinburgh University Press.
- Salazar, Danica 2014. *Lexical bundles in native and non-native scientific writing.* John Benjamins.
- Sinclair, John 1991. *Corpus, concordance, collocation.* Oxford University Press.
- Sinclair, John 2003. *Reading concordances. An introduction.* Pearson/Longman.
- Sinclair, John 2004. *Trust the text.* Routledge.
- Stefanowitsch, Anatol 2020. *Corpus linguistics. A guide to the methodology.* Language Science Press.
- Stubbs, Michael 2009. The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics* 30/1: 115–137.
- Szita Szilvia 2020. Korpuszépítés és korpuszhasználat alacsonyabb nyelvtudási szinteken. *Hungarológiai Évkönyv* 21/1–2: 173–179.
- Szita Szilvia 2017. Ön most eljött vagy megjött? Korpuszhasználat a magyar mint idegen nyelv tanításában. *Hungarológiai Évkönyv* 18/1: 81–93.
- Szita Szilvia 2018. Annak tűnik, aminek látszik? *Modern Nyelvoktatás* 2018 24/1: 31–47.
- Szita Szilvia – Pelcz Katalin 2017. Modellalapú nyelvoktatás, természetes nyelvhasználat. *THL2* 1(2), 262–269.
- Szita Szilvia – Pelcz Katalin 2013–2019. *MagyarOK A1–B2.* Pécsi Tudományegyetem.



- Szudarski, Pawel 2017. *Corpus linguistics for vocabulary*. Routledge.
- Tyne, Henry 2012. Corpus work with ordinary teachers: Data-driven learning activities. In: Thomas, J. – Boulton, A. (eds.): *Input, process and product: Developments in teaching and language corpora*. Masaryk University Press. 114–129.
- Widdowson, Henry 2003. *Defining issues in English language teaching*. Oxford University Press.

*Szita, Szilvia*

**On the use of the *MagyarOK* open pedagogical corpus**

The article presents the open pedagogical corpus called *MagyarOK A1–B2*, which was built for the Hungarian as a Second Language coursebook series with the same title, to enhance the teaching and learning of the language by providing (semi-)authentic texts so that users can study the actual language use of native speakers in specific communicative contexts and situations, and to provide language patterns and models which learners can integrate into their own linguistic output. The paper first gives a brief overview of the general criteria for pedagogical corpora and presents the contents of the corpus *MagyarOK*, then it proposes tasks for data-driven learning based on the corpus. It is suggested, on the one hand, that data-driven tasks can contribute in several ways to language acquisition through noticing, analyzing, and awareness-raising. On the other hand, active interaction with the corpus material can help learners memorize various language patterns.