

Szita Szilvia

Korpuszépítés és korpuszhasználat alacsonyabb nyelvtudási szinteken

1. Bevezetés

Nem csak a magyar mint idegen nyelv oktatásában kevésbé bevett gyakorlat a nyelvi korpuszok használata. Bár ezeknek az óriási szöveggyűjteményeknek a használata számos előnnyel jár, melyekből e cikk – terjedelme okán – csupán néhányat tud felvillantani, csak igen kevés tanár és diák dolgozik velük rendszeresen. Ennek oka leginkább abban keresendő, hogy a jelenlegi korpuszok (magyarból ugyanúgy, mint más nyelvekből) a nyelvészek elvárásainak és céljainak szem előtt tartásával készültek. Ebben a cikkben amellet kívánok érvelni, hogy amennyiben a korpuszhasználatot a nyelvtanítás számára is relevánssá és hatékonnyá kívánjuk tenni, akkor a nyelvtanuló igényeinek megfelelően megépített, úgynevezett pedagógiai korpuszokra van szükség. Egy ilyen pedagógiai adatbázis a folyamatosan bővülő *MagyarOK* korpusz, amely a Sketch Engine gyűjteményének nyilvános korpuszai között érhető el. A cikk második felében a korpusz tartalmát és a korpuszépítés háttérében álló módszertani megfontolásokat ismertetem röviden.

2. Eltérő célok, más-más korpusz

Az interneten nyilvánosan is elérhető nagy nyelvi adatbázisok – korpuszok – száma egyre gyarapszik. Ezek a „valóban használt nyelvet tartalmazó gyűjtemények” (Brazil 1995: 24) lehetőséget adnak arra, hogy megfigyeljük, hogyan használják az anyanyelvi beszélők a nyelvéket, mely elemek (lexikai egységek, grammatikai és pragmatikai megoldások) kerülnek előtérbe adott közléshelyzetben (pl. Hunston 2009: 145). Ezek a korpuszok általában több millió, sőt nem ritkán több milliárd szót tartalmaznak; az adatok nagy mennyisége a reprezentativitást hivatott szolgálni (Brezina 2018). Ilyen nyilvános óriáskorpusz magyar nyelvből a *Magyar Nemzeti Szövegtár* és a *huTenTen12* korpusz a Sketch Engine-en; az előbbi gyűjtemény leginkább a sajtó, a szépirodalom és a közösségi médiák anyagából merít, ez utóbbi pedig az interneten 2012-ig magyarul publikált oldalakat tartalmazza. Az efféle hatalmas nyelvi adatbázisok a nyelvészek igényeit kiszolgálják, a nyelvtanulók és a nyelvtanárok azonban a – főleg angol nyelvterületen – kidolgozott módszertani javaslatok ellenére (ld. pl. Boulton, 2017, Boulton–Thomas 2012) gyakran ódzkodnak használatuktól. Mi ennek az oka?

A nyelvész és a nyelvtanár, illetve -tanuló céljai nem ugyanazok a korpuszhasználat során (Boulton–Thomas 2012: 7; Römer 2006). A nyelvész az általa tanulmányozott nyelvről vagy nyelvváltozatról kíván empirikus úton minél teljesebb tudást szerezni, hogy ezáltal pontosabb, megbízhatóbb leírást készíthessen. Míg ő a korpuszok segítségével hipotéziseinek érvényét vizsgálja, a tanár és a diák elsősorban a szavak jelentésével és használatával kapcsolatos konkrét kérdésekre keres(ne) – szintén empirikus úton – választ. Vagyis a nyelvtanuló célja, ellentétben a nyelvessel, nem a nyelv leírása, hanem *az anyanyelvi beszélők nyelvhasználatának megfigyelése* a kompetens nyelvhasználóvá válás érdekében, a nyelvtanáré pedig, hogy a diákot ebben a folyamatban segítse. Mivel a két csoportot – nyelvész, tanár és diák – más-más célok motiválják a korpuszhasználatára, felmerül a kérdés, hogy vajon a nyelvelemzési célokra megépített korpuszok valóban kielégítik-e a tanulók és tanárok igényeit vagy új, más alapelvek mentén létrehozott korpuszokra van szükség.

A kifejezetten tanítási célra készülő, úgynevezett *pedagógiai korpuszok* létrehozásának szükségességét többek között Willis (2003: 16) és Braun is (2006, 2007) hangoztatja. Braun legfontosabb érve, hogy „[a] legtöbb nyilvános korpuszt a nyelvészeti kutatások érdekében, nem pedig pedagógiai megfontolások alapján hozták létre, így ezek *sem formailag, sem pedig tartalmilag nem felelnek meg a nyelvtanulók szükségleteinek*” (Braun 2007: 308, kiemelés a szerzőtől). A létező nagy nyelvi korpuszok egyrészt nehezen kezelhetők, másrészt a bennük rögzített megnyilvánulások sokszor még anyanyelvi beszélők számára is nehezen kontextualizálhatók (Widdowson 2003), valamint olyannyira komplexek, hogy a diáknak – különösen alacsonyabb nyelvtudási szinteken – nem igazán nyújtanak segítséget egy-egy problémás nyelvi jelenség feltérképezéséhez. Braun szerint ez lehet a magyarázata, hogy a tanárok és diákok körében a korpuszhasználat a mai napig nem terjedt el. Új, a tanulók szükségleteinek figyelembevételével összeállított korpuszokra lenne tehát szükség, amelyek leginkább azzal járulhatnak hozzá a nyelvi kompetencia fejlesztéséhez, hogy a diák számára *érthető és releváns nyelvi inputot* tartalmaznak (Willis 2003, Braun 2006). Az efféle korpuszok ugyanakkor hidat építhetnek a tanórai környezetben használt nyelv és a külvilág “cenzúrázatlan” nyelvhasználatára között.

Claire Kennedy és Tiziana Miceli 500 000 szavas nyelvi korpusza ezt a célt kívánja megvalósítani (2010, 2017). A két nyelvtanár-kutató olasz nyelvű, tematikus korpusza a B1–B2 szintre készült; a gyűjtemény szövegei olyan *adaptálható nyelvi modelleket tartalmaznak*, amelyek felhasználása a diákok nyelvi produktumainak minőségét javíthatja. A korpusz szavak és témák szerint is kereshető, és a diákok a konkordanciasorok mellett a teljes szövegekhez is hozzáférhetnek. Ennek a pedagógiai korpusznak tehát nem az az elsődleges célja, hogy az anyanyelvi beszélők nyelvhasználatát reprezentatív módon tükrözze – bár a megfelelő paraméterek mentén összeállított pedagógiai korpusz akár reprezentatív is lehet (Szita megjelenés előtt) –, hanem hogy a *természetes nyelvhasználat egyes szegmenseit mutassa be a diák számára is érthető példákon*. Hasonló célkitűzések jegyében született meg a MagyarOK nyilvános pedagógiai korpusz is, amelyet az alábbiakban mutatunk be.

3. A MagyarOK korpusz

A MagyarOK nyilvános írott nyelvi korpuszában összesen mintegy 210 000 token található. A gyűjtemény három alkorpuszból áll, melyek közül az első a tankönyvek teljes szövegét, azaz összesen körülbelül 150 000 tokent tartalmaz. A korpuszban kitöltött feladatok is szerepelnek, grammatikai feladatoknál a könnyebb eligazodás érdekében a számozást is megtartottuk. Az adatbázisban nyelvtudási szintek és tankönyvi fejezetek szerint is lehet keresni. A második alkorpuszban félautentikus narratív szövegeket találunk, amelyeket anyanyelvi beszélők alkottak a tankönyvben tárgyalt témákról¹. Ebben az adatbázisban mintegy 61 000 token található. A harmadik – egyelőre szerkesztés alatt álló – alkorpuszt pedig teljes mértékben autentikus szövegek – újságcikkek, blog- és fórumbejegyzések valamint személyes írásos anyagok (cset, e-mail) – alkotják majd. A szövegeket témájuk és nehézségi szintjük alapján válogattunk, a készülő adatbázis pillanatnyilag 500 000 tokent tartalmaz. A következőkben a két, már most is elérhető alkorpusz felhasználási lehetőségeivel foglalkozom részletesebben.

A MagyarOK tankönyvek készítését annak érdekében, hogy biztos alapot adhassanak az autentikus közléshelyzetekben való részvételhez, korpuszvizsgálatok előzték meg (erről bővebben ld. Szita 2014 és Szita megjelenés előtt). Ezeknek a korpuszinformált tankönyveknek az anyaga – vagyis félautentikusnak tekinthető nyelvi input – képezi az első alkorpuszt. A gondolat, miszerint *a tanórán feldolgozott szövegekből érdemes pedagógiai korpuszokat építeni*, nem új a szakirodalomban (ld. pl. Aston 2001; Braun 2006; Flowerdew 2009; Timmis

¹ Ezek hamarosan külön kiadvány formájában, feladatok kíséretében is elérhetőek lesznek.

2015; Tyne 2012; Widdowson 2003; Willis 2003). Timmis (2015: 3) Willist idézve azt javasolja, hogy a nyelvtanulási célokra összeállított korpuszban „azok a szövegek szerepeljenek, amellyel a diák a nyelvtanulás során találkozott”, mert a szövegkörnyezet ismerete sokat segít a konkordanciasorok elemzésében. Az efféle gyűjtemények másik előnye, hogy csupa olyan megnyilvánulást tartalmaznak, amelyek megfelelnek a diák nyelvtudási szintjének (Aston 2001).

A modellalapú nyelvtanítás kézikönyvében Pelcz Katalinnal arról írunk (Szita és Pelcz, kézirat), hogy a természetes nyelvhasználat elsajátításához – többek között – nagy nyelvi inputra is szükség van. Ez nem azt jelenti, hogy a diákok minden egyes órán sok-sok új szót tanulnak, hanem hogy a már ismert elemeket újabb és újabb kontextusban látják, újabb és újabb jelentéseket társítanak hozzájuk. (Példa erre a *Jól vagy?* mint kérdés és a *jól van* kifejezés mint a *rendben van* szinonimája, amelyek ugyanazon lemmákból építkeznek, de a jelentésük más.) Ez a meg gondolás vezérelt abban, hogy a korpuszba *modellszövegeket* is felvegyünk. Ezek a szövegek a fejezetek összefoglaló feladatára készült variációk, amelyeket tíz, a *MagyarOK* tankönyvcsaládot használó kolléga írt. Őket arra kértük, hogy a záró feladat kérdései alapján, a tankönyv szóanyagából minél többet felhasználva alkossák meg saját szövegeiket, ugyanakkor ne „butítsák le” a nyelvhasználatukat, hanem kifejezésmódjuk maradjon természetes². A kérés háttérében az a meg gondolás állt, hogy az egy-egy fejezethez készült szövegeknek megfelelő mennyiségű nyelvi ismétlődést és variációt kell tartalmazniuk ahhoz, hogy a gyakran használt nyelvi elemek hatékony rögzítéséhez hozzájárulhassanak (Taylor 2017, Hoey 2005). Vagyis abból indultunk ki, hogy kellő számú modell tanulmányozása esetén a tanuló nem két-három gyakori kifejezést fog szó szerint ismételtetni saját szövegében, hanem a megtanult elemeket kreatívan hasznosítja majd, hiszen használati sajátosságait – legalább részben – elsajátította.

Az A1-es alkorpuszban 80 – fejezetenként tíz-tíz – modellszöveg található. A szövegek átlagosan 150 szóból állnak, a teljes alkorpuszban így 12 000 token található. Az A2-es alkorpusz 80 180–250 szó hosszúságú szöveget tartalmaz, ezek összesen 16 000 tokent tesznek ki. A B1-es alkorpusz pedig 120 darab 250–300 szavas szöveget, azaz 33 600 szót tartalmaz.

A szövegek kellően rövidek ahhoz, hogy a diák teljes egészében elolvashassa őket, és elsődleges funkciójuk a szókincs rendszerezése és bővítése a már említett ismétlődések és variációk által. Összességük *komplex modelleket ad arra, amit a diák a nyelvtanulás egy-egy fázisában az adott témával kapcsolatban már el tud mondani*, ezenkívül a szövegszerkesztésre, a szövegkohéziós elemek használatára is számos példát nyújtanak. Pusztán a munkamenet szemléltetése céljából álljon itt kettő az A1-es könyv 2. fejezetének záró feladatához készült szövegek közül!

A nevem Boglárka. Magyar vagyok, de Szerbiában dolgozom. Belgrádban és Monoron élek. Monor egy kisváros Budapest mellett. Tanár vagyok a Belgrádi Egyetemen. Magyar nyelvet tanítok. Angolul jól beszélek, angoltanár is vagyok. Kicsit beszélek németül és törökül. Szerbül most tanulok. Két testvérem van. A húgom egyetemista Egerben, ő is angoltanár lesz. Egy kicsit beszél spanyolul is. Az öcsém Budapesten él. Egy irodában dolgozik. Most németül tanul, de egy kicsit beszél olaszul.

A nevem Vali vagy Valéria. Magyar vagyok, de nem Magyarországon, hanem Angliában élek. Londonban élek, mert itt dolgozom. Újságíró és tanár vagyok. A BBC-nél dolgozom, és a Külügyminisztériumban diplomátákat tanítok magyarul. Magyar az anyanyelvem, de angolul és spanyolul is beszélek, és egy kicsit franciául is. Nagyon érdekel a nyelvtanulás. Most olaszul tanulok, mert nagyon érdekel az olasz kultúra és a gasztronómia. A húgom Magyarországon, Gödöllőn él. Taxisofőr, és nagyon sokat dolgozik. Ritkán találkozunk, de minden nap beszélünk skájpon.

² A gyűjteményt a későbbiekben nem tanárok által írt szövegekkel is bővíteni kívánjuk.

A diák saját szövegének megírása előtt tanulmányozza a modellszövegeket, és már az olvasás során megjelölheti bennük azokat a kifejezéseket, amelyeket saját szövegében használni kíván. Az alábbi szöveget egy diák az A1-es tankönyv 2. fejezetének feldolgozása és a modellszövegek elolvasása után alkotta:

A nevem Nadia. Holland vagyok. Az egész családom holland. Az édesanyám, az édesapám, az öcsém és én Hollandiában, Leidenben élünk. Leiden egy szép város Hága mellett. Körülbelül százhuszonezer ember él ott. Nagyon érdekelnek a nyelvek és a nyelvtanulás. Egyetemista a Leideni Egyetemen és némettanár egy iskolában vagyok. Is franciatanár leszek. Beszélek nagyon jól németül, franciául és olaszul. Elég jól tudok angolul. Most tanulok magyarul, mert a magyar nyelv nagyon szép. Nekem elég egzotikus, és ezért érdekes. Holland az anyanyelvem. Otthon hollandul beszélünk.

Kennedy és Miceli kutatásaival (2010, 2017) összhangban saját tapasztalatunk is azt mutatja, hogy a modellszövegek elemeinek visszaforgatása lényeges mértékben javítja a diákok nyelvi produktumainak minőségét. Hadd idézzem egy A2 szintű nyelvtanuló visszajelzését, aki a modellszövegek hasznát következőképp fogalmazta meg: „Mielőtt modellszövegekkel dolgoztam, örültem, ha valamit le tudtam írni magyarul, ami nagyjából érthető. Fogalmam sem volt róla, hogy mennyi mindent tudok már mondani. Most többet, jobban és gyorsabban írok, és ez nagyon motiváló. A barátaim is azt mondták, hogy jobban beszélek magyarul.” Több diák visszajelzése is azt mutatja, hogy a modellszövegek nélkül kisebb arányban építenék be a tanult kifejezéseket saját nyelvi produktumaikba, valamint jó néhányan beszámoltak arról is, mekkora biztonságérzetet ad számukra, hogy tudják: amit leírnak, az helyes és érthető. A szövegekkel való munka a hasznos kifejezések, gyakori kollokációk fontosságára is ráirányítja a figyelmet.

Csak mellékesen jegyezzük meg, hogy a modellszövegek tanulmányozása a nyelvtanárnak és tankönyvíróknak is fontos tanulságokkal szolgálhat. Érdekes például megfigyelni, hogy a *van/nincs + -m* (*Van egy bátyám. Sajnos nincs testvérem.*) mindegyik szövegben szerepel, és a *szüleim, a nagyszüleim, a testvéreim* és a *barátaim* kifejezések is nagy gyakorisággal kerülnek elő, ugyanakkor a nyelvkönyvek – ebből a szempontból a *MagyarOK* sem nem kivétel – viszonylag későn vezetik be őket. Ezek az első modellszövegben még új, idegen elemek az ismétlődéseknek és variációknak köszönhetően a szövegek elolvasásával rögzülnek, és a diákok tapasztalataink szerint azonnal igen nagy biztonsággal tudják használni őket.

4. Lexikai és grammatikai elemek jelentésének és használatának feltérképezése

A pedagógiai korpusz nemcsak szöveggyűjteményként használható, hanem egy-egy lexikai elem használati sajátosságainak feltérképezésében is segíthet. Nézzünk két példát arra, hogyan használhatjuk a korpuszt egy-egy szó jelentésének tisztázására! A nehezen magyarázható és fordítható *pedig* szóra a korpusz 258 példát ad, ebből álljon itt tíz:

Itunk Cádizban, Spanyolországban. Gyönyörű volt az óceán, a régi város, a spanyolok	pedig	nagyon kedvesek. Szeretnék még egyszer odautazni. Amikor külföldön vagyok, múzeu
nem kelek korán, kávézom, elolvasom a híreket, felkészülök az óráimra, a tanítás után	pedig	sokszor sétálok egy nagyot az iskola melletti szép parkban. Néha fotózok is. Londonba
em tudtam igazán jól megtanulni, de szeretném folytatni a tanulást. A következő célo	pedig	az, hogy olaszul is megtanuljak. Szerintem nagyon igaz az a mondás, hogy annyi emb
; az angolt a tévéből és az internetről is meg lehet tanulni. Sajnos nekem nem sikerült,	pedig	néztem sok Spongya Bobot. Ezért most magánórákra járok, és próbálok az egyetem
általában a jó munkalehetőségek motiválják, Spanyolországban és Franciaországban	pedig	sok britnek van második otthona, és ez is segíti a nyelvtanulást. Nagyon sok iskolában
nül, angolul és franciául, de az általános iskolában oroszul tanultam, a gimnáziumban	pedig	olaszul. Még sok mindent értek olaszul, ha valaki lassan beszél. Nyelvtanár vagyok, m
oltam a világ legrosszabb nyelvtanulója! Soha nem tudtam leülni és szavakat magolni,	pedig	szeretem a nyelveket. Az iskolában németül és franciául tanultam, de végül angoltaná
álja megérteni az ő nyelvét. Ezért tudok néhány szót, mondatot kínaiul és arabul, most	pedig	szerbül és spanyolul tanulok. Spanyolul már régóta szeretnék megtanulni, ezért nagyoi
apom egyforma volt: 6 órától este 8 óráig dolgoztam. 5 óráig voltam az iskolában, 5-től	pedig	otthon tanítottam angolt és magyart. Ma már nincs átlagos munkanapom. Minden nap i
Mit parancsolnak? Adél: Én egy eszpresszót kérek, és két deci narancslevet. Nóra: Én	pedig	egy hosszúkávét és egy szénsavas ásványvizet. Pincér: Máris hozom. Adél: Azt ígérte

Az igen szerteágazó jelentésű *jó* melléknév használatára is számos – összesen 250 – találatot kapunk. Ebből szintén tíz véletlenszerűen választott példát közlünk:

iszéltük, hogy kedden találkozunk. Ne haragudj, de közbejött valami, és mégsem	jó	a kedd. Ráérnél esetleg pénteken délelőtt? Ákos: Egy pillanat, megnézem a napt
ragy! Én elég keveset keresek. 11. Jó érzés, hogy tudok másoknak segíteni. - De	jó	! Én nem tudok másoknak segíteni. Nem érzem, hogy hasznos, amit csinálok. Ez
rommal. A kislányom 13 éves, és iskolába jár. Hosszúhetényben van egy nagyon	jó	és viszonylag nagy iskola. Kb. 320 gyerek jár ide. Itt gyönyörű a környék és nagy
/ Brazíliában? A2_TK1/1. Tegeződjünk, jó? - Jó, nagyon szívesen. Tegeződjünk,	jó	? - Persze. - Tegeződjünk, jó? - Ne haragudjon, de nem szeretnék tegeződni. - At
öt eszem. Ma már vagy két kiló narancsot megettem. Hanna: Az ilyenkor nagyon	jó	, mert sok benne a C-vitamin. Meglátogatlak, jó? Tamás: Jaj, de jó! Társaságban
egyáltalán nem, de hát... De hát azért mégiscsak, mégis csak jó. Azt hiszem, ez	jó	lesz. Ja igen, ez talán jó lesz. Zsófi: Hát, most ez egy ilyen periódus, nem? Orsi: .
erjessel. Kertünk sajnos most nincs, pedig imádom kertészkedni. A lakásban van	jó	sok növény. Hágában sok kulturális lehetőség van és a közlekedés is kifejezetten
kor? Laura: Egy kis türelmet kérek, megnézem a naptáramat. A hétfő sajnos nem	jó	, mert nem vagyok az irodában. Mit szólna a kedd délelőtt tíz órához? János: Rer
zóta is sokszor eszembe jutnak a gyerekkori élmények, a konyha illata... Ildi: De	jó	lehetett! Én is gyerekkoromban tanultam meg sütni-főzni. A főzés ugyan nem a ke
mert igazságtalan volt. A2_TK5/4. Az az igazság, hogy nem voltam túl jó tanuló.	jó	tanáraim voltak, de valamiért mégsem élveztem az órákat. Azt hiszem, egyszerű

A konkordanciasorok, amelyek a valóságban egymástól távoleső megnyilvánulásokat hoznak egymás közelébe, a nyelv újfajta, nem lineáris olvasatát teszik lehetővé. Az egymás alatt megjelenő mondatok vertikális olvasata a kulcsszó jobb és bal oldalán elhelyezkedő nyelvi elemekben fellelhető ismétlődésekre, mintázatokra is ráirányítja a figyelmet (ld. pl. Sinclair 1991). Tíz példából még nem feltétlenül rajzolódna ki világosan a *pedig* vagy a *jó* jelentései és gyakori kollokációs partnerei (igaz, a *De jó! a nem/mégsem jó* és a *nagyon jó* kifejezések ismétlődése már itt is látszik), de 40–50 konkordanciasort átolvasva már igen. Ennyi példamondatot egy-egy gyakori szóval a tanuló saját nyelvtudási szintjén egyetlen szótárban sem talál – ebből a szempontból is igen értékesek lehetnek a pedagógiai korpuszok.

Végezetül villanjon fel egy lehetőség a korpuszhasználatra a grammatika tanításában! A magyar nyelvnek egyik alapvető, ugyanakkor nehezebben tanítható aspektusa a szórend. A konkordanciasorok a szórendi tendenciák feltérképezésében is segítséget nyújthatnak. Példaként nézzünk meg a *lakom* szó használatát, ezúttal csak az A1-es alkorpuszban! A 23 példa közül ismét tízet idézünk:

itt élni. Egyrészt mert biztonságos, másrészt mert nyugodt hely. Családi házban	lakom	a férjemmel. Kertünk sajnos most nincs, pedig imádok kertészkedni. A lakásban v
as, de amikor felnőttem, már jóval kisebbnek láttam. Most London belvárosában	lakom	, egy több emeletes sorházban. Sajnos, itt sem kertem nincs, sem erkély vagy ter
volt, ő most is minden évben meghív bennünket a születésnapjára. Kertvárosban	lakom	egy társasházban a két lányommal és a férjemmel. Lakástulajdonosok vagyunk. ,
rista jön télen és nyáron bort kóstolni. A szülővárosom Lenti. Már régóta nem ott	lakom	, de gyakran járok ott, mert ott él a családom. Lenti egy kisváros Nyugat-Magyar
pest is mindig kedves lesz számomra, hiszen ez az otthonom. Hosszúhetényben	lakom	. Ez egy kis falu Dél-Magyarországon, kb. 3000 ember él itt. Én egy szép új házba
ltban van bolt. A központban vannak boltok. A1_MF4/22. 1. Én a Balassi utcában	lakom	. Te is a Balassi utcában laksz? Péter is ott lakik. 2. Mi biciklivel közlekedünk. Ti is
n az anyukámnak. Sajnos hinta nem volt az udvarban. Egy kétszintes sorházban	lakom	a férjemmel. Sajnos nincsen kertünk, de van egy szép tetőteraszunk és két erkély
omnak van erkélye, ami azért jó, mert[3] nyáron ott lehet pihenni. Egy kis faluban	lakom	Pécs mellett. Az utolsó utcában, az utolsó házban. Ez egy családi ház nagy kertt
/szlát! A1_TK/33. Szia, Luca! Hogy vagy? Én remekül. Azért írok, mert új helyen	lakom	. Találtam egy gyönyörű lakást a belvárosban. Közel van a munkahelyemhez, gyé
önyvtár. A barcelonai egyetemeken sok külföldi diák tanul. Magyarországon élek/	lakom	. (ország) Sopronban élek/lakom. (város) A Petőfi utcában lakom. (utca) Egy szép

Már a fenti mondatokból is kibontakozik az a szórendi tendencia, hogy amikor a *lakom* igének két bővítménye van, az egyik bővítmény általában megelőzi, a másik pedig követi a szót: *egy kis faluban lakom Pécs mellett / kertvárosban lakom egy társasházban a két lányommal / családi házban lakom a férjemmel* stb. Az efféle, korpuszban tett “kirándulások” a diák számára igen hasznosak, hiszen a szórend nehezen megfogható voltát talán csak nagy számú, hasonló struktúrájú mondat tanulmányozása tudja világosabbá tenni.

5. Zárzó

Cikkemben a pedagógiai korpuszok építésének és felhasználásának néhány lehetséges módját ismertettem röviden, a *MagyarOK* nyilvános korpusz példáján. A korpusz megépítése során a következő alapelvet alkalmaztuk: (1) A gyűjtemény viszonylag kicsi és nagyrészt a tanórán feldolgozott tartalmakra épül, így a diák könnyen kontextualizálni tudja a találatokat. (2) A korpusz a diák számára releváns nyelvi tartalmakat mutat be. A témák és szövegtípusok megfelelnek az adott nyelvtudási szint KER előírásainak. (3) A nyelvészeti céllal készült korpuszokkal ellentétben a pedagógiai korpusz a teljes szövegekhez is hozzáférést biztosít. Ezek a szövegek egyrészt kontextusba helyezik a megfigyelt nyelvi elemeket, másrészt a diák modelleként használhatja őket saját nyelvi produktumai létrehozásánál.

A *MagyarOK* korpusz használható példamondatok és modellszövegek gyűjteményeként, amely a diákok saját mondatainak és szövegeinek megalkotásában segít, de egyes szavak jelentésének tisztázására és nyelvtani jelenségek megfigyelésére is. S mivel a nyilvánosan elérhető, felhasználóbarát pedagógiai korpuszok használatára véleményünk szerint egyértelmű, reményeink szerint ezek a gyűjtemények a jövőben a MID-tanárok eszköztárának magától értetődő alkotóelemét fogja képezni.

Irodalom

- Aston, Guy 2001. *Learning with Corpora*. Houston, Athelstan.
- Boulton, Alex 2017. Research timeline: Corpora in language teaching and learning. *Language Teaching* 50/4: 483–506.
- Boulton, Alex & Thomas, James 2012. Hands-on/hands-off: Alternative approaches to data-driven learning. In: James Thomas – Alex Boulton (szerk.): *Input, process and product: Developments in teaching and language corpora*. Brno, Masaryk University Press. 153–169.
- Brezina, Vaclav 2018. *Statistics in corpus linguistics. A practical guide*. Cambridge, Cambridge University Press.

- Braun, Sabine 2006. Getting past Groundhog day. Spoken multimedia corpora for student-centred corpus exploration. In: Harris, T. – Moleno Jaén, M. (szerk.): *Corpus linguistics in language teaching*. Frankfurt, Peter Lang, 75–98.
- Braun, Sabine 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL* 19/3: 307–328.
- Brazil, David 1995. *A grammar of speech*. Oxford, Oxford University Press.
- Flowerdew, John 2009. Corpora in language teaching. In: Long, H. – Doughty, C. J. (szerk.): *The handbook of language teaching*. London, Blackwell, 327–350.
- Hoey, Michael 2005. *Lexical priming: A new theory of words and language*. London, Routledge.
- Hunston, Susan 2009. The usefulness of corpus-based descriptions of English for learners: The case of relative frequency. In: Aijmer, K. (szerk.): *Corpora and language teaching*. Amsterdam, John Benjamins, 141–157.
- 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>
- Römer, Ute 2006. Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik* 54/2: 121–134.
- Sinclair, John 1991. *Corpus, concordance, collocation*. Oxford, Oxford University Press.
- Szita Szilvia (megjelenés előtt). *Corpus design and activities for lower proficiency levels*. Cambridge, Cambridge University Press.
- Szita Szilvia 2014. Invent content, not language. Meaningful interaction and natural language use in the classroom. In: Hegedűs R. – Görbe T. (szerk.): *Kleine Sprachen, was nun? Studies on language and culture in Central and Eastern Europe*. München, Kubon und Sagner, 112–127.
- Szita Szilvia – Pelcz Katalin 2017. Modellalapú nyelvoktatás, természetes nyelvhasználat. *THL2* 1/2: 262–269.
- Szita Szilvia – Pelcz Katalin (kézirat). *A modellalapú nyelvtanítás kézikönyve*.
- Taylor, John 2012. *The mental corpus. How language is represented in the mind*. Oxford, Oxford University Press.
- Timmis, Ivor 2015. *Corpus linguistics for ELT. Research and practice*. London, Routledge.
- Tyne, Henry 2012. Corpus work with ordinary teachers: Data-driven learning activities. In: Thomas, J. – Boulton, A. (szerk.): *Input, process and product: Developments in teaching and language corpora*. Brno, Masaryk University Press, 114–129.
- Widdowson, Henry 2003. *Defining issues in English language teaching*. Oxford, Oxford University Press.
- Willis, Dave 2003. *Rules, patterns and words: grammar and lexis in ELT*. Oxford, Oxford University Press.

Szita, Szilvia

Corpus building and the use of corpora at lower levels of language learning

The use of linguistic corpora is not a common practice in the teaching of Hungarian as a foreign language. Although the use of these large text collections has a number of benefits of which the present article can present only a few, teachers or students are often reluctant to consult them. The principal reason for this seems to be that most existing corpora were built with the expectations and aims of linguists and researchers in mind. This article argues that corpora can only be relevant and effective for language teaching purposes when they are built in accordance with the learners' needs. An example of pedagogical databases is the *MagyarOK* corpus for Hungarian, available among the public corpora on the Sketch Engine website. In the second half of the article, the content of this corpus and the methodological considerations behind its construction are briefly described.