

Szita Szilvia

Ön most eljött vagy megjött? Korpuszhasználat a magyar mint idegen nyelv tanításában*

1. Problémafelvetés

„Tanárnő, kérem, mit jelent pontosan az, hogy *megjön*? És az, hogy *eljön*? Melyiket mikor használjuk? Mi a különbség közöttük?”

A nyelvtanulók a tanulási folyamat során rendre szembesítenek bennünket olyan kérdésekkel, amelyekre nehéz egyértelmű választ adni. Az igekötős igék problémák tekintetében vitathatatlanul élen járnak: használatukra igen nehéz szabályokat megfogalmazni, és még nehezebb a szabályokat (rögtönzött) példákkal alátámasztani. Nyelvtanárként kézenfekvő, hogy első lépésben szabályt próbálunk adni a nyelvtanulónak – hiszen ez többnyire lerövidíti az adott nyelvi jelenség megértéséhez vezető utat –, majd a szabályra példákat keresünk. Ez a módszer számos nyelvi jelenség (többes szám, múlt idő stb.) esetében jól működik, vannak azonban a nyelvnek olyan jellemzői, amelyek nem feltétlenül írhatók le egyértelmű szabályokkal. Ezek közé tartozik az igekötős igék használata. Márpedig ha egy jelenségre nem tudunk pontos szabályt adni, még nagyobb bizonytalanságban hagyjuk a tanulót, mint mielőtt feltette kérdését: az lehet az érzése, hogy „itt nincs szabály,” „ezt érezni kell”, „ezt csak az anyanyelvi beszélők tudják helyesen használni.”

Ebben a cikkben emellett kívánok érvelni, hogy azokban az esetekben, amikor nehezen körülírható, nem egyértelmű jelentéssel bíró nyelvi elemről van szó, ne elvont szabályok, hanem autentikus, anyanyelvi beszélőktől származó példák felől közelítsük meg a kérdést. Ahelyett, hogy először szabályokat, majd példákat próbálnánk megfogalmazni a nyelvtanuló számára, fordítsuk meg a munkafolyamatot: először adjunk nagy mennyiségű, autentikus példát, amelyeken keresztül a nyelvtanuló megfigyelheti a használati tendenciákat. Vagyis adjunk lehetőséget arra, hogy a tanuló a példákon keresztül olyan nyelvi tapasztalatot építsen fel, amellyel az anyanyelvi beszélő implicit, nem reflektált módon eleve rendelkezik, és amelynek birtokában (többnyire) helyesen használja anyanyelvét. Az anyanyelvi beszélőnek – hacsak nem nyelvtanár – ugyanis nem a szabályokról van elméleti, rendszerezett tudása, hanem a használaton keresztül rendelkezik az adott lexikai elemről való gyakorlati tudással. Ha a nyelvtanuló először nagy számú példát tanulmányoz és elemez, és csak ezután próbálja levonni a használati tendenciákat, hasonló módon fejleszti nyelvi kompetenciáját, mint ahogyan az anyanyelvi beszélő szerezte meg azt.

A modern technológiának, elsősorban az internetnek köszönhetően korábban soha nem látott méretű nyelvi adatbázisok (irodalmi szövegek, fórumok, posztok, blogok stb.) állnak rendelkezésünkre, és teszik lehetővé az élőnyelv vizsgálatát. Ezek a korpuszok számtalan autentikus példát kínálnak egy-egy lexikai elem használatára, amelyeket megfelelő didaktizálás mellett már alacsonyabb szinteken is bemutathatunk a nyelvtanulóknak. Az alábbiakban arra kívánok egy lehetséges munkafolyamatot bemutatni, hogyan dolgozhatunk az interneten elérhető magyar nyelvű korpuszokkal, milyen válaszokat nyerhetünk ki belőlük az igekötős igék használatával kapcsolatban, és ezeket hogyan prezentálhatjuk a diákoknak. A módszer szemléltetéséül a *megjön* és az *eljön* igék vizsgálatát választottam.

2. *A tankönyvek és a nyelvtanár lehetőségei*

A tankönyvek az élőnyelvről legfeljebb keresztmetszeti képet adhatnak (Leńko-Szymańska–Boulton 2015, 12). Még ha a dialógusok, narratív szövegek és kapcsolódó feladatok a valós nyelvhasználatot tükrözik is, a tankönyv – terjedelménél fogva – nem adhat elegendő példát egy-egy lexikai elem használatára, hiszen ezek nem szerepelhetnek benne azzal a gyakorisággal, amellyel a mindennapokban előfordulnak. Ez azért rendkívül lényeges, mert mindenfajta nyelvtanulásra igaz, hogy minél több példán keresztül találkozunk egy szóval, annál inkább rögzülnek a hozzá tartozó, jellemző kontextusok is. **A szó ismerete tehát egyre inkább magában foglalja a használatáról való tudást is**, vagyis azt, hogy milyen más szavak társaságában, milyen összefüggésben fordul elő jellemző módon (Hoey 2005, 8).

A nagy számú példa hiánya olyan, egyértelmű jelentéssel bíró szavak esetében, mint *asztal, kávé, piros, sétál* stb., nem jelent problémát, hiszen a nyelvtanuló néhány példából megérti a pontos jelentést, levonja a fontosabb használati szabályszerűségeket, majd ennek a tudásnak a birtokában többnyire helyes összefüggésben használja a szót. A kevésbé egyértelmű jelentésű lexikai elemek használatával kapcsolatban – ide tartoznak az igekötők, illetve az igekötős igék – azonban a tankönyvek nem adhatnak elegendő számú példát, amelyekkel találkozáskor a nyelvtanuló megszerezheti a már említett, anyanyelvi beszélőhöz hasonló nyelvi kompetenciát. A nyelvtanári példák – még ha hozzájárulnak is a tanuló adott lexikai elemről való tudásának bővítéséhez – a biztos és helyes használathoz mennyiségüknél fogva nem adhatnak elegendő fogódzót. Ennek elsajátításához ugyanis a nyelvtanulónak az adott elemet sok-sok különféle kontextusban kellene megfigyelnie, csak így térképezheti fel az alakhoz tartozó használati lehetőségeket. Ebben segítenek a korpuszok.

3. *A korpuszok*

3.1. *A korpusznyelvészet és az anyanyelvi beszélők nyelvhasználata*

A korpusznyelvészet a valóban használt nyelv tanulmányozhatósága tekintetében vitathatatlanul új fejezetet nyitott (pl. Boulton–Tyne 2014; Biber 2009). Az ágazat az alkalmazott nyelvészet egyik legdinamikusabban fejlődő területe, a kutatások túlnyomó része az angol nyelvre irányul, és az angolszász országokban zajlik. Ez a nyelv mára számos írott és beszélt nyelvi korpuszsal (*British National Corpus, American National Corpus, Brown Corpus* stb.) és korpuszalapú nyelveírással rendelkezik (ld. pl. Burnard–McEnery 2000; Carter–McCarthy 2014; Gray–Biber 2016; McCarthy–Dell 2016).

A korpusznyelvészet egyik legfontosabb felismerése, hogy valós nyelvhasználatunk általában nem eredeti és nem korlátlanul kreatív. Éppen ellenkezőleg: az anyanyelvi beszélők természetes nyelvhasználatának egyik jellemzője, hogy egy-egy közléshelyzet megvalósítására megvannak a maguk preferált kifejezési módjai. Ahelyett, hogy megnyilatkozásukat minden alkalommal egyes elemekből építenék fel, előszeretettel használnak részben vagy teljes egészében rögzített lexikai egységeket. Ezeknek előnye, hogy leegyszerűsítik a kognitív folyamatokat, gördülékennyé teszik a beszélgetést, és olyan közös teret hoznak létre, amelyet a beszélgetőpartnerek mindegyike magáénak ismer el (Davis–Kryszewska 2013, 8). Az eredetiség a szépirodalom, ezen belül is elsősorban a költészet ismertetőjegye: ez különbözteti meg minden más műfajtól, regisztertől, amelyekre az eredetiség és a kreatív nyelvhasználat sokkal kevésbé jellemző.

Annak a ténynek, hogy az élőnyelvi megnyilvánulások jelentős hányadát részben vagy teljes egészében rögzített kifejezések alkotják, az idegennyelv-tanítás gyakorlatában is figyelmet kell szentelni. Ha a nyelvtanuló egy adott szót önmagában sajátít el, az jó eséllyel

vezet nyelvhasználati hibákhoz, hiszen a tanuló éppen a leglényegesebb információval nem rendelkezik, nevezetesen azzal, hogy a szó milyen kontextus(ok)ban fordul elő jellemző módon. Ahhoz, hogy a hibákat elkerülje, azt is tudnia kell, **mi veszi körül a szót**. Ha tehát nyelvtanárként az a célunk, hogy megkönnyítsük diákjaink számára a természetes nyelvhasználat elsajátítását, akkor lényeges, hogy valószerű, anyanyelvi beszélők által használt nyelvi eszközöket tanítsunk, és ezekre külön is felhívjuk a diákok figyelmét (Lewis 1997, 21; Davis–Kryszewska 2013, 10). A preferált kifejezésmódok ismerete nélkül ugyanis a nyelvtanuló képtelen az anyanyelvi beszélőkhöz hasonló, természetes nyelvhasználatot elsajátítani, holott ez volna a nyelvtanulás elsődleges célja (vö. pl. Cowie 2002, 3; Pace-Sigge 2013; Boulton–Tyne 2014, 23). A természetes nyelvhasználat megfigyelésére, a preferált kifejezésmódok összegyűjtésére és tanulmányozására az interneten található, hatalmas élőnyelvi adatbázisok adnak egyedülálló lehetőséget.

3.2. *A felhasznált magyar nyelvű korpuszok*

A korpusz „ténylegesen előforduló írott, vagy lejegyzett élőnyelvi adatok gyűjteménye” (<http://corpus.nytud.hu/mnsz/>). Ebben az értelemben minden szövegeken alapuló, kellő méretű adatbázis tekinthető korpusznak. Az annotált korpuszok „a szövegeket valamilyen szempont szerint válogatják és rendezik is. (...) Nem csak tárháza a szövegeknek, hanem tartalmazza azok bibliográfiai adatait, bejelöli a szerkezeti egységeket (...). Emellett pedig feltünteti a szavak mellett szófaji kódjukat is” (uo.).

Annotált korpuszból magyar nyelven mindössze egyetlenegy áll rendelkezésünkre. A 2005-ben létrehozott és azóta folyamatosan bővülő *Magyar Nemzeti Szövegtár*, amely „az írott köznyelv általános célú reprezentatív korpusza” kíván lenni (uo.), jelenleg 187,6 millió szövegszót tartalmaz. Alkorpuszai a sajtóból vett, szépirodalmi, tudományos, hivatalos és személyes szövegeket tartalmaznak. A keresett szóval gyakorisági listát és konkordanciát készíthetünk, valamint kollokációs vizsgálatokat végezhetünk.

A másik nagyméretű adatbázis, a *Hunglish* korpusza magyar–angol nyelvű, és 120 millió szövegszót tartalmaz. Szótári bejegyzéseket, filmfeliratokat, szépirodalmi szövegeket, számítástechnikával kapcsolatos, sajtó-, jogi és tudományos szövegeket is találunk benne. A filmfeliratokat tartalmazó alkorpusz különösen érdekes lehet a nyelvtanár számára, hiszen – jellegéből adódóan – kevés narratív szöveget tartalmaz, ehelyett inkább az interakciókban használt, többnyire a beszélt nyelvet tükröző mondatok dominálnak. Ugyan korpuszelemzésre ezen az oldalon nincs lehetőség, a filmfeliratok rövid, viszonylag egyszerű mondatait magunk is könnyen megszerezhetjük. Jól használható az adatbázis szépirodalmi szöveggyűjteménye is, mert számos könny(ebb)en érthető művet tartalmaz, amelyek nyelvhasználata közelít a beszélt nyelvhez.¹

4. *A korpuszhasználat előnyei a nyelvoktatásban*

A korpuszok nyelvoktatásban való alkalmazásának több előnye is van: (1) belőlük statisztikai elemzéssel megbízható adatok nyerhetők egy-egy szó előfordulásának gyakoriságáról; (2) feltérképezhetőek azok a szókapcsolatok, amelyek a választott lexikai elemet tartalmazzák, s annak jelentését újabb és újabb árnyalatokkal bővítik; valamint (3) a szó mondatbeli viselkedésének grammatikai, szintaktikai és pragmatikai jellemzői is leírhatóak. A korpuszhasználat tehát összességében abban segít, hogy a nyelvtanuló egy-egy lexikai elem használata-

¹ Ez nyelvtanulói szempontból azért lényeges, mert a Közös Európai Referenciakeret az A1–B1-es szinten a mindennapi nyelvhasználat, rutinszerű kommunikációs helyzetekben való megnyilvánulások ismeretét irányozza elő, így ebből a korpuszból már alacsonyabb szinten is releváns példák meríthetők.

tát az élőnyelvből vett nagy mennyiségű, rendszerezett példán figyelhesse meg, és ne példákából levont, gyakran csak nehezen értelmezhető használati szabályok felől közelítse meg. A nyelvtanuló ezáltal autentikus mondatok birtokába kerül, ezeken keresztül vizsgálhatja a nyelvnek egy olyan aspektusát, amely a tanulás adott fázisában releváns számára. Ennek köszönhetően az élőnyelv működéséről szerzett tudása is mélyül, finomodik, nyelvi tudatossága erősödik (Leńko-Szymańska–Boulton 2015, 3).

Ebben a munkafolyamatban a konkrét példától haladunk az absztrakció felé, ám a folyamatnak csupán egyik eredménye a használati tendenciák rendszerezése. Legalább ugyanennyire lényeges, hogy a tanuló újfajta módon közelít a nyelvhez: megtanulja, hogy bizonyos esetekben kellő mennyiségű, releváns példa vizsgálata jobb eredményre vezet, mint egy-egy elvont szabály értelmezése.

Az alább ismertetendő munkafolyamatban az *eljön* és a *megjön* igék példáján azt kívánom bemutatni, milyen információk nyerhetők ki a választott korpuszból, amelyek segíthetnek két hasonló jelentésű szó (jelen esetben ige) használati jellemzőinek megfigyelésében és elsajátításában. A vizsgálathoz a *Magyar Nemzeti Szövegtár* magyarországi irodalmi és sajtóban megjelent szövegeit, fórumbejegyzéseit, valamint a *Hunglish* filmfeliratainak és irodalmi szövegeinek gyűjteményét használtam fel. A gyakorisági eloszlás megfigyeléséhez két helyen a *Google* keresőprogramján keresztül elérhető dokumentumokat is bevontam a vizsgálatba.

5. A munkafolyamat leírása

A választott lexikai elemeket a javasolt munkamenetben három lépésben vizsgáljuk: (1) előfordulási gyakoriságuk, (2) jellemző környezetük, szemantikai tartományuk és (3) grammatikai, szintaktikai, pragmatikai sajátosságaik. Mivel jelen cikk elsődleges célja, hogy a munkamódszert bemutassa, lemondunk a sajátosságok mindenre kiterjedő leírásáról, csak megjegyezzük, hogy ezek mindegyike tovább pontosítható.

5.1. Az igeidők eloszlása gyakoriság alapján

Az összehasonlító vizsgálat első lépésében a gyakoriságot vesszük nagyító alá. Egy egyszerű keresés a jelen és a múlt idejű, egyes szám harmadik személyű alakok gyakoriságára a *Magyar Nemzeti Szövegtárban* az alábbi eredményeket hozza:

1. táblázat: A használati arányok eloszlása a Magyar Nemzeti Szövegtárban

| megjött | megjön | eljött | eljön |
|--------------------------------|-------------------------------|--------------------------------|--------------------------------|
| 7989 | 4059 | 17623 | 15338 |
| ebből 2364 mondatkezdő pozíció | ebből 207 mondatkezdő pozíció | ebből 3772 mondatkezdő pozíció | ebből 1378 mondatkezdő pozíció |
| 66% | 34% | 53,5 % | 46,5 % |
| Összes találat: 12 048 | | Összes találat: 32 961 | |

A *Google* keresőprogramjában folytatott keresés egy jóval nagyobb adatbázisban hasonló eredményt mutat:

2. táblázat: Használati arányok a Google keresőprogramban

| megjött | megjön | eljött | eljön |
|---------------------------|----------------|---------------------------|----------------|
| 899 000 | 297 000 | 1 320 000 | 917 000 |
| 75,16 % | 24,84 % | 59 % | 41 % |
| Összes találat: 1 196 000 | | Összes találat: 2 237 000 | |

A fenti adatokból kiderül, hogy az *eljön* és *eljött* együttesen mindkét korpuszban legalább kétszer olyan gyakran fordul elő, mint a *megjön* és a *megjött* igék. Az is látható, hogy a *megjön* ige esetében egyértelműen a múlt idejű alakok dominálnak, míg az *eljön* esetében a múlt idejű alakból mindössze néhány százalékkal van több, mint a jelen idejű alakból.

Ezzel az egyszerű vizsgálattal máris egy fontos jellemzőt tártunk fel a két ige kontrasztív leírásához: amellet, hogy gyakoriság szempontjából az *eljön* ige alakjai vannak túlsúlyban, az is kiderül, hogy a *megjön* igével túlnyomórészt múltban lezárult eseményeket írunk le, míg az *eljön* ige esetében a jelen és a múlt idejű használat is gyakori.

Ezeket a vizsgálati eredményeket első lépésben a tanulókkal is megoszthatjuk, hiszen a jellemző igeidők ismerete is hozzájárul az ige használatáról való tudáshoz. A teljes munkafolyamaton is végigvezethetjük őket, megtaníthatjuk, hogyan végezhetnek tanári segítség nélkül hasonló vizsgálatokat.

5.2. A jelentések eloszlása gyakoriság alapján

A következő lépésben először érdemes önmagunk számára feltérképezni a két ige lehetséges jelentéseit. Nézzük meg ehhez *A magyar nyelv értelmező szótárának* bejegyzéseit az *eljön* és a *megjön* igékhez!² *Eljön*: 1. a beszélő közelébe, lakásába vagy valamely szóban forgó helyre jön, s oda megérkezik; jön és ideérkezik: *Eljön vhova; eljön vkiért, vmiért; gyalog, autón, vonaton jön el. Eljössz hozzám? Jöjjön el holnap újra. Sokan ott voltak az ünnepélyen, kár, hogy ti nem jöttetek el.* Elmegy v. elmenni készül oda, ahova a beszélő is megy: *Eljössz velünk kirándulni? Ki jön el velem? Ő is eljön velünk az ünnepélyre.* 2. Valahonnan útnak indul a beszélő felé, s így tartózkodási helyét elhagyja: *Már eljött hazulról. Mikor visszamentem, még ott volt, csak később jött el. Eljött egész a sarokig.* 3. (átvitt értelemben) <Időpont, -szak> bekövetkezik. *Eljön vminek az ideje, az órája. Végre eljött az ő napja is. Eljött a cselekvés órája. Eljött a szocializmus ideje.*

Ezek után nézzük meg, hogyan csoportosítja az értelmező szótár a *megjön* ige jelentéseit! 1. <Személy dolga végeztével> valahonnan visszatér, megérkezik otthonába, ahol a beszélő vagy az író is tartózkodik. *Megjön az iskolából, a városból, a mezőről, a mulatságból. Megjött a fogságból, a harctérről, külföldi úttjáról. Megjöttek a fecskék, a gólyák.* 2. <Kül. várt, beígérkezett személy> megérkezik oda, ahol várják, v. ahova érkezni szándékozik, s ahol a beszélő v. az író is van. *Megjött a vendég, az új munkatárs. Csakhogy megjöttetek! Megjött a csomag, a levél, a távirat, a rendelt kocsi. Mikor jön meg a levélre a válasz?* 3. (átvitt értelemben) <Valamely rendszeresen visszatérő időszak, jelenség, állapot> bekövetkezik, újra itt van. *Megjön a tavasz, a nyár, az ősz, a tél; megjön a farsang; megjön az őszi esőzés.* 4. (átvitt értelemben) <Várt dolog vagy jelenség> megvalósul, létrejön, megjelenik, bekövetkezik. *Megjön majd az is! Hajnalban végre megjött az eső.* <Rendsz. állandósult szókapcsolatokban: vkinek vmely képessége, kedvező testi v. lelki állapota> kifejlődik, létrejön v. újra megjelenik, helyreáll. *Megjön a bátorsága; megjön az esze; megjön az étvágya; megjön a kedve; megjön a szava.*

A fenti – használatra és jelentésre vonatkozó – rövid leírások az anyanyelvi beszélő számára teljesen világosak, hiszen olyan implicit, nem reflektált, spontán módon alkalmazott tudást fogalmaznak meg, amellyel már a szócikk elolvasása előtt is rendelkezett. A példák pedig mindössze arra szolgálnak, hogy felidézzék számára a már eleve ismert jelentéseket (Kiss 1999).

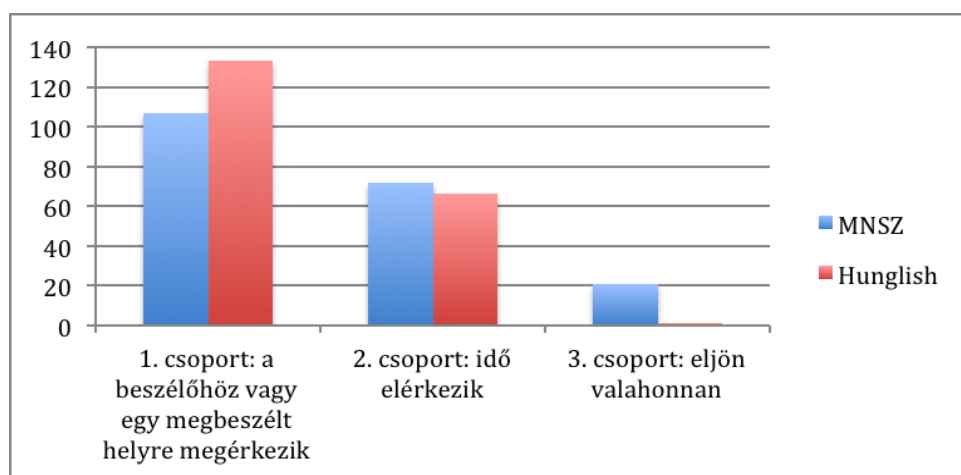
A nyelvtanuló azonban egy efféle összefoglalással csak nehezen boldogul, hiszen általában az anyanyelvi beszélőhöz képest jóval kevesebb tapasztalata van a nyelvhasználatot illetően. Egyrészt jóval több példára van szüksége, mint amennyit a szótárban talál, másrészt olyan jellegűekre, amelyek a jellemző használatot, nyelvi környezetet is bemutatják, és amelyek

² Ezen a helyen nem idézem azokat a példákat, amelyek szépirodalmi szövegekből vett idézetekben találhatók.

mennyisége és sorrendje tükrözi a jelentések gyakoriságát. Mint feljebb már említettem, az értelmező szótár elsősorban nekünk, nyelvtanároknak segít a jelentések feltérképezésében. A szótári definíciókat tehát ne ebben a lépésben ismertessük meg a tanulókkal, hanem csak azután, ha a korpuszból vett, nagy mennyiségű, rendszerezett példán keresztül megfelelő mennyiségű nyelvi inputot szereztek, és a példákat már tanulmányozták, elemezték (4.3–4.5). Azt is látni fogjuk, hogy a szótár definíciói a korpuszokból vett példákhoz viszonyítva tovább finomíthatóak és számos olyan, a nyelvtanuló számára hasznos információkkal bővíthetőek, amelyek sem a szótárban nem szerepelnek, sem az anyanyelvi beszélő nem képes saját nyelvi tapasztalatai alapján levonni.

Ha jelentésük szerint osztályozzuk a *Magyar Nemzeti Szövegtár* első kétszáz példáját, akkor azt látjuk, hogy az *eljön* ige az esetek több mint felében a szótárban első jelentésként definiált 'megérkezik valahova' jelentésben fordul elő, egészen pontosan 107-szer. A második leggyakoribb jelentés a szótár által harmadik, átvitt jelentésként megadott jelentés (időszak, idő elérkezik), amely 72-szer szerepel a korpuszban. A másodikként megadott 'eljön valahonnan' jelentés a korpuszban viszonylag ritka, 21 előfordulással az összes használatnak kb. 10 százalékát teszi ki. A *Hunglish* korpuszában az eloszlások aránya 133, 66 és 1.³

1. ábra: Az *eljön* ige szótárban feltüntetett jelentéseinek eloszlása a *Magyar Nemzeti Szövegtárban* és a *Hunglish* korpuszában⁴



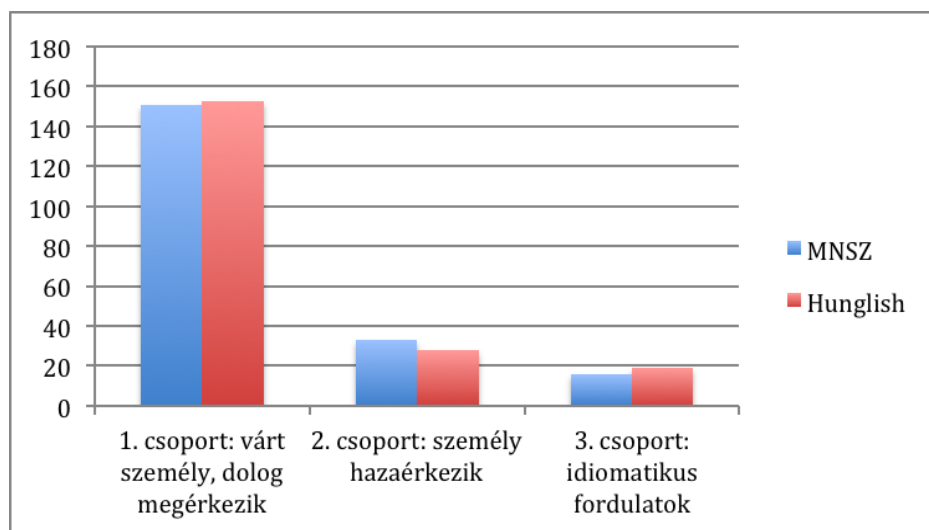
A 2. ábrán a *megjön* ige jelentéseinek eloszlását tanulmányozva azt látjuk, hogy a két korpusz erős hasonlóságot mutat. Idiomatikus fordulatokból van a legkevesebb (MNSZ: 16, Hunglish: 19 találat), de a 'hazatér, hazaérkezik' jelentés is viszonylag ritkán fordul elő (MNSZ: 33, Hunglish: 28 találat). A legtöbb példát a 'várt személy, dolog megérkezik' jelentéshez találjuk (MNSZ: 151, Hunglish: 153 találat).⁵

³ Az első és a harmadik jelentés eloszlásának különbségét a szövegek regiszterbeli különbsége magyarázhatja. A *Hunglish* korpuszában nagyobb számmal található informális kommunikációs helyzetekre jellemző megnyilvánulások, mint a *Magyar Nemzeti Szövegtárban*. A hipotézis igazolására további vizsgálatokra volna szükség.

⁴ A MNSZ korpuszában a találatokat a még pontosabb eredmények érdekében igeidők, szám, személy és mód szerint is lekérdeztük.

⁵ Az 'ismétlődő időszak visszatérése' jelentést a 'várt személy, dolog megérkezik' jelentés alá soroltuk be, hiszen valójában ennek egyik alcsoportját képezi.

2. ábra: A megjön ige szótárban feltüntetett jelentéseinek eloszlása a Magyar Nemzeti Szövegtárban és a Hunglish korpuszában



5.3. A példák rendszerezése: eljön

A Magyar Nemzeti Szövegtár és a Hunglish korpuszában több ezer példát találunk az *eljön* igével. A merítés elsődleges kritériuma, hogy ebből a nagyon bőséges anyagból olyan mondatokat gyűjtsünk ki, amelyek a gyakori, jellemző használati lehetőségeket mutatják. Amint korábban már megállapítottuk, a nyelvtanulónak nagy számú példára van szüksége ahhoz, hogy a nehezen körülírható jelentésű lexikai elemek használatát el tudja sajátítani. Ha a jellemző használatra keresünk példákat, természetes, hogy a választott mondatok között ismétlődéseket, részleges átfedéseket fogunk tapasztalni. A tipikus kontextus(ok) rögzítéséhez szükség van az ismétlődésre, de mivel a mondatok az átfedések mellett mindig tartalmaznak némi variációt is, újabb és újabb – hasonló, de nem teljesen azonos – nyelvi környezet megfigyelését is lehetővé teszik. Sinclair szerint „tíz példa szegényes minta; legalább ötvenre van szükség, hogy egy szó jelentéseit körvonalazhassuk, és százötvenre ahhoz, hogy megbízhatóan számoljunk be róluk” (Sinclair 1993, 7; idézi Szirmai 2005, 29). Mivel tapasztalataim szerint száz–százötven példával a nyelvtanulók – kivált alacsonyabb nyelvi szinteken – nemigen tudnak mit kezdeni, a nekik szóló példasor összeállításánál az ötven mondatot tartom irányadónak.

A következő lépésben az általunk összeállított példatárral dolgozunk. Az alábbi mondatok a Magyar Nemzeti Szövegtár sajtó, személyes és irodalmi szövegek valamint a Hunglish filmfelirat és szépirodalmi szövegek alkorpuszaiból származnak, amelyeket szemantikai alapon csoportosítottam. A különböző jelentéseket nem adtam meg – ezeket az elemzés során derítik majd ki a nyelvtanulók. Az első fázisban a példák megfigyelésén, a sokrétű nyelvi tapasztalat felépítésén van a hangsúly. Az *eljön* igével az alábbi ötven példát választottam, a példák mennyiségét a jelentések gyakoriságának megfelelően:

1. csoport

Ugye, eljössz?

Eljössz?

Eljössz megnézni a meccset?

Eljössz velünk?

Eljöttél!

Hát eljöttél.

Eljött hozzám.

Tudtam, hogy eljössz!

Remélem, el tudtok jönni.

Kösz, hogy eljöttél.

Kösz, hogy eljöttetek.

Köszönöm, hogy eljöttél.

*Jó, hogy eljöttél.
Jól tetted, hogy eljöttél.
Örülök, hogy eljöttél.
Örülnénk, ha eljönnél.
Szeretném megköszönni, hogy eljöttetek a dr.
Meinheimer tiszteletére adott fogadásunkra.
Személyesen is eljöttem, hogy megköszönjem.
Nem tudom, jó döntés volt-e, hogy eljöttem.
Eljött, ahogy ígérte.
Ha te ott leszel, talán a többiek is eljönnek.
Egyszer eljöttek az unokatestvéreim, és nekik is
megmutattam a házat.
Eljöttél meglátogatni bennünket?*

2. csoport

*Eljön még a mi időnk!
Eljön a te időd.
Eljön mindennek az ideje.
Tudtam, hogy egyszer eljön ez a nap.
Eljött az én időm.
Eljött a nagy nap.
Eljött a búcsú ideje.*

3. csoport

*Eljössz velem Párizsba, vagy nem?
Eljössz velem moziba?*

4. csoport

Eljött értem apukám, és hazavitt.

5. csoport

Láttam, hogy nincs szükség rám, ezért eljöttem.

*Szerinted eljön?
Az anyja és az apja is eljött, hogy beszéljen
velünk.
Tegnap volt a születésnapja. Eljött hozzá az
anyja és az öccse.
Eljöttél meglátogatni a barátaidat?
Sokan eljöttek a bulira?
Szeretném, ha eljönne erre az útra.
Amikor eljött hozzám, már tudta, hogy el fog
utazni.
Sajnos nem tudtam eljönni.
El tudsz jönni?
Nem tudok eljönni.*

*Eljött a döntő pillanat.
Eljött a karácsony is.
Eljött a születésnapom... és nem kellett újra
ugyanazt kívánnom.
Eljött augusztus eleje, semmi. Eljött a közepe is,
de semmi.*

Ágnes is eljött velünk túrázni.

Eljött a gyűrűért.

Eljöttem otthonról.

A vizsgált korpuszok szerint az *eljön* legtöbb esetben a szótárban elsőként megadott ('a beszélő közelébe (...) vagy valamely szóban forgó helyre jön s oda megérkezik') jelentésben fordul elő, így a legtöbb példamondatban ezzel találkozunk a nyelvtanuló.

A második csoportba azok a mondatok tartoznak, amelyeknek alanya idővel kapcsolatos kifejezés: pl. *Eljött a nagy nap. Majd eljön a te időd* stb.

A harmadik és negyedik csoport az első szótári jelentés két alcsoportjának felel meg (ezeket az 1. ábrán az első jelentés alá soroltuk).

A szótárban másodikként megadott ('valahonnan útnak indul a beszélő felé, s így tartózkodási helyét elhagyja') jelentés a korpuszban igen ritka: *Láttam, hogy nincs szükség rám, ezért eljöttem. Eljöttem otthonról.*

A mondatokat tanulmányozva azt is megállapíthatjuk, hogy a szótár definícióit több helyütt is tovább pontosíthatjuk. Például az első csoport leírását kiegészíthetjük azzal a megfigyeléssel, hogy az itt felsorolt mondatok többsége valamilyen meghívásra adott válasz, reakció: *Megígérte, hogy eljön. Azt mondta, hogy eljön. Köszönöm, hogy eljött. Örülnénk, ha eljönnél* stb.

Az elemzés során az is világossá válik, miért választottuk külön a harmadik és a negyedik csoportot a korpuszból nyert információk alapján. Bár a szótár ezeket is az első jelentés alá sorolja, látjuk, hogy az ige bővítményei mindhárom jelentésben különbözőek, illetve a csoportok között más eltérések is feltárhatóak; ilyen például a jellemző mondatfajta és a pragmatikai jellemzők. Az *eljön valakivel valahova* gyakran – igaz, nem kizárólag – olyan jelen idejű kérdésekben fordul elő, amelyekkel elhívunk valakit valahova: *Eljössz velem*

moziba? Eljössz velem Párizsba? stb. Az *eljön valakiért valahova* jelentés pedig egyenesen új aspektussal bővíti az ige szemantikai tartományát: *Eljött értem apukám, és hazavitt. Eljött a gyűrűért.* Az *eljön* ige itt ugyanis nem a valahova való tartós megérkezést jelöli, mint az 1–3. csoport mondatainak esetében, hanem valaminek vagy valakinek egy bizonyos helyről való elhozatalát. Rövid ideig tartó cselekvést ír le, amelyet távozás követ. Az első csoportot a nyelvhasználati jellemzők világos prezentálása érdekében volt érdemes további alcsoportokra osztani.

Ezeket a jellemzőket a nyelvtanulók vonják le. Ehhez először tanulmányozzák a mondatokat, majd megpróbálják összegyűjteni az egyes csoportok jellemzőit. A megfigyelés irányítását kérdésekkel is segíthetjük: pl. *Milyen szó áll gyakran az ige előtt? Hol áll az ige a mondatban? Melyik igeidő/személy dominál? Miben hasonlítanak a csoport mondatainak alanyai? Hogyan írhatjuk körül az ige jelentését az egyes csoportokban? Milyen bővítményei vannak az igének?* stb. Ismét emlékeztetnék arra, mennyire fontos tudatosítani a nyelvtanulókkal, hogy a példamondatok gondos átolvasása legalább annyira lényeges, mint a csoportok jellemzőinek összegyűjtése és megfogalmazása, hiszen ez utóbbi művelet szükségszerűen leegyszerűsíti és leszűkíti a példából nyerhető információkat. A példák ismételt átolvasása újabb és újabb aspektusokkal gazdagíthatja az igeről való gyakorlati tudást, még akkor is, ha ennek a tudásnak nem minden részletét tudjuk explicit módon leírni. A munkafolyamat következő lépésében a tanulók által leszűrt jellemzőket összehasonlítjuk a pontosításokkal kiegészített szótári definíciókkal.

5.4. A példák rendszerezése: megjön

Hasonlóan járunk el a *megjön* ige esetében is. Ezzel az igével a Magyar Nemzeti Szövegtár és a Hunglish korpusza alapján az alábbi ötven példát választottuk, szemantikai kritériumok szerint a következőképpen csoportosítva:

1. csoport

Itt van, megjött a válasz.

Végre megjött az italunk.

Ha megjönnek az eredmények, kivel kell beszélnem?

Legyenek készen, mire megjön az engedély!

Délutánra szokott megjönni a kenyér.

Amikor megjött az értesítés, hogy az apja eltűnt, nagyon megijedt.

Megjöttek az első gólyák.

Nézzétek, gólyák! Megjöttek végre!

Na, megjött a láda Pestről.

Megjött a titkos üzenet.

Végre megjött a pizza!

És a pénz is megjön majd a jövő héten.

Na, jó, ülhetsz a fotelban, amíg megjön a taxi, de aludni nem engedlek.

Ha megjön a tavasz, elutazom Kiskunmajsára.

Megjött a fagy is.

Megjött az első hó.

Napok múlva megjött az eső.

Holnapra biztosan megjön a szél.

Mikó úr, megjött a felesége.

Úgy látom, megjött a gondnok.

Ha megjön Peti, mit mondasz neki?

Na végre, megjöttél!

Szevasz, Joci! Hát megjöttél?

Megjött a postás.

Megjött az orvos.

Végre megjött Rosemary.

Ilona néni, megjött a Pista!

Amíg a doktor úr megjött, ott maradtam Ilona néni mellett.

Itt van Süsü! Megjött Süsü! Sziasztok!

Jó, hogy megjöttél.

Csak hogy megjöttél!

Öt órát vártam, míg végre megjött.

Örülök, hogy megjöttél.

Amint a professzor megjön, értekezletet tartunk.

Menned kéne, mielőtt Harald megjön.

Beszélek a sráccal, ha megjön.

Elnézést, de megjött az újságíró.

Azt mondta, parkoljunk itt háromkor, és várjunk, amíg megjön.

2. csoport*Megjöttem!**Megjöttél?**Angéla, megjöttem!**Apám szeptemberben jött meg a frontról.**Reménykedtek, hogy majdcsak megjön a rég elment családtag.**Van, amikor két hét, van, amikor egy hét múlva jön haza, de azért megjön mindig.**Elmegy sétálni, de már jóval hat előtt megjön. Itthon családiasabb.**Mindig olyan lehetetlen időpontokban tudsz megjönni!***3. csoport***Házasodjunk össze, mielőtt megjön az eszed.**Végre megjött az önbizalma.**Hogy megjött a hangja!**Majd megjön a bátorságod.*

A *megjön* igét tartalmazó mondatokat a korpusz alapján alapvetően egy domináns és két ritkábban előforduló jelentés köré csoportosíthatjuk: (1) várt személy vagy esemény elérkezik, (2) valahonnan visszatér, megérkezik otthonába, (3) idiomatikus fordulatok. A példából az is világos, hogy tulajdonképpen mindhárom esetben arról van szó, hogy a beszélő várta az adott cselekvés bekövetkezését, egy adott személy vagy dolog érkezését. A jelentést a mondatokban igen gyakran előforduló *végre* határozószó is erősíti.¹ Ebben magyarázatot találunk a dominánsan múlt idejű előfordulásra is, hiszen a mondatok többsége egy-egy esemény bekövetkeztét, lezárulását adja hírül.

5.5. Példa a kollokációs vizsgálatokra

A Magyar Nemzeti Szövegtár keresőfelülete kollokációs vizsgálatokra is lehetőséget ad. Ezeket elvégezve pontosabb adatok nyerhetők arról, ami „a szó körül van,” azaz felállíthatjuk a vizsgált szót jellemzően kísérő szavak és kifejezések listáját, és ezek mondatbeli helyéről, azaz a szórendről (erről a magyar nyelvben olyan problémásnak tartott aspektusról) is levonhatunk következtetéseket. Például felállíthatjuk az igét jellemzően megelőző idő- és módhatározók, tipikus mondatkezdetek gyakorisági sorrendjét, és – legalább részben – leírhatjuk a velük járó szórendet:

eljött + alany (leggyakrabban személy)**eljön + alany (leggyakrabban személy)**ÖRÜLÖK, HOGY / JÓ, HOGY / KÖSZÖNÖM, HOGY *eljött + alany* (leggyakrabban személy)REMÉLEM, HOGY *eljön + alany* (leggyakrabban személy)HA *eljön + alany* (leggyakrabban idő)MAJD (egyszer) *eljön + alany* (leggyakrabban idő)VÉGRE *eljött + alany* (leggyakrabban idő)ÚJRA *eljön + alany* (leggyakrabban idő)LASSAN *eljön + alany* (leggyakrabban idő)**megjött + alany (személy vagy dolog)**VÉGRE *megjött + alany* (személy vagy dolog)HA *megjön + alany* (személy vagy dolog)MAJD(csak) *megjön + alany* (személy vagy dolog)MIRE *megjön + alany* (személy vagy dolog)

Fontos hangsúlyozni, hogy a fenti információk kizárólag korpuszból nyerhetők ki. Korpusz nélkül még az anyanyelvi beszélő sem rendelkezik megbízható mennyiségű és minőségű statisztikai adatokkal, valamint olyan eszközökkel, amelyek alapján a vizsgált lexikai elemmel gyakran együtt használt szavak listáját összeállíthatná. Márpedig – amint a cikk

¹ Az a tény, hogy a *megjön* szűkebb jelentésmezővel bír, magyarázatul szolgálhat arra, miért fordul elő ritkábban a korpuszokban, mint az *eljön*. A hipotézis megerősítésére további vizsgálatokra volna szükség.

elején utaltunk rá – éppen a gyakori szókapcsolatok, preferált kifejezőmódok jelenléte az, amely az anyanyelvi beszélő természetes nyelvhasználatát jellemzi (még ha ennek az anyanyelvi beszélő nincs is feltétlenül tudatában), és amelyekre a nem anyanyelvi beszélőnek a nyelvtanulás során olyan nagy szüksége van.

Elemzésünket tovább folytathatjuk, ha a rendelkezésünkre álló példákon további szintaktikai-grammatikai vizsgálatokat végzünk. Az egyszerű mondatok szerkezetét összehasonlítva látjuk például, hogy mindkét ige – ahogy sok más igekötős ige is – általában mondatkezdő (hangsúlyos) pozícióban áll: *Eljött a nagy nap. Megjött a pizza* stb.

Ha a mondatokat bővítjük, a gyakran használt határozók tekintetében ugyan különbségek vannak, ám ezek mondatbeli pozíciója és funkciója (legtöbbször mód- és időhatározó, amely az ige előtt áll) megegyezik: *ha / majd (egyszer) / lassan + eljön + alany* (személy vagy tárgy), *végre / ha / majd(csak) / mire + megjött + alany* (személy vagy tárgy). Az igékkel használt helyhatározók listája részben fedi egymást, a személyes névmás ragozott alakja (*hozzám* stb.) azonban jóformán csak az *eljön* igével használatos.²

A segédigékkel alkotott szerkezeteket vizsgálva kiderül, hogy leginkább a *tud* igét használjuk, elsősorban az *eljön* igével együtt: *El tudsz jönni? Nem tudok eljönni* stb., a *megjön* igével csak igen ritkán, speciális jelentésben fordul elő, mégpedig ha szemrehányást teszünk: *Mindig olyan lehetetlen időpontokban tudsz megjönni!*

Már ez a – teljesség igénye nélkül készített – leírás is mutatja, hogy más-más jelentés esetén nemcsak az ige szemantikai, de grammatikai, szintaktikai, sőt pragmatikai jellemzői is eltérnek.

Vizsgálataink alapján megállapíthatjuk, hogy bár a *megjön* és az *eljön* jelentése első látásra igen hasonló, a mondatban nagyon is eltérően működnek. Minél részletesebben térképezzük fel a két ige viselkedését, annál egyértelműbben választhatók el egymástól használati jellemzőik, és annál több fogódzót kap a nyelvtanuló, hogy a két igét megfelelő kontextusban tudja használni.

5.6. A munkafolyamat zárása: a rendszerezés

A munkafolyamat záró lépéseként egy rövidebb, rendszerezett áttekintést készítünk a jelentésekről. Ennek funkciója, hogy – csupán jelzés szintjén – felidézze azt a nagy mennyiségű példát, amelyekkel a nyelvtanuló az előző lépésekben dolgozott, és rögzítse a belőlük levont megfigyeléseket. Az áttekintéshez az értelmező szótár definícióit vehetjük alapul, amit a korpuszvizsgálat során nyert adatokkal egészítünk ki. Az *eljön* ige első jelentéséről és használati jellemzőiről készített összesítés ennek alapján így mutathat:

Személy valahova, ahol a beszélő van (általában meghívásra) ELJÖN.

ÖRÜLÖK, HOGY / JÓ, HOGY / KÖSZÖNÖM, HOGY *eljött* + alany

REMÉLEM, HOGY *eljön* + alany

Örülök, hogy eljöttél. Köszönöm, hogy eljöttek. Nem tudom, miért nem jött el. Ne haragudj, de nem tudok eljönni a bulira. Mikor jössz el hozzánk?

A rendszerező összefoglalás mindegyik csoport esetében hasonló módon építhető: (1) definíció, (2) jellemző szöveggörnyezet szörendi információkkal, (3) tipikus példák.

A nyelvtanuló az összesítéssel természetesen a példatárat is megőrzi, hiszen a két dokumentum együttesen tükrözi a tanulási folyamatot. A nyelvtanuló folyamatosan vissza tud kanyarodni a példákhoz, sőt a példatárat bővítheti is. A módszert elsajátítva pedig egy-egy felmerülő kérdéssel kapcsolatosan önállóan is végezhet egyszerűbb vizsgálatokat. Gyakorlás céljára további korpuszalapú mondatokkal könnyen készíthetőek feladatok is, amelyekben

² A *megjött hozzám* a Google keresőprogramjában 948 találatot mutatott, míg az *eljött hozzám* 56 700 találatot.

például az igét vagy az ige előtt álló szót kell kiegészíteni, a helyes igeidőt vagy igemódot kell eldönteni, szöveget kell írni megadott szókapcsolatokkal vagy mondatokkal stb.

6. Összegzés

Ebben az írásban azt kívántam bemutatni, hogyan segíthetnek a korpuszok a természetes nyelvhasználat jellemzőinek, különös tekintettel az igekötős igék használati tendenciáinak feltérképezésében és elsajátításában. A nagy nyelvi korpuszok – mint az anyanyelvi beszélők valóban elhangzott mondatainak gyűjteményei – megmutatják egy-egy szó előfordulási gyakoriságát és tipikus nyelvi környezetét. Annak szemléltetésére, hogy milyen információkat nyerhetünk ki ezekből az adatbázisokból, és ezeket milyen módszerrel dolgozhatjuk fel a magyaróra keretében, az *eljön* és a *megjön* igéket választottam.

A bemutatott munkafolyamatban a nyelvtanuló először a Magyar Nemzeti Szövegtár és a Hunglish korpuszából vett, nagy számú, jelentés szerint rendszerezett példát tanulmányozza, ezeken figyeli meg az adott szó környezetét, viselkedését. A használati tendenciákra, lehetséges jelentésekre vonatkozó tudnivalókat csupán ezután vonja le. Vagyis nem szabályokhoz rendel példákat, hanem példákhoz használati tendenciákat valamint egyéb – grammatikai, szintaktikai, esetenként pragmatikai – jellemzőket. A jelentések csoportosításához az értelmező szótár definícióit vettük alapul, de – amint láttuk – ezek kiegészítést, pontosítást igényelnek, és a gyakoriságot figyelembe véve sorrendjükön is változtatnunk kell. A bemutatott munkafolyamatban a nyelvtanuló olyan széles körű nyelvi tapasztalatot szerez, amelynek köszönhetően a vizsgált nyelvi elemet sikeresebben képes beépíteni nyelvhasználatába, és ezáltal fontos lépéseket tesz az anyanyelvi beszélőéhez hasonló nyelvi kompetencia megszerzése felé.

Irodalom

- Bárcki, Géza – Ország, László (szerk.) 1959–1962. *A magyar nyelv értelmező szótára*, I–VII. kötet. Akadémiai Kiadó, Budapest.
<http://mek.oszk.hu/adatbazis/magyar-nyelv-ertelmezo-szotara/elolap.php>
- Boulton, Alex – Tyne, Henry 2014. *Des documents authentiques aux corpus*. Collection langues et didactique. Les Édition Didier, Paris.
- Burnard, L. – McEnery, T. (eds.) 2000. *Rethinking language pedagogy from a corpus perspective*. Peter Lang, Frankfurt am Main.
- Carter, Ronald – McCarthy, Michael 2014. *Cambridge grammar of English*. Cambridge University Press, Cambridge.
- Conrad, Susan 2004. Corpus linguistics, language variation, and language teaching. In: Sinclair, John (ed.) *How to Use Corpora in Language Teaching. Studies in Corpus Linguistics*. John Benjamins Publishing Company, Amsterdam/Philadelphia. 67–85.
- Cowie, Anthony P. 1992. Multiword lexical units and communicative language teaching. In: Arnaud, Pierre J. L. – Henri Béjoint (eds.) *Vocabulary and applied linguistics*. Palgrave Macmillan, London. 1–12.
- Davis, Paul – Kryszewska, Hanna 2013. *The company words keep*. Delta Publishing, London.
- Dóla, Mónika 2014. Lexikon és grammatika kapcsolatáról – különös tekintettel az idegennyelv-tanulásra. *Hungarológiai Évkönyv* 15: 8–25.
- Gray, Bethany – Biber, Douglas 2016. *Grammatical complexity in academic English*. Cambridge University Press, Cambridge.
- Firth, John Rupert 1957. *Papers in Linguistics 1934–1951*. Oxford University Press, Oxford.
- Hoey, Michael 2005. *Lexical priming*. Routledge, London.
- Kiss, Jenő 1999. Nyelvi intuíció és elfogadhatósági ítéletek. In: *Magyar Nyelv* XCV/2: 129–137.
<http://www.magyar nyelv.c3.hu/99-2/kissj.html>

- Leńko-Szymańska, Agnieszka – Boulton, Alex 2015. Data-driven learning in language pedagogy. In: Leńko-Szymańska, Agnieszka – Boulton, Alex (eds.) 2015. *Multiple affordances of language corpora for data-driven learning*. John Benjamins Publishing Company, Amsterdam/Philadelphia. 1–14.
- Lewis, Michael 1993. *The lexical approach*. Language Teaching Publications, Hove.
- McCarthy, M. – O'Dell, Felicity 2016. *Academic vocabulary in use*. Cambridge University Press, Cambridge.
- O'Keeffe, Anne – McCarty, Michael – Carter, Ronald 2007. *From corpus to classroom*. Cambridge University Press, Cambridge.
- Pace-Sigge, Michael T-L. 2013. *Lexical priming in spoken English usage*. Palgrave MacMillan.
- Sinclair, John 1991. *Corpus concordance and collocation*. Oxford University Press, Oxford.
- Szirmai, Mónika 2005. *Bevezetés a korpusnyelvészetbe. A korpusnyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában*. Tinta Könyvkiadó, Budapest.

Felhasznált korpuszok

Hunglish: www.hunglish.hu

Magyar Nemzeti Szövegtár: <http://corpus.nytud.hu/mnsz/>

Szita, Szilvia

Those verbs with a prefix!

The use of corpora in the teaching of Hungarian as a foreign language

This article presents a possible classroom approach to verbs with a prefix, through the use of corpora. Language learners are considered to benefit significantly more from observing the behaviour of real-life utterances than from memorising abstract rules for their use. This is all the more relevant as the precise meaning of certain lexical items – and the way they are used by native speakers – cannot always be captured satisfactorily with rules alone. Examples taken from large linguistic databases and processed with simple statistical tools are, however, capable of providing first-hand information about the use of such items in real-life situations. Not only does an example-based approach allow us to analyse the occurring forms of words, but it also makes possible the observation of the context(s) in which they are used. The study of lists of examples, therefore, enables learners to refine and expand their understanding of how language actually works.

The present paper focuses on Hungarian verbs with a prefix since they seem to present extreme problems for language learners. As a matter of fact, in lack of a comprehensive linguistic description, prefixed Hungarian verbs are a frequent source of error for learners trying to employ them. For the present purpose, two verbs with similar meanings (*eljön* and *megjön* – lit. *arrive*) have been selected to illustrate the relevance and usefulness of the proposed corpus-based analysis. The method presented is aimed to provide Hungarian as a foreign language teachers with practical tools to help their learners build real-life linguistic experience and understand the insights that can be gained from the use of corpora.