

Pelcz Katalin

A *Csak természetesen!* speciális korpusz születése, tematikus szókincsének rövid vizsgálata

1. A korpusz kerete, a képzés rövid ismertetése

A *Csak természetesen! Modellalapú nyelvtanítás kezdő szinten* című online képzésnek az volt a célja, hogy bevezessen a modellalapú oktatás alapjaiba, s a magyar mint idegen nyelv oktatásához adjon átfogó támogatást a nyelvtanárok számára, különös tekintettel az A1-es szint oktatására. A kurzus általános oktatásmódszertani alapelveket és kezdő szinten átadandó ismereteket tárgyalt hat modulon keresztül. A 40 órás kurzus a magyar mint idegen nyelv tanításának fontosabb kérdéseit és részterületeit járta körül, melyet a résztvevők 12 hét alatt dolgozhattak fel.

A modulok oktatóvideókra épültek, majd minden videót rövid, az ismereteket ellenőrző teszt követett. A modulokhoz fórumok kapcsolódtak: az együtt gondolkodás és a tapasztalatcsere platformjai. A fórum felületére töltötték fel a résztvevők a modul elemeihez tartozó kifejtendő kérdésekre adott válaszokat is. Minden fórumot egy-egy szakember kísért, akik értékelték a válaszokat, és segítettek, támogatták a kurzus elvégzését. A modulokat értékelés zárta. Minden modulhoz kapcsolódott szakirodalmi ajánló is.

A kurzus a PTE Általános Orvostudományi Karán működő Nemzetközi Oktatási Központ munkatársaival együttműködésben készült. A munkát a Tempus Közalapítvány koordinálta, melynek Canva felületén zajlott a munka. A tartalom kidolgozója Szita Szilvia és Pelcz Katalin volt, a mentori munkát pedig Antal Zsófia, Arató Zsófia és Baumann Tímea végezték.

A képzés először a 2021/2022-es évben került meghirdetésre, majd a túljelentkezés, érdeklődés és a pozitív visszajelzések okán újra meghirdetésre került a 2022/2023-as tanévben. A képzés az első meghirdetés idején a Stipendium Hungaricum Ösztöndíjprogram keretén belül Magyarországon tanuló ösztöndíjasokat tanító nyelvtanárok munkáját volt hivatott segíteni, majd később külföldön dolgozó kollégák számára is elérhetővé vált az online kurzus. Az előző év tapasztalatait végiggondolva ugyan frissítettünk néhány elemet (például beépítettünk konzultációt is a képzési portfólióba), ugyanakkor a képzés célja, a kifejtendő kérdések tartalma és a feldolgozás módja változatlan maradt, így mindkét képzésen született szövegek bekerülhettek a *Csak természetesen!* speciális korpuszába.

2. A képzés tartalmának bemutatása

A következőkben a hat képzési modul tartalmát kívánom bemutatni. Az első, bevezető modul célja az volt, hogy meghatározzuk a képzés általános célkitűzéseit, keretét és lehetőségeit. Ebben a modulban részletesen áttekintettük, hogy milyen tanulási célokat szeretnénk elérni, és hogyan illeszkednek mindezek a résztvevők szakmai fejlődési céljaihoz. Megvizsgáltuk a képzés lehetőségeit a rendelkezésre álló erőforrások, a technológiai eszközök és az időbeli keretek tekintetében. Ezen kívül tisztáztuk az alapfogalmakat, és bemutattuk a képzés struktúráját, a modulok közötti kapcsolatokat, hogy a résztvevők átfogó képet kapjanak a teljes képzési folyamatról.

A második modul középpontjában a tematikus feldolgozás módszertana állt. Hogyan tanítsunk hatékonyan, a természetes nyelvhasználat szem előtt tartásával? Hogyan dolgozzunk nyelvi modellekkel? Hogyan dolgozzunk tematikusan felépített tananyaggal? Hogyan tanítsunk lexikát? Hogyan iskolázzuk a készségeket? Ebben a részben ismertettük, hogyan dolgozzuk fel a nyelvoktatás során az egyes témákat, és hogyan építjük be ezeket az oktatási folyamatba. Különös figyelmet fordítottunk a készségek fejlesztésére, a tematikus egységek logikai sorrendjére és az egymásra épülő ismeretek rendszerére. A modul célja az volt, hogy a résztvevők megértsék a tematikus feldolgozás alapelveit, és képesek legyenek alkalmazni azokat saját oktatási gyakorlatukban.

A harmadik modulban a résztvevők néhány kihívást jelentő grammatikai jelenséggel ismerkedtek meg. Hogyan tanítsuk a grammatika kezdő szinten elsajátítandó jelenségeit, és hogyan kapcsoljuk össze a grammatika gyakorlását értelmes kommunikációs helyzetekkel? A modul célja az volt, hogy a résztvevők mélyebb ismereteket szerezzenek a tipikusan kihívást jelentő grammatikai jelenségekről, és hatékony módszereket sajátítsanak el azok tanítására. Ebben a modulban kapott helyet a lexikogrammatikai megközelítés bevezetése (Biber és mtsai. 1998: 84), mely azt tűzi ki célul, hogy nyelvtan és tartalmat egységben tanítsunk. A gyakorlatok és példák segítettek a résztvevőknek abban, hogy magabiztosan alkalmazzák a tanultakat saját oktatási helyzeteikben.

A negyedik modul a kiejtés tanításának módszertanát tárgyalta. Hogyan fejlesszük hatékonyan a kiejtést? A modellalapú oktatás módszertét követve kiemelt fontossággal kezeljük a kiejtés oktatását. Ebben a részben bemutattuk a helyes kiejtés elsajátításának technikáit és gyakorlatait. A modul célja az volt, hogy a résztvevők pontos ismeretekkel rendelkezzenek a kiejtés fejlesztésének lehetőségei területén, hatékony módszereket sajátítsanak el a kiejtés javítására. Külön hangsúlyt fektettünk a fonetikai és fonológiai ismeretek alkalmazására, valamint a szöveghallgatási- és beszédgyakorlatokra, amelyek elősegítik a kiejtési készségek fejlesztését.

Az ötödik modulban a visszajelzések adásának, a tesztelésnek és a hibajavításnak a módszereit tárgyaltuk. Ebben a részben bemutattuk a hatékony visszajelzés elveit és technikáit, valamint a különböző tesztelési módszereket, amelyek segítenek a résztvevők nyelvi fejlődésének értékelésében. A modul célja az volt, hogy a résztvevők megtanulják, hogyan adjanak konstruktív visszajelzést, hogyan előzhetnek meg egyes hibákat, s hogyan javítsák ki hatékonyan a hibákat, valamint hogyan alkalmazzák a tesztelés eredményeit az oktatási folyamat javítására.

A résztvevők több lépcsőben adtak visszajelzést, értékelték a kurzust. Az értékelés során érdekes megfigyelni, hogy a képzés mely elemeit tartották a legfigyelemreméltóbbnak, ugyanis erre az értékelő kérdésre születtek a legszerteágazóbb válaszok. Például:

„Sok fontos útravalót kaptam, nehéz egyetlen egyet kiemelni. Nagyon tetszett annak hangsúlyozása, hogy a nyelv nem állandó szó- és szabálygyűjtemény. Ugyan a nyelvtan a nyelv szabályrendszere, tartószervezete, de a szabályokból nem kapunk használható nyelvtudást.” Vagy: *„...úgy érzem, ez a kurzus nagyon sokat segített és segít abban, hogy a magyar mint idegen nyelv tanításának napi gyakorlatában soha ne feledkezzem meg a nagy képről, hogy mit, miért és azt hogyan tegyem. A természetes nyelvhasználat szem előtt tartása, a modellek alkalmazása és a nagy nyelvi inputra való odafigyelés mellett a mindennapi munkámat nagyban segíti az a módszertani alapelv, hogy a téma bevezetése és feldolgozása során szekvenciában gondolkozzam, a feladatok lehetőség szerint mindig épüljenek egymásra.”*

3. A korpusz szövegeinek születése

A bemutatott tartalmakhoz kifejtendő kérdések kapcsolódtak, melyeket csoportban vagy egyedül dolgozhattak ki a résztvevők. A képzésnek 2021/22-ben 67 résztvevője volt, akik közül 39 vált adatközlővé. A következő kurzuson ez a szám nem változott, 2022/23-ban 67 résztvevőből 39 adatközlő szövegei kerültek be a korpuszba. Az adatközlők legnagyobb hányada több nyelvet beszélő, gyakorló nyelvtanár (1. ábra).



1. ábra: Néhány adat a kurzus 2021/22-es résztvevőiről

Mindkét képzés lezárultával elindult a korpuszépítő, adatgyűjtő munka is. A képzés résztvevői nyilatkoztak, hogy hozzájárulnak-e fórumbejegyzéseik, kérdéseik és válaszaik korpuszba emeléséhez. A gyűjtésben részt lehetett venni névvel vagy név nélkül. A korpuszépítés első lépése a nyers szövegek összegyűjtése és strukturált formába rendezése volt. A szövegeket változatlan formában egy táblázatba vittük fel, amely a modulok szerint tartalmazza az adatokat. Minden modult tovább tagoltunk a fórumkérdések és az arra adott reakciók szerint, ezáltal a szövegek átlátható és könnyebben kezelhető formában kerültek rögzítésre. Az így kialakított teljes táblázatot elküldtük az adatközlőknek, hogy annak tartalma később is elérhető legyen számukra, és időkeret nélkül tudják használni módszertani segédletként a munkájuk során. A korpuszépítés második lépése a nyers szövegek feldolgozása volt, amely több műveletet foglalt magában.

Elsőként a személyes adatok védelme érdekében az összegyűjtött szövegek anonimizálását végeztük el. Ennek során eltávolítottuk vagy átalakítottuk azokat a szövegrészeket, amelyek az adatközlők közül egyértelműen azonosíthatóvá tettek volna bárkit is. Az anonimizált szövegeket feltöltöttük a *Sketch Engine* felületére, amely automatikusan annotálta a korpuszt, elvégezte a szófajok szerinti címkézést, a tokenizációt és a lemmák megjelölését. A szövegekhez metaadatokat is hozzárendeltünk, ezek tartalmazzák például a szövegek keletkezési idejét, a

szervezők (anonimizált) azonosítóját és a fórumok témáját. A metaadatok hozzáadása lehetővé tette a szövegek kontextusának pontosabb megértését és a későbbi elemzések során történő hatékonyabb feldolgozását. A szövegeket végül egy ellenőrzési folyamatnak vetettük alá, hogy biztosítsuk a feldolgozás pontosságát és következetességét. Az ellenőrzés befejezése után a korpusz készen állt a további nyelvészeti elemzésekre és kutatásokra.

4. A speciális korpusz indokoltsága és jelentősége

Meggyőződésünk, hogy a nyelvtanítás és -tanulás elsődleges célja hosszú távon, hogy a diákokat elvezesse a minőségi, önálló nyelvhasználatig. E gondolatot is szem előtt tartva a kurzus módszertani segítséget adott ahhoz, hogy a diákokat hatékonyan segíthessük abban, hogy már az első lépéseiket ezzel a célkitűzéssel összhangban tegyék meg.

A kurzus a szerzők által kidolgozott modellalapú nyelvoktatás módszertana is azt a használatalapú nyelvészeti felfogást követi (Dörnyei 2009; Ellis 2002; Tomasello 2003), mely szerint nyelvtudás csak nyelvhasználatból születhet. A használatalapú nyelvfelfogás szerint, amikor találkozunk egy nyelvi példával, az agyunk összehasonlítja ezt a példát a korábbi példákkal, amelyekkel már találkoztunk. Ez azt is jelenti, hogy egy szó vagy kifejezés jelentését az határozza meg, hogy miként és milyen kontextusban találkoztunk vele, használjuk azt. Így a nyelvi fejlődés azoknak a találkozásoknak az eredménye, amelyek a nyelv használata közben alakulnak ki. Ahogy egy nyelvi elemmel egyre többször találkozunk, az ismétlések révén az elérésük gyorsabbá és hatékonyabbá válik. Ezért egy gyakran használt szó vagy szerkezet könnyebben és gyorsabban előhívható a memóriából. A használatalapú nyelvfelfogás hangsúlyozza, hogy a nyelvi tudás nem statikus szabályok gyűjteménye, hanem dinamikus rendszer, amely a nyelvhasználat során szerzett tapasztalatok alapján folyamatosan alakul és fejlődik. Ez a megközelítés komoly felelősséggel ruházza fel a nyelvtanárt: minden az órán szerzett nyelvi tapasztalat számít. Ugyanakkora ahhoz is vezérfonalat kapunk, hogyan tehetjük a nyelvórát a természetes nyelvhasználat helyszínévé.

A *Csak természetesen!* képzési program során született szövegek maguk is visszaforgathatóak célzott, tematikus korpuszá, melynek elemzése segíti a nyelvoktatást és a tanárképzést is.

A speciális vagy célzott korpuszok, egy adott nyelvi regiszterhez, témához vagy műfajhoz tartozó szövegeket gyűjtenek egybe. Olyan nyelvi adatbázisok, amelyek meghatározott regiszterű, témájú és műfajú szövegeket tartalmaznak. A speciális korpusz létrehozását az indokolja, hogy lehetővé teszik a specifikus nyelvi minták és használati sajátosságok tanulmányozását. Olyan vizsgálatokat tesznek lehetővé, amelyek célzottan szolgálnak – jelen esetben – a nyelvoktatás fejlesztését, s amelyeket egy általános korpusz nem feltétlenül tudna biztosítani. Ugyanakkor az általános és a speciális korpuszok közötti különbség nem jelent éles határt, hanem a különbség inkább céljukban rejlik. Egy általános korpusz célja, hogy bemutasson egy nyelvet vagy nyelvváltozatot, míg egy speciális korpusz arra szolgál, hogy egy adott témával kapcsolatos kérdésekre adjon választ. Az általános korpuszokat gyakran használjuk referenciakorpuszként a korpusznyelvészeti kutatásokban, hogy egy speciális korpuszsal összehasonlítva, például egy adott nyelvi elem gyakoriságában tapasztalható különbségeket figyeljük meg (Hunston 2002: 21–23), ahogy ez alább is megfigyelhető lesz.

McEnery és Wilson (2001) szerint a speciális korpuszok nélkülözhetetlen eszközök a nyelvészeti kutatásokban, mivel lehetővé teszik, hogy mélyreható és célzott vizsgálatokat végezzünk egy adott nyelvi kontextusban. Hozzájárulnak a nyelvi variabilitás és a kontextusfüggő nyelvhasználat jobb megértéséhez, lehetővé teszik a célzott nyelvi jelenségek részletes vizsgálatát, összehasonlítását és elemzését, hozzájárulva a nyelvi jelenségek mélyebb megértéséhez és a nyelvoktatás fejlesztéséhez. A speciális korpuszok felhasználhatók az oktatásban, mivel autentikus nyelvi adatokat biztosítanak a tananyafejlesztéshez, a nyelvtanároknak és a tanu-

lóknak egyaránt. Hunston (2002) hangsúlyozza, hogy a korpuszok autentikus nyelvi adatokat biztosítanak, amelyek gazdag forrást jelentenek a nyelvoktatás számára. Ezek az adatok segítenek a tanulóknak megérteni, hogyan használják a nyelvet valós helyzetekben. Az autentikus szövegek elemzése lehetővé teszi a tanulók számára, hogy találkozzanak a nyelv valódi használatával, szemben az idealizált vagy mesterségesen létrehozott példákkal.

5. A célkorpusz: Miért tanítunk, miért tanítanak magyarul?

Vizsgáljuk meg a következőkben a *Csak természetesen* korpusz egy elemét az első modulból. A választott rész az első modul 1.2. részéhez kapcsolódó kérdések és válaszok. A modul így épül fel:

- 1.1. A modul áttekintése. A tananyag előkészítéséről; célok, lehetséges módszerek a kezdő szinten. Módszertani alapfogalmak
- 1.2. **Miért tanítunk, miért tanítanak magyarul? Milyen reális célkitűzések várhatók el egy év alatt?**
 - Ön milyen ok(ok)ból tanul(t) idegen nyelveket?
 - Milyen szinten beszél ma ezeket a nyelveket, és mire használja őket?
 - Ön szerint mik lehetnek a reális célkitűzések magyar nyelvből egy kezdő nyelvkurzusnál, 90-120 tanóra alatt?
 - Mit gondol a kollégái meglátásairól?
- 1.3. Általános módszertani alapfogalmak: modellalapú nyelvtanítás, természetes nyelvhasználat

A kiválasztott rész 22.375 tokennek, 18.580 szóval 1097 mondat került feltöltésre a *Sketch Engine* rendszerébe; ez alkotja jelen vizsgálatunk tárgyát. Azért ennek a résznek az elemzésére fókuszálunk, mert a téma egyszerre készíttet önvizsgálatra, és válik didaktizálható tartalommal.

6. Tematikus alapszókincs feltérképezése

A választott rész szókincsének vizsgálatára teszünk kísérletet, a vizsgált részkorpusz alapszókincsét és többszavas egységeit fogjuk górcső alá venni. A tematikus alapszókincsnek feltérképezése azért hasznos a nyelvoktatás számára, mert egy adott téma kapcsán célzottan segíti a kommunikációt. A tematikus alapszókincsre épül a természetes nyelvhasználat, s a korpuszvizsgálaton alapuló szókincs használata segíthet a hatékonyabb kommunikációban, valamint a nyelvi félreértések minimalizálásában.

A tematikus alapszókincset olyan szavak halmazaként határozhatjuk meg, amelyek egy bizonyos témához kapcsolódnak, vagy ahhoz szükségesek (Sinclair 1991). Az alapszókincs egy bizonyos területhez, témakörhöz vagy egy adott terület szaknyelvéhez kapcsolódik: jelen esetben a nyelvtanulási célok és lehetőségek témájáról szóló szövegek képezik a vizsgálat tárgyát. A korpusz segítségével pontos információk nyerhetők a szavak gyakoriságáról, jelentéseiről és kollokációiról.

Az elemzés első lépéseként nézzük meg a *Wordlist* funkció alkalmazásával, hogy a célkorpuszon belüli szavak gyakorisága milyen eredményeket mutat (2. ábra). Megfigyeljük a szófajok szerinti gyakorisági mutatókat is. (Mindeközben kritikusan szemléljük a kapott eredményeket, hiszen a program nagyon sokféle vizsgálatot és megfigyelést tesz lehetővé, de „vét hibákat”, például a szavakként értelmezett írásjeleket, a *nyelv* és a *nyelvet* szavak külön listázását vagy a szófaji besorolást tekintve is.)

| Word | Frequency ? ↓ | Word | Frequency ? ↓ | Word | Frequency ? ↓ | Word | Frequency ? ↓ |
|------|---------------|-------|---------------|------|---------------|------|---------------|
| 1 | , | 1,913 | ... | 14 | de | 127 | ... |
| 2 | a | 1,350 | ... | 15 | (| 123 | ... |
| 3 | . | 1,108 | ... | 16 | - | 121 | ... |
| 4 | az | 664 | ... | 17 | tanultam | 117 | ... |
| 5 | és | 448 | ... | 18 | már | 111 | ... |
| 6 | is | 370 | ... | 19 | ez | 109 | ... |
| 7 | hogy | 271 | ... | 20 | idegen | 108 | ... |
| 8 | nem | 238 | ... | 21 | tanulni | 106 | ... |
| 9 | volt | 226 | ... | 22 | angol | 104 | ... |
| 10 | nyelvet | 172 | ... | 23 | nagyon | 103 | ... |
| 11 | egy | 166 | ... | 24 | " | 102 | ... |
| 12 | nyelvv | 149 | ... | 25 | : | 91 | ... |
| 13 |) | 134 | ... | 26 | német | 88 | ... |
| | | | | 27 | meg | 84 | ... |
| | | | | 28 | mert | 82 | ... |
| | | | | 29 | el | 78 | ... |
| | | | | 30 | kezdtam | 77 | ... |
| | | | | 31 | orosz | 76 | ... |
| | | | | 32 | csak | 75 | ... |
| | | | | 33 | így | 73 | ... |
| | | | | 34 | magyar | 69 | ... |
| | | | | 35 | általános | 63 | ... |
| | | | | 36 | szinten | 61 | ... |
| | | | | 37 | ezt | 59 | ... |
| | | | | 38 | mint | 58 | ... |
| | | | | 39 | kötelező | 58 | ... |
| | | | | 40 | még | 57 | ... |
| | | | | 41 | angolul | 54 | ... |
| | | | | 42 | ha | 51 | ... |
| | | | | 43 | iskolában | 51 | ... |
| | | | | 44 | vagy | 50 | ... |
| | | | | 45 | van | 49 | ... |
| | | | | 46 | egyetemen | 47 | ... |
| | | | | 47 | ami | 47 | ... |
| | | | | 48 | amit | 47 | ... |
| | | | | 49 | jó | 45 | ... |
| | | | | 50 | első | 45 | ... |

2. ábra: A célkorpusz szavainak gyakorisági mutatója

Láthatjuk, hogy a leggyakrabban használt szavak a szövegkohéziót szolgálják. Tanulságos továbbá, hogy milyen gyakori elem az *is*:

- a családommal *is* rendszeresen részt vettünk a különböző programokon
- az őseim között *is* akadnak szlovák felmenők
- a német nyelv *is* kicsit a mindennapjaim része

A választott példák hatékonyan iskolázzák például a tipikus szórend rögzítését is.

Meglepő, hogy a *nem* a nyolcadik helyen szerepel, amiből feltételezhetjük, hogy a témában sok olyan szövegrész szerepel, ami a nyelvtanulással kapcsolatban azt fogalmazza meg, mi az, amit nem sikerült megtanulni, elvégezni, megélni. Érdekes megfigyelni, hogy milyen gyakran követi egymást két olyan tagmondat, amelyben helyet kap a tagadás:

- *nem* volt könnyű, *nem* igazán kedveltem ezt a nyelvet
- *nem* voltam eléggé motivált, *nem* szerettem tanulni
- messze *nem* volt eredményes, ugyanis *nem* tudok olyan szinten kommunikálni
- *Nem* bántam meg, mivel ma már *nem* lehet nemzetközi szakirodalmat olvasni
- hogy a *nem* magyar családtagokkal megértessem magam, még ha *nem* is anyanyelvi szinten

A program listázza a téma leggyakoribb főneveit és igéit (3. és 4. ábra). A leggyakoribb főnév a várakozással megegyezően a *nyelv* szó, a témához kapcsolódó leggyakoribb ige a *tanul* (a *van* igét kizárhatjuk, ugyanis ez általában minden szövegfajta leggyakoribb igéje). E három lista mutatja meg számunkra a tematikus alapszókincset, melyet további összehasonlítás és vizsgálat alá vehetünk.

WORDLIST Csak természetesen!

Get more space

noun (1,064 items | 4,723 total frequency)

| Noun | Frequency ? ↓ | Noun | Frequency ? ↓ | Noun | Frequency ? ↓ | Noun | Frequency ? ↓ |
|----------------|---------------|-----------------|---------------|---------------|---------------|----------------|---------------|
| 1 nyelv | 576 ... | 14 motiváció | 40 ... | 27 tudás | 28 ... | 40 nyelvtanár | 18 ... |
| 2 év | 122 ... | 15 nyelvvizsga | 40 ... | 28 osztály | 27 ... | 41 diploma | 17 ... |
| 3 szint | 92 ... | 16 tanár | 38 ... | 29 nyelvtudás | 26 ... | 42 alap | 17 ... |
| 4 idegen | 80 ... | 17 lehetőség | 38 ... | 30 érdeklődés | 25 ... | 43 irodalom | 17 ... |
| 5 iskola | 70 ... | 18 óra | 37 ... | 31 rész | 24 ... | 44 nyelvtan | 17 ... |
| 6 egyetem | 62 ... | 19 idő | 37 ... | 32 eset | 22 ... | 45 tapasztalat | 16 ... |
| 7 nyelvtanulás | 57 ... | 20 szak | 36 ... | 33 ok | 21 ... | 46 kurzus | 16 ... |
| 8 kultúra | 56 ... | 21 kor | 35 ... | 34 nap | 19 ... | 47 kérdés | 16 ... |
| 9 tanulás | 52 ... | 22 munka | 34 ... | 35 élet | 19 ... | 48 sikerélmény | 16 ... |
| 10 általános | 49 ... | 23 szó | 31 ... | 36 világ | 19 ... | 49 kiejtés | 16 ... |
| 11 gimnázium | 48 ... | 24 kommunikáció | 31 ... | 37 kapcsolat | 19 ... | 50 cél | 16 ... |
| 12 sor | 48 ... | 25 ha | 29 ... | 38 b2 | 18 ... | | |
| 13 középiskola | 43 ... | 26 diák | 28 ... | 39 tanítás | 18 ... | | |

3. ábra: A célkorpusz főneveinek gyakorisági mutatója

WORDLIST Csak természetesen!

Get more space

verb (503 items | 2,925 total frequency)

| Verb | Frequency ? ↓ | Verb | Frequency ? ↓ | Verb | Frequency ? ↓ | Verb | Frequency ? ↓ |
|------------|---------------|----------------|---------------|---------------|---------------|-------------|---------------|
| 1 van | 464 ... | 14 jár | 26 ... | 27 kap | 15 ... | 40 hallgat | 12 ... |
| 2 tanul | 308 ... | 15 ért | 24 ... | 28 jön | 15 ... | 41 tart | 12 ... |
| 3 tud | 102 ... | 16 megtanul | 24 ... | 29 sikerül | 15 ... | 42 fejleszt | 11 ... |
| 4 használ | 91 ... | 17 tesz | 24 ... | 30 érez | 14 ... | 43 tölt | 11 ... |
| 5 kezd | 88 ... | 18 gondol | 23 ... | 31 ad | 14 ... | 44 megért | 11 ... |
| 6 beszél | 83 ... | 19 segít | 22 ... | 32 dolgozik | 14 ... | 45 elég | 11 ... |
| 7 szeret | 59 ... | 20 foglalkozik | 21 ... | 33 lát | 14 ... | 46 tetszik | 11 ... |
| 8 kell | 52 ... | 21 folytat | 20 ... | 34 ír | 14 ... | 47 jelent | 11 ... |
| 9 tanít | 44 ... | 22 vesz | 19 ... | 35 megy | 14 ... | 48 jut | 10 ... |
| 10 lesz | 35 ... | 23 elsajátít | 17 ... | 36 ismerkedik | 14 ... | 49 érzik | 10 ... |
| 11 választ | 32 ... | 24 olvas | 17 ... | 37 kerül | 14 ... | 50 megéri | 10 ... |
| 12 elkezd | 31 ... | 25 találkozik | 17 ... | 38 emlékszik | 12 ... | | |
| 13 él | 27 ... | 26 mond | 16 ... | 39 kommunikál | 12 ... | | |

4. ábra: A célkorpusz igéinek gyakorisági mutatója

A tematikus alapszókincs feltérképezésében értékes adatokkal szolgál, ha egy általános korpuszal össze tudjuk vetni a szavak relatív gyakoriságát. A relatív gyakoriság aránya jelzi, hogy egy adott szó vagy kifejezés milyen gyakran fordul elő a nyelvi anyagban, és mennyire jellemző az adott nyelvre általában. A relatív gyakoriság számítása a szó előfordulásainak számában és a korpuszban található összes szószámában rejlik. Ezután az eredményt százalékos értékben kapjuk meg, ami azt jelenti, hogy a szó vagy kifejezés a korpuszban található összes szóból hány százalékot tesz ki. A relatív gyakoriság lehetővé teszi, hogy megértsük, milyen gyakran használják egy adott szót vagy kifejezést a nyelvi anyagban, és hogy ez az előfordulási gyakoriság mennyiben jellemző a vizsgált nyelvre vagy szövegtípusra. Ha az alábbi táblázat utolsó oszlopának értéke magasabb, mint 100, akkor gyakrabban fordul elő az adott szó vizsgált

korpuszban, mint az általános korpuszban, ha pedig kisebb, akkor ritkábban. Az itt következő táblázatban csoportosítottam a szavakat, s kiemelem a különösen alacsony értéket mutató szavakat, tehát azokat a szavakat, amelyek más szövegösszefüggésben is gyakran használatosak.

Az alábbi táblázatban – valamint az azt követő összehasonlításban is – a *Skech Engine* felületén megtalálható 5.164.717.029 szímszámú *HunTenTen* általános korpuszsal vetjük össze a *Csak természetesen! 1.2* tematikus alkörpuszát (1. táblázat).

| Item | Frequency (focus) | Frequency (reference) | Relative frequency (focus) | Relative frequency (reference) | Score |
|----------------|-------------------|-----------------------|----------------------------|--------------------------------|---------------|
| nyelvtanulás | 57 | 12458 | 2547,4861 | 3,94001 | 515,887 |
| nyelvvizsga | 40 | 30253 | 1787,7095 | 9,56792 | 169,258 |
| nyelv | 576 | 661520 | 25743,018 | 209,21463 | 122,465 |
| motiváció | 40 | 48968 | 1787,7095 | 15,48679 | 108,493 |
| | | | | | |
| középiskola | 43 | 63030 | 1921,7877 | 19,93409 | 91,85 |
| gimnázium | 48 | 109868 | 2145,2515 | 34,74724 | 60,04 |
| egyetem | 62 | 494602 | 2770,9497 | 156,42456 | 17,608 |
| iskola | 70 | 967812 | 3128,4917 | 306,08362 | 10,191 |
| tanulás | 52 | 144970 | 2324,0225 | 45,84872 | 49,628 |
| kultúra | 56 | 371098 | 2502,7932 | 117,36475 | 21,153 |
| tanár | 38 | 360382 | 1698,324 | 113,97568 | 14,78 |
| | | | | | |
| tanít | 44 | 240446 | 1966,4805 | 76,0443 | 25,537 |
| tanul | 308 | 604087 | 13765,363 | 191,05067 | 71,681 |
| | | | | | |
| angol | 221 | 530652 | 9877,0947 | 167,82587 | 58,511 |
| orosz | 133 | 377556 | 5944,1343 | 119,40719 | 49,375 |
| spanyol | 62 | 205966 | 2770,9497 | 65,13953 | 41,911 |
| német | 149 | 696427 | 6659,2178 | 220,25446 | 30,102 |
| olasz | 62 | 295753 | 2770,9497 | 93,53588 | 29,322 |
| francia | 49 | 422261 | 2189,9441 | 133,54575 | 16,284 |

1. táblázat: A szavak relatív gyakoriságának összehasonlítása a célkorpuszban és az általános korpuszban

7. A többszavas egységek vizsgálata

A leggyakoribb főnevek és igék tanulmányozása után a többszavas egységek tematikus vizsgálatának szentelünk figyelmet, mivel ezek fontos szerepet játszanak a nyelv természetes használatában és a jelentés kialakításában. A *többszavas egységek* a korpusznyelvészetben sokat kutatott téma: azoknak a szócsoportoknak a vizsgálata, amelyek a véletlen valószínűsénél statisztikailag nagyobb valószínűséggel fordulnak elő egy szövegben (vö. Sinclair 1991; Hoey 1991: 5–6). Ezeknek a többé-kevésbé stabil formáknak, általában együtt előforduló szókapcsolatoknak (kollokációk, idiómák, állandó szókapcsolatok) a vizsgálata alapvető fontosságú a nyelv természetes használatának megértéséhez; kulcsszerepet játszanak a jelentés kialakításában és a nyelvi kommunikációban, ezért jelentős hatással van a nyelvtanulásra.

Az *N-grams* funkció használatával vizsgálhatjuk a korpusz többszavas kifejezéseit (5. ábra). Az egyszerű kereséssel megadjuk, hogy hány elemre szeretnénk keresni, és a funkció használatával megkapjuk a vizsgált korpusz leggyakoribb kollokációit:

The screenshot shows the N-GRAMS tool interface. At the top, there is a search bar with the text "Csak természetesen!" and a search icon. Below the search bar, the title "3-4-grams, word" is displayed, followed by "(Items: 24, total frequency: 167)". The main content is a table with three columns: "N-gram" and "Frequency?". The table lists 24 items, each with a number in the first column, the N-gram text, and its frequency. The items are sorted by frequency in descending order.

| | N-gram | Frequency ? | | N-gram | Frequency ? | | N-gram | Frequency ? |
|----|--------------------------|-------------|----|------------------------------|-------------|----|-------------------|-------------|
| 1 | kezdttem el tanulni | 17 ... | 11 | magyar mint idegen nyelv | 6 ... | 21 | nyelv és irodalom | 5 ... |
| 2 | ezt a nyelvet | 13 ... | 12 | az orosz nyelv | 6 ... | 22 | az orosz nyelvvel | 5 ... |
| 3 | a mai napig | 11 ... | 13 | az angol nyelvet | 6 ... | 23 | a munkám során | 5 ... |
| 4 | az angol nyelv | 8 ... | 14 | mint idegen nyelv | 6 ... | 24 | Az első idegen | 5 ... |
| 5 | mert kötelező volt | 8 ... | 15 | a mindennapi kommunikációban | 6 ... | | | |
| 6 | magyar mint idegen | 7 ... | 16 | az általános iskolában | 6 ... | | | |
| 7 | az orosz volt | 7 ... | 17 | Az általános iskolában | 6 ... | | | |
| 8 | első idegen nyelv | 7 ... | 18 | kötelező volt az | 5 ... | | | |
| 9 | kezdttem angolul tanulni | 6 ... | 19 | a magyar nyelvet | 5 ... | | | |
| 10 | a német nyelv | 6 ... | 20 | Az első idegen nyelv | 5 ... | | | |

5. ábra: A célkorpusz többszavas egységeinek gyakorisági mutatója

Ha pedig az általános nyelvi korpuszal összevetésben vizsgáljuk a gyakori szókapcsolatokat, a következő eredményt kapjuk. Ez az elemzés arra hívja fel a figyelmet, hogy melyek azok a kifejezések, melyek témaspecifikusak, de gyakran fordulnak elő további szövegekben is. A kulcsszóvizsgálattal összehasonlíthatjuk a célkorpusz szókincsét a referencia-korpusz szókincsével (6. ábra). Ennek alapján azonosítjuk azokat a szavakat, amelyek szignifikánsan gyakrabban fordulnak elő a célkorpuszban, mint a referencia-korpuszban.

reference corpus: Hungarian Web 2012 (huTenTen12) (Items: 3,756)

| Term | Term | Term | Term |
|-------------------------------|----------------------------------|--|----------------------------------|
| 1 nyelv tanulás ... | 14 második idegen nyelv ... | 27 német nyelv ... | 40 iskola ötödik osztály ... |
| 2 első idegen nyelv ... | 15 szakirodalom olvasás ... | 28 nyelv elsajátítás ... | 41 másik nyelv ... |
| 3 első idegen ... | 16 anyanyelvű diák ... | 29 kezdő szint ... | 42 egyetemi tanulmányaim sor ... |
| 4 második nyelv ... | 17 középfokú nyelvvizsga ... | 30 angol nyelv tanulás ... | 43 másik idegen nyelv ... |
| 5 idegen nyelv ... | 18 munkám sor ... | 31 kötelező tárgy ... | 44 b2 szint ... |
| 6 közvetítő nyelv ... | 19 belső motiváció ... | 32 adott nyelv ... | 45 nyelv tanítás ... |
| 7 mindennapi kommunikáció ... | 20 jó szint ... | 33 olasz nyelv ... | 46 kultúra iránti érdeklődés ... |
| 8 orosz nyelv ... | 21 nyelv megtanulás ... | 34 tanulmányaim sor ... | 47 hétköznapi kommunikáció ... |
| 9 új nyelv ... | 22 magyar szak ... | 35 angol nyelv ... | 48 francia nyelv ... |
| 10 spanyol nyelv ... | 23 stipendium hungaricum ... | 36 angol b2 ... | 49 orosz szak ... |
| 11 harmadik idegen nyelv ... | 24 nyelvek iránti érdeklődés ... | 37 c1-es szint ... | 50 ország nyelv ... |
| 12 harmadik idegen ... | 25 francia nyelv tanulás ... | 38 orosz tanulás ... | |
| 13 második idegen ... | 26 finn nyelv ... | 39 általános iskola ötödik osztály ... | |

6. ábra: A célkorpusz többszavas egységeinek gyakorisági összehasonlítása az általános korpuszal

Az ilyen többszavas egységek beépítése a nyelvtanulásba hozzájárul a fluens kommunikációhoz és a megértéshez (O'Keeffe és mtsai. 2007: 77). A Hoey-féle *Lexical Priming* elmélet szerint amikor egy szót elsajátítunk, az mentálisan „primerizálódik” a kollokációs használatra. Michael Hoey *Lexical Priming: A New Theory of Words and Language* (2005) című könyvének első három fejezetében leírja, hogy amikor egy szót elsajátítunk, azt nemcsak a jelentésével együtt tanuljuk meg, hanem a kollokációival, grammatikai környezetével, szövegtípusokkal és pozíciójával is. A „primerizálás” lényege, hogy a szavak ismételt használata révén mentális mintázatok alakulnak ki, amelyek alapján bizonyos kontextusokban természetesen használjuk őket. Tehát egy szó által, hogy ismételten találkozunk, különböző kontextusokban és szavakkal együtt, fokozatosan feltöltődik az összefüggésekkel és társításokkal. Így, amikor egy szót megtanulunk, célravezető, ha az együtt fordul elő más szavakkal különböző kontextusokban. Ezek az ismétlődő társulások alakítják a szó használatának módját. Ebből következően sokkal hatékonyabb, ha a tanított szókapcsolatok számát növeljük a tanítás során, s érdemes mindig szövegben is megvizsgálni ezeket az elemeket, mivel a szavak és szókapcsolatok használatának helye a szövegben is meghatározó. Az ismételt találkozások során egyre több információt jegyzünk meg a szóról. Megjegyezzük a **kollokációit**, hogy milyen más szavakkal fordul elő gyakran, illetve a **kolligációit** is, hogy milyen nyelvtani szerkezetekben jelenik meg, valamint a **szemantikai és pragmatikai asszociációit**. Emellett még a szavak jellegzetes diskurzusbeli megjelenéséről is gyűjtünk információkat, és azt is azonosítjuk, hogy mely nyelvi regiszterben, stílusokban vagy szituációkban használjuk jellemzően az adott szót, szókapcsolatot (Hoey 2005: 9–13). Tehát a szavak használatát nemcsak a szótári jelentésük határozza meg, hanem az is, hogy milyen más szavakkal, milyen kontextusokban találkozunk velük ismételten.

Mivel a nyelvórán a találkozások száma korlátozott, ezért jól átgondoltnak kell lennie minden megnyilvánulásnak explicit magyarázatnak, példának. Lényeges a több szóból álló egységek kiemelése, a használati gyakoriságára vonatkozó információk beépítése, a tanulási folyamat és gyakori nyelvhasználati lehetőségek biztosítása esetén lehet megfelelően hatékony (Dörnyei 2009: 207).

Az alapszókinszavainak kontextusát további módszerekkel tárhatjuk fel. A konkordancia-keresés lehetővé teszi, hogy egy adott szó vagy kifejezés összes előfordulását megtaláljuk a korpuszban, és lássuk azokat közvetlen kontextusukban. Az így nyert konkordanciasorok vizuális elrendezésük révén is segítik a megfigyelést: a középen elhelyezkedő, pirossal jelölt elem körül jobbra és balra meghatározott számú karakter található, amelyek általában nem alkotnak teljes mondatokat, de további kattintást követően megjeleníthetjük a hosszabb szövegrészt is. A konkordanciasorok segítségével kirajzolódnak a használati mintázatok, és megfigyelhetjük a vizsgált elem szemantikai asszociációit (Hoey 2005). Hasonló jelentésű szavak hasonló mintázatok mutatnak, és a mintázatok eltérései gyakran a jelentésbeli különbségekre utalnak (Sinclair 1991).

Tehát a kulcsszavak azonosítását követően lehetőségünk van arra, hogy megvizsgáljuk a szavak előfordulásának kontextusát. Két példát hozok erre, a *nyelvtanulás* és a *tanul* szavak konkordanciasorait (7. és 8. ábra):

| | |
|---|---|
| való szoros kötődésem miatt sokszor gondolok arra, hogy újra nekiállok a nyelvtanulásnak , de az igazi nagy elhatározás még várat magára. | Második ide |
| zni egy ösztöndíjprogrammal, ami szintén ösztönzően hatott, hiszen már a nyelvtanulás egy eléggé korai szakaszában anyanyelvi közegbe, célnyelvi kultúrába | |
| al és a nyelvekért (meg a nyelvtudásomból) élek, a kezdeti találkozásaim a nyelvtanulással nem voltak pozitívak. | Az első tanult nyelvem 10 éves koromtól |
| ni, még a kazah hallgatók kedvéért sem. | Különböző okaim voltak a nyelvtanulás kapcsán. |
| lás kapcsán. | Az általános, valamint középiskolában kötelező volt a nyelvtanulás . |
| számomra új nyelvet tanulni, főleg felnőttként, hiszen újra szembesültem a nyelvtanulás során felmerülő nehézségekkel, problémákkal. | Angolul és fran |
| retek a nyelvbe vagy a kultúrába. | Ettől függetlenül fokozatosan a nyelvtanulás által az új nyelvek és kultúrák "hatása alá kerültem". |
| Ezek megértése tovább tudott motiválni, hogy komolyabban is folytassam a nyelvtanulást , miközben erőteljesen bővíthetett tovább a szókincesem. | Tehát |
| gismerése, kultúrák és kulturális tartalmak megismerése/ megértése volt a nyelvtanulás célja. | Szerencsére magánóra keretein belül is tanulhattam a ny |
| lizációnak köszönhetően heti 5*45 percben. | Ebben a két esetben a nyelvtanulás még csupán egy kezdődő érdeklődés kezdete volt, illetve kötelező rés |
| lva, hogy "Annyi ember vagy, ahány nyelven beszélsz" folyamatosan fűti a nyelvtanulás szeretetét. | Először németet tanultam az általános iskola 2. oszt |
| ir saját elvárásaimnak akartam megfelelni, s ezért igyekeztem elmélyülni a nyelvtanulásban . | Külföldi munkatapasztalataimnak köszönhetően, a nyelvek és |
| ir saját elvárásaimnak akartam megfelelni, s ezért igyekeztem elmélyülni a nyelvtanulásban . | Külföldi munkatapasztalataimnak köszönhetően, a nyelvek és |
| ra lépve nyílt meg a "szabad idegen nyelv választás" lehetősége. | A nyelvtanulás egyébként mindig is vonzott, szüleim bátorítására végül kéttannyelvű c |
| al és a nyelvekért (meg a nyelvtudásomból) élek, a kezdeti találkozásaim a nyelvtanulással nem voltak pozitívak. | Az első tanult nyelvem 10 éves koromtól |
| ek. | Hallgatóink oktatásában a néhány évvel ezelőtt elkezdett angol nyelvtanulással szerzett ismereteim, illetve spanyol nyelvi ismereteim segítenek (ez ut |
| ra elutazni a Szovjetunióba. | Tehát már a kezdetekben megfogott a nyelvtanulás . |
| használattal célzó vállalkozás volt. | Eszter: Mindenekelőtt a kötelező nyelvtanulás miatt kezdtem el oroszul tanulni az általános iskola 4. osztályában. |
| le érettségi vizsgát tenni. | Az egyetemen tovább folytattam a német nyelvtanulást . |
| | Többször jártam Németországban. |
| | Hangsúlyosabban |

7. ábra: Bepillantás a célkorpuszba, néhány példa a *nyelvtanulás* konkordanciasorai közül

8. Összefoglalás

Egy olyan témához kapcsolódó speciális korpuszt mutattam be, mely a nyelvtanárok és a nyelvtanulók számára egyaránt releváns. A korpusz vizsgálata tanulságokkal szolgál a lexika megfigyeléséről, tanításáról is. A témával összefüggésben arra világítottam rá, hogy a lexika tanításának hatékony módja az, ha a tanulók valós nyelvi kontextusokban tanulják meg a szavakat és a kifejezéseket. Ebből következik, hogy (1) a lexikai egységeket nem izoláltan, hanem valódi példamondatokban, szövegekben, kommunikációs helyzetekben és szituációkban mutatjuk be, s (2) a tananyag didaktizálása során is érdemes olyan szövegeket alapul venni, amelyek a valós nyelvhasználaton, autentikus szövegek elemzésén alapulnak.

Irodalom

- Biber, Douglas – Conrad, Susan – Reppen, Randi 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge University Press, Cambridge.
- Biber, Douglas 2009. A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14 (3): 275–311.
- Dörnyei Zoltán 2009. *The Psychology of Second Language Acquisition*. Oxford Applied Linguistics. Oxford University Press, Oxford.
- Ellis, Nick C. 2002. Frequency effects in language processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition* 24 (2): 143–188.
- Hoey, Michael 1991. *Patterns of Lexis in Text*. Oxford University Press, Oxford.
- Hoey, Michael 2005. *Lexical priming: A new theory of words and language*. Routledge, Abingdon, England.
- Hunston, Susan 2002. *Corpora in Applied Linguistics*. Cambridge University Press, Cambridge.
- HuTenTen: *Corpus of the Hungarian Web* (é.n.). <https://www.sketchengine.eu/hutentenhungarian-corpus>
- McEnery, Tony – Wilson, Andrew 2001. *Corpus Linguistics: An Introduction*. Edinburgh University Press, Edinburgh.
- O’Keeffe, Anne – McCarthy, Michael – Carter, Ronald 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press, Cambridge.
- Sinclair, John 1991. *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- Tomasello, Michael 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge, MA.

Pelcz, Katalin

Creating a special corpus for language learning and a short study of its thematic vocabulary

The paper introduces a specialized corpus relevant to language teachers and learners. The corpus consists of writings from participants of the online teacher training course titled *Naturally! Model-based Language Teaching at Beginner Level*. Over two years, 78 participants developed and submitted their responses to the course's essay questions, resulting in a specialized corpus that explores various issues related to language education. This study focuses on a portion of the corpus that is particularly applicable to language teaching. The article details the process of creating the corpus and demonstrates how to uncover basic vocabulary and multi-word units related to the subject matter. The analysis of the corpus provides insights into the observation and teaching of the lexicon. It highlights that an effective method of teaching vocabulary involves learners acquiring words and expressions within real-life language contexts.