

Pécsi Tudományegyetem Bölcsészettudományi Kar

Nyelvtudományi Doktori Iskola

Alkalmazott Nyelvészeti Doktori Program



Vermeki Boglárka

## **Korpuszépítéstől a tananyagfejlesztésig**

**Korpusznyelvészeti vizsgálatok és az eredmények felhasználása  
egy gyermekeknek szóló tananyagban**

Doktori (PhD) értekezés

Témavezető: Dr. Pelcz Katalin

Pécs

2023

## Köszönetnyilvánítás

Közel négy éve kezdtem el az értekezésben bemutatott kutatással és a dolgozat írásával foglalkozni. Ez alatt az idő alatt rengeteg támogatást és segítséget kaptam, ezért itt szeretném megragadni az alkalmat, hogy köszönetemet fejezzem ki mindazoknak, akik valamilyen módon hozzájárultak a disszertáció megvalósulásához.

Nem tudom eléggé kifejezni, hogy mennyire hálás vagyok témavezetőmnek, *Dr. Pelcz Katalinnak*, aki számos értékes ötlettel, tapasztalataival, segítőkész hozzáállásával és tanácsaival jelentős mértékben hozzájárult a disszertációm elkészüléséhez. Mindenkinek ilyen témavezetőt kívánok és remélem, hogy egyszer lesz lehetőségem a tőle tanultak szerint segíteni mások munkáját.

Szeretném kifejezni köszönetemet és hálámat *Dr. Szita Szilviának* és ismét *Dr. Pelcz Katalinnak*, akik hosszú idő óta inspiráltak a munkájukkal, támogattak, és a Korpusznyelvészeti és Szakmódszertani Munkacsoport (KorSzak) megalapításával kiemelkedő segítséget nyújtottak a disszertációm megszületéséhez.

Köszönöm a segítséget a *KorSzak tagjainak*, különösen *Arató Zsófiának* és *Dr. Szabó Veronikának*, valamint *édesanyámnak*, akik nagyban hozzájárultak a felvételek elkészüléséhez, valamint *az adatközlő gyermekeknek*, akik nélkül nem jöhetett volna létre a KorSzak Gyermeknyelvi Korpusz.

Nagyon köszönöm *Kasikci Katalinnak* a sok közös munkát, a kreatív ötleteit és a baráti támogatását.

Szívből köszönöm *Domonkos Dóra* értékes baráti támogatását, és hogy mindig számíthatok rá.

Nagyon hálás vagyok *Dr. Schmidt Ildikónak*, hogy számos értékes tanáccsal látott el. Köszönöm a szakmai támogatását, a sok közös munkát.

Emellett szeretném kifejezni a hálámat minden volt tanáromnak, különösen *Dr. Nádor Orsolyának*, aki a magyar mint idegen nyelv felé terelt az egyetemi éveim alatt. A *Károli Gáspár Református Egyetem angol szakán* oktatóknak, akiknél elsajátíthattam a tudatos tananyagkészítés alapjait, valamint rengeteg szakmódszertani ismeretet.

Szeretném megköszönni a *családomnak*, *édesanyámnak*, *húgomnak* és *öcsémnek*, valamint *barátaimnak* az állandó támogatást és biztatást.

# Tartalom

|   |    |
|---|----|
| <b>I. Bevezetés</b> .....   | 1  |
| 1. Az értekezés hipotézisei.....  | 2  |
| 2. A kutatás célja.....   | 3  |
| 2.1 Miért szükséges egy új gyermekeknek szóló magyar nyelvi tananyag?.....                                    | 4  |
| 3. Kutatási kérdések.....   | 7  |
| 4. Néhány kulcsfogalom meghatározása.....   | 9  |
| 4.1 Formulaszerű nyelvhasználat, formulaszerű szekvenciák, formulaszerű egységek ..                           | 9  |
| 4.2 Természetes nyelvhasználat, anyanyelvi beszélő, kompetens nyelvhasználó .....                             | 10 |
| 4.3 Nyelvhasználati mintázatok .....  | 11 |
| 4.4 A pedagógiai korpusz .....  | 12 |
| 5. A doktori értekezés szerkezete.....  | 13 |
| 5.1 Az értekezésben idézett fordításokkal, rövidítésekkel és hivatkozásokkal<br>kapcsolatos információk ..... | 14 |
| <b>II. Elméleti kutatási keret</b> .....  | 17 |
| 1. Bevezetés az elméleti kutatási keret meghatározásához.....   | 17 |
| 2. A használatalapú nyelvelméleti modellek .....  | 18 |
| 2.1 Gyermekek anyanyelv-elsajátítása a használatalapú elméletek alapján .....                                 | 20 |
| 2.2 Nyelvi észlelés és nyelvhasználat a használatalapú nyelvelméleti keretben .....                           | 22 |
| 2.3 A használatalapú elméletek a nyelvoktatásban.....   | 24 |
| 3. A lexikai előfeszítés elmélete .....   | 26 |
| 3.1 Pszicholingvisztikai előzmények.....  | 26 |
| 3.1.1 Szematikai előfeszítés.....   | 27 |
| 3.1.2 Ismétléses előfeszítés .....  | 29 |
| 3.2 Korpusznyelvészeti előzmények.....  | 30 |
| 3.2.1 Lexika, grammatika és szemantika kapcsolata a korpusznyelvészeti kutatások<br>alapján.....              | 31 |
| 3.3 A lexikai előfeszítés elmélete és legfontosabb kutatásai .....  | 32 |
| 3.3.1 A lexikai előfeszítés a beszélt nyelvben .....  | 36 |
| 3.3.2 Lehet-e vizsgálni a primingot korpusznyelvészeti eszközökkel? .....                                     | 37 |
| 3.4 A lexikai előfeszítés jelentősége a nyelvoktatásban .....   | 37 |
| 4. A formulaszerű nyelvhasználat.....   | 39 |
| 4.1 A formulaszerű egységek jellemzői.....  | 40 |

|   |     |
|---|-----|
| 4.2 A formulaszerű nyelvhasználat jelentősége a nyelvoktatásban .....   | 45  |
| 5. A beszédfolyékonyság .....   | 48  |
| 5.1 A beszédfolyékonyság meghatározása .....  | 48  |
| 5.2 A beszédfolyékonyság kutatásának mai megközelítései .....   | 52  |
| 5.2.1 A Segalowitz-féle hármass modells.....  | 52  |
| 5.2.2 Beszédproduktios modellek .....   | 54  |
| 5.2.3 Automatizálás .....   | 57  |
| 5.3 A beszédfolyékonyság kurrens meghatározása .....  | 58  |
| 5.4 A beszédfolyékonyság a nyelvoktatásban .....  | 60  |
| 5.4.1 A beszédfolyékonyság kapcsolata az első és másodikk nyelvvel.....   | 61  |
| 5.4.2 A beszédfolyékonyság értékelése a nyelvvizsgán .....  | 63  |
| 6. A korpusznyelvészetről.....  | 67  |
| 6.1 A korpusz definíciója .....   | 67  |
| 6.1.1 A korpusz reprezentativitása és mérete.....   | 68  |
| 6.1.2 A korpuszok típusai .....   | 70  |
| 6.2 Gyermeknyelvi korpuszok .....   | 72  |
| 6.2.1 CHILDES.....  | 74  |
| 6.2.2 Magyar gyermeknyelvi korpuszok .....  | 78  |
| 6.3 A pedagógiai célú korpuszok.....  | 81  |
| 6.3.2 Gyermekek számára készült pedagógiai korpuszok .....  | 82  |
| 6.3.3 Korpusznyelvészeti kutatások gyermekek számára készült pedagógiai korpuszon.....  | 86  |
| 6.4 A korpusznyelvészet módszer vagy elmélet? .....   | 88  |
| 6.5 A korpusznyelvészet alkalmazása a nyelvoktatásban.....  | 89  |
| 6.5.1 Korpuszinformált tananyagok.....  | 95  |
| 6.5.1.1 A korpuszok felhasználási lehetőségei a tananyagok írásánál.....  | 97  |
| 6.5.2 Korpuszinformált tananyagok a magyar mint idegen nyelv oktatásának területén.....   | 101 |
| 6.5.2.1 A hasznos lexika kiválasztása a tananyag számára.....   | 101 |
| 6.5.2.2 A tanított nyelvi formák arányának és sorrendjének kialakítása, egy szintezett lexikogrammatikai tanmenet elkészítése .....                   | 101 |
| 6.5.2.3 A témákhoz kapcsolódó lexika, választvariációk és szövegtípusok kapcsolatának beépítése a tananyagba .....                                    | 102 |
| 6.5.2.4 Valós és a tanulók számára releváns szituációk beépítése a tananyagba. ....   | 104 |
| 6.5.2.5 Saját intuíción helyett a korpuszadatokra támaszkodva készített, a természetes nyelvhasználatot tükröző szövegek beépítése a tananyagba ..... | 105 |

|   |     |
|---|-----|
| <b>III. A kutatás</b> .....   | 108 |
| 1. A KorSzak Gyermeknyelvi Korpusz .....  | 108 |
| 1.1 Néhány szó a Korpusznyelvészeti és Szakmódszertani Munkacsoportról .....                                | 108 |
| 1.2 A KorSzak Gyermeknyelvi Korpusz építésének célja és folyamata .....                                     | 109 |
| 1.2.1 A témakörök.....  | 110 |
| 1.2.2 A felvételek módja.....   | 113 |
| 1.2.3 A természetes nyelvhasználat megfigyelhetőségének kérdése .....                                       | 114 |
| 1.3 A KorSzak Gyermeknyelvi Korpusz általános adatai és jellemzői .....                                     | 115 |
| 1.4 A KorSzak Gyermeknyelvi Korpusz összeállításának módja .....  | 117 |
| 1.4.1 A beszélt nyelvi szövegek átírása .....   | 117 |
| 1.4.2 Annotáció.....  | 118 |
| 2. A kutatás módszertana .....  | 119 |
| 2.1 Sketch Engine .....   | 122 |
| 2.1.1 Korpuszépítés a Sketch Engine felületén.....  | 122 |
| 2.1.2 Korpuszeszközök a Sketch Engine felületén .....   | 126 |
| 2.1.3 A kutatás kvantitatív részéhez használt eszközök .....  | 127 |
| 2.1.3.1 Wordlist – gyakorisági listák készítése.....  | 128 |
| 2.1.3.2 N-grams – több elemből álló szekvenciák keresése .....  | 131 |
| 2.1.3.3 Word Sketch – kollokációk és szókombinációk keresése.....   | 133 |
| 2.1.3.4 Word Sketch Difference – két nyelvi elem kollokációk általi<br>összehasonlítása .....               | 137 |
| 2.1.3.5 Text type analysis – a teljes korpusz statisztikájának áttekintése a<br>szövegtípusok alapján ..... | 140 |
| 2.1.4 A kutatás kvalitatív részéhez használt eszközök .....   | 141 |
| 2.1.4.1 Concordance – konkordanciasorok készítése .....   | 142 |
| 2.2 Praat .....   | 152 |
| 3. A formulaszerű egységek beazonosítása .....  | 156 |
| 3.1 Formulaszerű egységek vizsgálata az agglutináló nyelvekben.....   | 157 |
| 3.1.1 A finn nyelvvel kapcsolatos kutatások .....   | 158 |
| 3.1.2 A török nyelvvel kapcsolatos kutatások .....  | 160 |
| 3.1.3 A magyar nyelvvel kapcsolatos kutatások .....   | 162 |
| 4. A jelen kutatásban meghatározott feltételrendszer .....  | 165 |
| 4.1 Gyakoriság .....  | 165 |
| 4.2 A formulaszerű egységek mérete .....  | 166 |
| 4.3 Funkció és jelentéstartalom.....  | 167 |

|  |            |
|--|------------|
| 4.4 Feldolgozás és produkció.....  | 168        |
| 5. Kutatásetika.....   | 169        |
| 5.1 A korpuszépítéssel, a korpuszépítőkkal kapcsolatos etikai irányelvek .....                       | 169        |
| 5.2 A beszélt nyelvi korpuszok adatközlőivel kapcsolatos etikai irányelvek .....                     | 170        |
| 5.3 A korpusz megosztásával, hozzáférhetőségével kapcsolatos etikai irányelvek.....                  | 171        |
| 5.4 A korpuszban összegyűjtött adatokat felhasználókkal kapcsolatos etikai irányelvek .....          | 172        |
| <b>IV. A kutatás menete és eredményei .....</b>  | <b>173</b> |
| 1. Bevezetés .....   | 173        |
| 2. Egyelemű lexémák.....   | 175        |
| 2.1 A szógyakorisági vizsgálatok eredményei.....   | 175        |
| 2.2 Szófajonkénti (főnév, melléknév, ige) eredmények .....   | 183        |
| 3. A két- vagy többelemű lexikai egységek.....   | 190        |
| 3.1 A leggyakoribb két- vagy többelemű lexikai egységek.....   | 196        |
| 3.1.1 Kételemű lexikai egységek .....  | 196        |
| 3.1.2 Háromelemű lexikai egységek.....   | 197        |
| 3.1.3 Négyelemű lexikai egységek .....   | 198        |
| 4. A KorSzak Gyermeeknyelvi Korpuszban található gyakori lexikai egységek használati mintázatai..... | 201        |
| 4.1 Az <i>ugye</i> használati mintázatai.....  | 201        |
| 4.2 A NEM TUD használati mintázatai.....   | 205        |
| 4.3 A MI és TUD használati mintázatai .....  | 210        |
| 4.4 Az <i>annyira nem</i> használati mintázatai .....  | 214        |
| 4.5 A ILYEN és OLYAN használati mintázatai .....   | 216        |
| 4.6 Az <i>igazából</i> használati mintázatai .....   | 222        |
| 5. A két- és többelemű lexikai egységek hosszabb egységekbe történő beágyazódása...226               |            |
| 5.1 Az <i>én is</i> beágyazódása .....   | 226        |
| 5.2 Az <i>így vagyok</i> és az <i>úgy vagyok</i> beágyazódása .....                                  | 228        |
| 5.3 A <i>mit tudom én</i> beágyazódása.....  | 231        |
| 6. A két- és többelemű lexikai egységek szövegben betöltött szerepük szerint .....                   | 235        |
| 6.1 Álláspont kifejezése .....   | 235        |
| 6.2 Diskurzusszervező kifejezések .....  | 238        |
| 6.3 Hivatkozás kifejezése.....   | 240        |
| 6.4 A lexikai egységek szövegben betöltött szerepeinek összefoglalása.....                           | 242        |
| <b>V. A kutatási eredmények felhasználása .....</b>  | <b>244</b> |

|  |            |
|--|------------|
| 1. A Magyar felfedezőről.....  | 244        |
| 2. Hogyan használható a KorSzak Gyermeknyelvi Korpuszon végzett korpuszalapú kutatás a tananyag elkészítése során? .....                         | 245        |
| 2.1 A hasznos lexika kiválasztása a tananyag számára.....  | 245        |
| 2.2 A tanított nyelvi formák arányának és sorrendjének kialakítása, egy szintezett lexikogrammatikai tanmenet elkészítése .....                  | 251        |
| 2.3 A témákhoz kapcsolódó lexika, válaszvariációk és szövegtípus kapcsolatának beépítése a tananyagba.....                                       | 252        |
| 2.4 Valós és a tanulók számára releváns szituációk beépítése a tananyagba.....   | 255        |
| 2.5 Saját intuíciók helyett a korpuszadatokra támaszkodva készített, a természetes nyelvhasználatot tükröző szövegek beépítése a tananyagba..... | 256        |
| 3. Néhány gondolat a projektalapú nyelvtanításról.....   | 262        |
| 4. A tananyag szerkezete.....  | 265        |
| <b>VI. Összegzés és lezárás .....</b>  | <b>268</b> |
| Referenciák .....  | 273        |
| Korpuszok .....  | 288        |
| Ábrák és diagrammok jegyzéke.....  | 289        |
| Táblázatok jegyzéke.....   | 292        |
| English summary .....  | 295        |
| Mellékletek .....  | 330        |
| Melléklet A: A Korszak Gyermeknyelvi Korpuszban lévő kételemű lexikai egységek .....   | 330        |
| Melléklet B: A Korszak Gyermeknyelvi Korpuszban lévő háromelemű lexikai egységek.....  | 343        |
| Melléklet C: A Korszak Gyermeknyelvi Korpuszban lévő négyelemű lexikai egységek .....  | 359        |
| Melléklet D: Szünetek helye (Praat vizualizációk).....   | 361        |
| Melléklet E: Példa egy konkordanciasorra (VAN EGY ILYEN) .....   | 375        |
| Melléklet F: A szülői hozzájáruló nyilatkozat .....  | 376        |
| Melléklet G: Példa a Magyar felfedezőhöz készített kifejezésbankokból.....   | 377        |
| Melléklet H: Példa a Magyar felfedezőhöz készített Tanári (A1) kézikönyvből .....  | 392        |
| Melléklet I: Példa a Magyar felfedezőhöz készített kiegészített feladatlapokból (B1).....  | 394        |

## I. Bevezetés

A korpusznyelvészeti kutatások az utóbbi évtizedekben egyre nagyobb figyelmet kaptak, egyre szélesebb körben használják őket az alkalmazott nyelvészet különböző területein. Ez annak köszönhetően bebizonyosodott, hogy a nagyszámú adatokra támaszkodó elemzések számos haszonnal járnak a nyelvhasználat megfigyelésére irányuló kutatások során. Ugyanakkor elmondható az is, hogy a korpuszok alkalmazása nem korlátozódik kizárólag a nyelvtudományra, jelentősége a nyelvoktatásban is egyértelművé vált. Többek között azért, mert a bennük található nyelvi minták megismerésével a diákok olyan nyelvi szerkezeteket tanulhatnak meg, amelyeket az anyanyelvi beszélők valóban használnak a mindennapi kommunikációjuk során. Az életszerű nyelvi példák felhasználása révén jobban megérthetik, hogy bizonyos nyelvi elemeket valójában hogyan és milyen kontextusban használnak. Ezáltal az ő nyelvhasználatuk is természetesebbé és pontosabbá válik.

A korpuszok valós nyelvi adatok széles köréhez biztosítanak hozzáférést, amelyek megfigyelése és empirikus elemzése olyan tananyagok létrehozását segíti, amelyek a témákat tekintve relevánsabban a nyelvtanulók számára, a bennük található szövegek pedig a természetes nyelvhasználatot tükrözik. Az így készült tananyagok a nyelvtanulók motivációjához is hozzájárulnak, valamint vonzóbbá tehetik a tanulási folyamatot is. A fent említett okok alapján megállapítható, hogy a korpuszok és a korpuszinformált tananyagok alkalmazása a gyermekek nyelvtanítása során is különösen előnyös lehet.

Az alkalmazott nyelvészeti kutatások, ideértve a korpusznyelvészeti vizsgálatok is, arra törekszenek, hogy a kutatás során nyert eredményeket a való életben felmerülő problémák megoldására használják fel. Ennek megfelelően, célkitűzésem már a doktori tanulmányaim kezdetétől fogva az volt, hogy olyan kutatást végezzek, amely hasznos lehet a nyelvoktatás területén, s olyan disszertációt készítek, mely párhuzamosan mutatja be a kutatást és annak gyakorlati felhasználását. Ennek következtében a jelen korpusznyelvészeti kutatás célja a vizsgálatok megalapozását követően az eredmények egy magyar mint idegen nyelvi tananyag formájában történő megjelenése, amely segítheti mind a magyarul tanuló gyermekek, mind a nyelvtanárok munkáját.



## 1. Az értekezés hipotézisei

Azt feltételezzük, hogy a jelen doktori értekezés a korpusznyelvészet szempontjából értékes hozzájárulást jelenthet az agglutináló nyelvek korpuszalapú kutatásainak bővítéséhez. Az alkalmazott nyelvészet és különösen a korpusznyelvészet területén az angol nyelv kutatása a leggyakoribb, beleértve a jelen kutatásban is vizsgált formulaszerű nyelvhasználatot. Ez a kutatás példaként szolgálhat arra is, hogyan lehet összeegyeztetni és alkalmazni az angol nyelvre kifejlesztett kutatási módszereket és paradigmákat egy agglutináló nyelv vizsgálatával.

Emellett azt is meg kell jegyeznünk, hogy a gyermekek nyelvhasználatát vizsgáló kutatások általában a nyelvfejlődést helyezik előtérbe. Azonban tananyagkészítés céljával készülő, a 11–15 éves korosztály nyelvhasználatában megfigyelhető formulaszerű egységek kutatása nívumnak számít a magyar nyelv esetében. Ezért a jelen kutatás jelentős hozzájárulást jelenthet az ilyen vizsgálatok bővítése terén is. Továbbá az értekezésben bemutatott pedagógiai korpusz építési folyamatának kidolgozása és ismertetése további kutatások számára nyújthat segítséget.

A nyelvoktatás szempontjából azt feltételezzük, hogy egy az anyanyelvi beszélők természetes nyelvhasználatát tükröző tananyag létrehozása érdekében nem elegendő a saját megérzéseinkre és nyelvtudásunkra támaszkodnunk (McCarthy – McCarten, 2022). Ennek megvalósítása érdekében célravezető módszer lehet, ha először pedagógiai korpuszt építünk, amely a tananyagkészítés céljával jön létre. A tudományos és tananyagfejlesztő munka megkezdésének első lépéseként Azt gondoljuk, kiemelkedően fontos a célcsoporthoz hasonló korú anyanyelvi beszélők nyelvhasználatának megfigyelése és az általuk preferált beszélgetési témák megismerése, amelyek feltételezéseink szerint a nyelvtanulók számára is relevánsak lehetnek. A releváns témák kiválasztása pedig kulcsfontosságú a nyelvtanulók, jelen esetben a gyermekek motivációjának fenntartása és a hatékony tanulási folyamat biztosítása érdekében (McCarten, 2010; McCarthy – McCarten, 2022). A fenti feltételezéseknek megfelelő kutatási módszerek alkalmazása hozzájárulhat egy olyan tananyag létrehozásához, amely tükrözi a természetes nyelvhasználatot és a nyelvtanulók igényeihez is igazodik.

## 2. A kutatás célja

A doktori értekezésben bemutatott korpuszalapú kutatás során magyar anyanyelvű gyermekek természetes nyelvhasználatának megfigyelésére és vizsgálatára töreksem abból a célból, hogy az eredményeket felhasználva egy korpuszinformált tananyag készülhessen magyarul tanuló gyermekek számára. A kutatás Hoey (2003, 2004, 2005) lexikai előfeszítés elméletét alapul véve, a gyermekek spontánbeszédében megjelenő formulaszerű lexikai elemekre (vö. Wray, 2002, 2019) koncentrálok, amelyek az első nyelv elsajátításában kiemelkedő szereppel bírnak, valamint rendkívül hasznosak a nyelvtanulók számára is. A kutatás célja az adatközlő gyermekek nyelvhasználatában lévő formulaszerű egység általános feltérképezése, valamint nyelvhasználatukra jellemző lexikai egységek használati mintázatainak megismerése.

A jelen kutatás során egy olyan pedagógiai célú beszélt nyelvi korpuszra – a KorSzak Gyermeknyelvi Korpuszra – támaszkodtunk, amelyet kifejezetten tananyagkészítés céljából a gyermekek nyelvhasználatának megfigyelésére és elemzésére terveztek és hoztak létre, így a 2020-ban kezdődött korpuszépítés folyamatának bemutatása is fontos részét képezi az értekezésnek. Meg kell jegyeznünk azonban azt is, hogy a korpuszépítés nem célja a kutatásnak, csupán egy eszköz ahhoz, hogy az erre épülő kutatás eredményeinek felhasználásával egy hatékony tananyag készüljön gyermekek számára. Fontos megemlítenünk még, hogy nemcsak a magyar mint idegen nyelv, hanem más nyelvek oktatásának területén sem elterjedt olyan tananyagok készítése, amelyek a gyermekek nyelvhasználatát vizsgáló korpuszalapú kutatásokra épülnek. Olyan pedagógiai célú korpuszokban sem bővelkedünk, amelyeket gyermekek nyelvoktatása során lehetne használni. Bár léteznek kimondottan a gyermekek számára szerkesztett pedagógiai korpuszok (lásd például SACODEYL, CorpusMate), az ezek használatával történő nyelvoktatáshoz vagy egyáltalán nem készültek, vagy csak korlátozott mennyiségű feladatlap érhető el. A korpuszokban található szövegeket nem didaktizálták, nem készült belőlük komplex tananyag.

## 2.1 Miért szükséges egy új gyermekeknek szóló magyar nyelvi tananyag?

A magyar mint idegen nyelv oktatása az elmúlt években jelentős átalakuláson ment keresztül. Az egyre növekvő igények, a célcsoportok életkorának változása és az ezzel járó kihívások új módszerek és eszközök alkalmazását igénylik. Annak ellenére, hogy ma már egyre több gyermek tanul magyarul, megállapítható, hogy a számukra szóló tananyagok aránya még ma is kevésbé jelentős a felnőtteknek készütekkel összehasonlítva (Vermeki, 2021b). Számukra készült korpuszinformált tananyag pedig a tudomásom szerint még nem létezik. A nyelvtanítás során szerzett tapasztalataim alapján megállapítható, hogy a tananyag biztosítása szempontjából a legnagyobb kihívást a pedagógusok számára jelenleg a fiatalabb tinédzser (11–14 éves) korosztály jelenti, mert a gyermekeknek szóló tananyagok már nem, a felnőtteknek készütek pedig még nem relevánsak számukra. Az említett korcsoportnak különösen fontos lenne az is a nyelvtanulás során, hogy megismerkedjenek velük egykorú célnyelven beszélő társaik nyelvhasználatával. A jelen kutatás célja a fent említettek miatt az, hogy egy olyan korpuszinformált tananyag készüljön az eredményeinek felhasználásával, amely ennek a korosztálynak a magyar mint idegen nyelv oktatásában jól használható, és amely kifejezetten a magyar anyanyelvű gyermekek nyelvhasználatát veszi figyelembe.

Amikor a magyar nyelvet tanulók célcsoportjáról beszélünk, gyakran hajlamosak vagyunk csak két kategóriát megkülönböztetni: gyermekek és felnőttek. A jelenleg forgalomban lévő gyermekek számára készült tankönyvek és tananyagok (1. táblázat) jelentős része így egy nagyobb korcsoport (például 8–14 éves vagy 10–18 éves korosztály), vagyis a gyermekek igényeit kívánja lefedni (Vermeki, 2021ab). Ugyanakkor meg kell jegyeznünk, hogy a gyermekek korcsoportot több okból sem lehet homogén csoportnak tekinteni (vö. Pinter, 2011), így ennek a csoportnak az igényeit megoldhatatlanul nagy kihívás lenne egyetlen tankönyvben teljesen kielégíteni. Ennek a fejezetnek nem célja, hogy részletesen beszámoljon a különböző korosztályok idegen, másod- vagy származásnyelvvél való kapcsolatáról, ezért erre nem is térnék ki, azonban azt fontos megjegyeznünk, hogy a gyermek-idegennyelvpedagógia kutatásait figyelembe véve, ha csak két tényezőre koncentrálnunk például a tanulási stratégiára (globális – analitikus) és a motivációra (belső – külső) (Kovács, 2009: 20–25), már akkor is legalább három részre (6–10 évesek, 11–14 évesek, 15–18 évesek) kellene osztanunk az iskoláskorú gyermekek korcsoportját.

**1. táblázat** Magyar mint idegen nyelvi és származásnyelvi tananyagok életkoronként és nyelvi szintenként (Vermeki, 2021b: 156)

| 6-10 éves                                      | 11-14 éves                                     | 15-18 éves  | 18+ éves   |
|--|--|---|--|
| Kiliki 1. (7–11 éves) A1–A2                    |  | Hungarian the Easy Way (nem meghatározott, témájából ítélve: 15–18+ éves) A1–A2 |  |
|  | Kiliki 2. (10–14 éves) B1–B2                   |   |  |
|  | Zahra és Zia (8–14 éves) A1–A2                 |   |  |
|  | Balassi-füzetek: Kalandra fel! (10–18 éves) A1 |   |  |
| Balassi-füzetek (7–14 éves) <sup>1</sup> B1–B2 |  |   | Halló, itt Magyarország<br>Hungarolingua<br>Lépésenként magyarul<br>MagyarOK<br>Multikulti |

Véleményem és tanítási tapasztalataim szerint a legkritikusabb korosztály, amelynél problémát jelent, hogy nem készült még kifejezetten számukra, az ő érdeklődési körüket megcélzó, módszertanában hozzájuk igazodó komplex tananyagcsomag, az a fiatalabb tinédzser korosztály, vagyis a 11–14 évesek. Azok a tanárok, akik hasonló korú gyermekeket tanítanak vagy a fiatalabb korosztálynak íródott tankönyveket használhatják, vagy az idősebbek számára készült könyvek részleteit, valamint saját fejlesztésű feladatlapjaikat tudják bevinni az óráikra. Ez több szempontból is megnehezíti a munkájukat. Egyrészt rendkívül időigényes az órákra való felkészülés, ha minden alkalommal ők állítják össze a feladatlapokat. Egy nyelvtanártól nem is lenne elvárható, és nem is tartozik a feladatai közé, hogy minden alkalomra magas színvonalú, spirális visszatérést biztosító, teljes tanulási szakaszt felölelő, a tanulást szisztematikusan támogató tananyagot szerkesszen. Másrészt pedig nehézséget okozhat a tanulók motivációjának fenntartása, ha számukra nem érdekes, kisgyermeknek vagy felnőtteknek szóló anyagokat használnak az órán (Vermeki, 2021b).

A tinédzser korosztály sokszor jelent kihívást a nyelvtanárok számára, de ha megfelelően motiváljuk őket, a kisgyermeknél és a felnőttéknél is nagyobb lehetőségek rejlenek a tanításukban (Ur, 2012: 286–301). Ahogy minden korosztálynál, náluk is rendkívül fontos, hogy aktív részesei lehessenek a tanulási folyamatoknak, de ez sem

<sup>1</sup> Miénk a vár!, Ünnepeljünk együtt!, Mesélj nekem!, Játssz velünk! (magyar származású gyermekek számára készült füzetek)

történik meg, ha nem megfelelően motiváltak. Motiválásukhoz pedig elengedhetetlen, hogy számukra releváns témákkal foglalkozzunk, amelyekhez tudnak kapcsolódni. Az ő tanításuk azért is igényel másfajta megközelítést, mert a gondolkodásuk egyre jobban hasonlít a felnőttek analitikus gondolkodásához, szeretnek komolyabb témákról is beszélgetni, valamint korábbi tanulmányaikra is támaszkodhatunk az órákon, miközben továbbra is élvezik a játékokat és a kreatív feladatokat, ha az megfelelő kihívást jelent számukra. Éppen emiatt a kettősség miatt érdemes olyan tanulóközpontú módszereket alkalmazni a nyelvórán, amelyek során jelentéssel bíró feladatok megvalósításán keresztül tanulhatják a nyelvet. Következésképpen kijelenthetjük, hogy a fenti okokból nem elégséges számukra egy kisgyermekeknek vagy egy felnőtteknek szóló tankönyv, hanem szükség van egy kifejezetten számukra készült tananyag létrehozása, amely ezekhez a tanulók – és tanárok – által elvárt igényekhez és a szükségleteikhez tud alkalmazkodni.

### 3. Kutatási kérdések

A doktori értekezésben bemutatott korpuszalapú kutatás (vö. Szirmai, 2005; Stefanowitsch, 2020; Hunston, 2022) tehát a KorSzak Gyermeknyelvi Korpuszra alapul, és legfontosabb célkitűzése, hogy feltérképezze az adatközlő gyermekek nyelvhasználatában lévő leggyakoribb formulaszerű egységeket (vö. Wray, 2002; Wray – Fitzpatrick, 2008; Wray, 2019), valamint tanulmányozza ezeknek a lexikai egységeknek a használati mintázatait (vö. Hoey, 2005), majd az eredmények felhasználásával korpuszinformált magyar nyelvi tananyag készüljön gyermekek számára. A korpuszban található lexikai egységek megfigyelésére azért van szükség, mert ezeknek a gyakori lexikai egységeknek fontos szerepük van nemcsak a nyelvelsajátításban, hanem a nyelvtanulásban is (vö. Tomasello, 2003, 2013, 2015; Bybee, 2008; Ortega et al., 2016; Ellis et al., 2016; Pelcz, 2023), hiszen az ismeretük és megfelelő alkalmazásuk által fluensebbé válik a nyelvhasználatunk (vö. Tavakoli, 2020).

1. Melyek a leggyakoribb egyelemű lexémák a KorSzak Gyermeknyelvi Korpuszban?
2. Milyen a leggyakoribb egyelemű lexémák szófaji (főnév, melléknév, ige) eloszlása a KorSzak Gyermeknyelvi Korpuszban?
3. Melyek a leggyakoribb két- vagy többelemű lexikai egységek a KorSzak Gyermeknyelvi Korpuszban?
4. Hogyan használják a KorSzak Gyermeknyelvi Korpuszban szereplő gyermekek a kiválasztott lexikai egységeket? Milyen használati mintázatok figyelhetők meg a korpuszban?
5. Hogyan ágyazódnak be a két- vagy többelemű lexikai egységek hosszabb szekvenciákba? A beágyazódott lexikai egységek milyen használati mintázatot mutatnak?
6. Mi a szövegben betöltött szerepe a KorSzak Gyermeknyelvi Korpuszban lévő két- és többelemű lexikai egységeknek?

A KorSzak Gyermeknyelvi Korpuszon végzett kutatás tehát a fenti hat kutatási kérdésre fókuszál. A kérdésekből láthatjuk, hogy a kutatás során kvantitatív vizsgálatokat kvalitatív elemzések követnek annak érdekében, hogy minél átfogóbb képet nyújtsanak a vizsgált

jelenségekről. A kutatás több szekvenciából tevődik össze, amelyek alapvetően két meghatározott pontból indultak ki: az egyelemű lexémák és a két- vagy több elemből álló lexikai egységek vizsgálatából. Az első két kérdés az egyelemű lexémákkal foglalkozik, és arra keresi a választ, hogy melyek a leggyakoribbak a korpuszban, illetve, hogy a gyakoriságukat tekintve vannak-e különbségek az egyes szófajok között. A harmadik kérdés a két- vagy többelemű lexikai egységekre összpontosít, és a leggyakoribb ilyen egységek meghatározására irányul. A negyedik és az ötödik kérdés a lexikai egységek használati mintázataira és beágyazódási viszonyaira koncentrál, kiemelve egy-egy nyelvhasználati példát, amely jellemzően megjelenik a korpuszban szereplő gyermekek beszédprodukcióiban. Végül a hatodik kérdés az adott lexikai egységek szerepét és használatát vizsgálja a szövegekben. Ezeknek a kutatási kérdéseknek a megválaszolása segíti a korpuszban található 11–14 éves gyermekek nyelvhasználatának jobb megismerését, és hozzájárul a magyarul tanuló gyermekeknek szóló tananyag elkészítéséhez.

A korpuszalapú kutatások hagyományosan a kvantitatív kutatómódszertant alkalmazták, mivel az elsődleges céljuk az volt, hogy az emberi intuíció korlátait meghaladva (ld. Stefanowitsch, 2020: 15–17), számokkal igazolható adatokat szolgáltatassanak a nyelvészek számára. Ugyanakkor a kapott eredmények megfelelő értelmezése és hatékony felhasználása érdekében szükséges a kvalitatív kutatómódszertan alkalmazása is (Martinez, 2020: 211). Ezt figyelembe véve, az átfogóbb kép elérése érdekében a kutatás egy magyarázó szekvenciális tipológia alapján történt (vö. Dörnyei, 2007), amely során kvantitatív vizsgálatokat követően kvalitatív elemzéseket végeztem.

## **4. Néhány kulcsfogalom meghatározása**

Az értelmezési eltérések és a terminológiai bizonytalanságok gyakran előfordulnak a tudományos szövegekben, különösen akkor, ha a fogalmak definíciójáról van szó. A pontos és egyértelmű fogalmi keretek meghatározása azonban kulcsfontosságú a kutatásokban, hiszen ezek határozzák meg a vizsgálati eredmények érvényességét és általánosíthatóságát (vö. Paltridge – Phakiti 2015). Ebben az alfejezetben az alapvető fogalmakat és kifejezéseket határozzuk meg, amelyekhez a továbbiakban a kutatás során hivatkozni fogunk, és amelyeknek a tisztázása segíti a kutatás megértését. Az alfejezet célkitűzése annak tisztázása, hogy a kulcsfogalmak a doktori értekezésben milyen értelmezésben jelennek meg, részletesebb szakirodalmi áttekintésükre az elméleti keretet bemutató részben kerül sor.

### **4.1 Formulaszerű nyelvhasználat, formulaszerű szekvenciák, formulaszerű egységek**

Az értekezés egyik kulcsfogalma a formulaszerű nyelvhasználat, amely iránt az érdeklődés az elmúlt évtizedekben jelentősen megnőtt az alkalmazott nyelvészeti kutatásokban, és amelynek meghatározására számos definíció született (Nelson, 2018: 433). Nagyszámú kutatás bizonyította, hogy a spontán beszédfeldolgozással járó kognitív terhelés miatt a beszédprodukciók „korlátozott és ismétlődő repertoárral” (Biber et al., 1999: 1049) rendelkeznek, amely főként formulaszerű, előre megszerkesztett lexikai egységekből állnak (Wray, 2002, 2019). Továbbá, megállapították azt is, hogy az anyanyelvi beszélők diskurzusában nagy arányban szerepelnek a formulaszerű lexikai elemek (lásd Altenberg, 1998; Biber et al., 1999; Pawley – Syder, 1998). A jelenség leírására rengeteg kifejezés található a szakirodalomban, ilyenek például a legtöbbit használt lexikai egységek [lexical chunks], a lexikai kötegek [lexical bundles] és a többszavas kifejezések [multi-word expressions] is. Maga a formulaszerű nyelvhasználat kifejezés számos esetben átfogó gyűjtőfogalomként [umbrella term] funkcionál, amely magában foglalja az összes olyan jelenséget, amely olyan szavak csoportjaira utal, amelyek gyakran fordulnak elő együtt, egymás társaságában (vö. Firth, 1957; Sinclair, 1991; Hoey, 2005). A formulaszerű nyelvhasználat során használt formulaszerű szekvenciák, vagy formulaszerű egységek



lehetnek többszavasak (három vagy négy szóból állóak), vagy ennél rövidebb sorozatok, például a kétszavas kollokációk (Nelson, 2018: 433). Wray szerint a formulaszerű szekvencia kifejezés minden olyan többszavas láncra utal, amelyet a nyelvhasználó úgy érzékel, hogy egyetlen lexikai egységként azonosítható vagy hasznosítható (2019: 267). Szerinte ezek a szekvenciák holisztikusan tárolódnak és hívódnak elő a memóriánkból, csökkentve ezzel a kognitív terhelést (2002: 9). Erre azonban még nem találtak bizonyítékot (Siyanova-Chanturia – Pellicer-Sánchez, 2019: 5). Siyanova-Chanturia és Pellicer-Sánchez, (2019: 2) tanácsára azonban az értekezésben a formulaszerű egységek közül nem zárjuk ki az egyszavas formulaszerű egységeket sem (például *ugye*).

A jelen doktori értekezésben a formulaszerűséget elsősorban a korpuszban való nagy gyakoriság alapján határozzuk meg (vö. Hoey, 2005; Sinclair, 2004; Stubbs, 1996). Azonban a formulaszerű egységek kiválasztására vonatkozó kritériumok nemcsak a gyakoriságukra, hanem azok tulajdonságaira (átláthatóság, gyakoriság, gyors feldolgozás, jelentéstartalom, kiszámíthatóság, konvencionizáltság, megjósolhatóság, pragmatikai funkció, rögzítettség, sematikusság); vö. Carrol – Conklin, 2020: 96.) is épülnek, valamint figyelembe veszik az agglutináló nyelvek jellemzőit is. A formulaszerű nyelvhasználatra bővebben a II. 4. fejezetben térünk ki.

## **4.2 Természetes nyelvhasználat, anyanyelvi beszélő, kompetens nyelvhasználó**

A természetes nyelvhasználat a jelen doktori értekezésben arra a nyelvhasználatra utal, amelyet az anyanyelvi beszélők a mindennapi kommunikációjuk során, hétköznapi szituációkban, tudatos erőfeszítés nélkül alkalmaznak. McCarthy (2009: 118–146) meghatározása szerint a természetes nyelvhasználat számos szinten működik, beleértve a megnyilatkozások, a szövegek és a beszélők közötti kapcsolatokat is. Magában foglalja azokat a nyelvi eszközöket, amelyeket az anyanyelvi beszélők spontán használnak a kommunikáció során, valamint a nyelvhasználati szokásokat, a beszélők közötti interakciós mintákat és a szociokulturális kontextusokat, amelyek befolyásolják a nyelvhasználatot (vö. Timmis, 2022). Ahogy az előző alfejezetben láthattuk, az anyanyelvi beszélők nyelvhasználatára kerülnek a formulaszerű egységek (vö. Erman – Warren, 2000; Altenberg, 1998) is. Ezek a formulaszerű egységek elősegítik a beszéd folyékonyságát, és lehetővé teszik a beszélők számára, hogy könnyen átadjanak

információt, kifejezzék véleményüket vagy reagáljanak a másik fél beszédére (McCarthy, 2009: 122), ezért ezek a formulaszerű egységeket jelentősen hozzájárulnak a természetes nyelvhasználatához.

A fentiekhez kapcsolódóan, a természetes nyelvhasználat mellett az értekezésben többször is előforduló anyanyelvi beszélő és kompetens nyelvhasználó kifejezéseket a kutatás során úgy határozzuk meg, hogy olyan személy, aki magas szintű jártassággal rendelkezik egy nyelvben, képes összetett megnyilatkozásokat megérteni és előállítani, társadalmi interakciókban részt venni, és hatékonyan ki tudja fejezni a gondolatait és a véleményét. A kompetens nyelvhasználó fogalmi értelmezéseinek más aspektusaira a jelen kutatás során nem térünk ki (vö. például Busch, 2015).

### **4.3 Nyelvhasználati mintázatok**

A jelen doktori értekezésben megjelenő nyelvhasználati mintázatok és használati mintázatok kifejezések egy fogalomra utalnak, amelynek meghatározása Sinclair (1991) elméletén alapul, amely szerint a nyelvi tudás nagymértékben épül a nyelvi mintázatok [idiom patterns] ismeretére. A nyelvi mintázatok olyan frazeológiai elemek, amelyek a lexika és a grammatikai szerkezetek kombinációjából származnak. Sinclair (2004: 164) szerint a szavak kombinációi együtt alkotnak jelentést, nem pedig a különböző szójelentések adódnak össze egy-egy szerkezetben. Ennek a jelenségnek a leírására alkották meg a lexikogrammatika [lexico-grammar] fogalmát (Halliday, 1985). Ezek a mintázatok lehetővé teszik, hogy megértsük, hogyan kapcsolódnak a szavak egymáshoz, és hogyan hoznak létre értelmes egységeket (Ellis et al., 2016: 31). Hunston és Francis (2000) korpuszalapú elemzéseikkel, amelyek megerősítették Sinclair elméletét, két fontos eredményre jutottak. Az egyik, hogy a többjelentésű szavak különböző jelentéseikhez különböző mintázatok társulnak. A másik, hogy "bizonyos mintázatok hajlamosak olyan lexikai elemekhez kapcsolódnak, amelyeknek meghatározott jelentése van" (2000: 83). Hoey (2005) lexikai előfeszítés elmélete is ezen a gondolaton alapul, valamint ezt kiegészítve azon a megfigyelésen, amely szerint bizonyos szavak, illetve azok formái vagy vonzzák, vagy megpróbálják elkerülni egymást. Kutatásában Hoey három egymással összefüggő jelenséget figyelt meg: az általa használt korpuszban lévő kollokációkat, amelyek olyan szavak, amelyek gyakran megjelennek egymás társaságában; a kolligációkat, vagyis milyen grammatikai szerkezetekben jelennek meg a szavak; valamint a szemantikai

asszociációkat, amely kifejezéssel Hoey a korábbi szemantikai prozódia és a szemantikai preferencia fogalmát vonta össze (2005: 22–24). A kollokáció és a kolligáció kölcsönösen függenek egymástól, és mindkettő hozzájárul a szöveg jelentéséhez. Szemantikai asszociációról akkor beszélhetünk, ha „egy szó vagy egy szósorozat a nyelvhasználó elméjében egy jelentéshalmazhoz, vagy osztályhoz kapcsolódik” (Hoey, 2005: 24). A jelen doktori értekezésben a használati mintázatok megfigyelése alatt tehát ezeknek a jelenségeknek a megfigyelésére kerül sor.

#### **4.4 A pedagógiai korpusz**

A KorSzak Gyermeknyelvi Korpusz egy pedagógiai célokat szolgáló korpusz, amelynek célja, hogy a gyermekek számára készült tananyagokhoz és az annak elkészülését megelőző korpuszalapú kutatásokhoz autentikus szövegeket biztosítson. A jelen doktori értekezésben a pedagógiai korpusz fogalmát nyelvoktatásra készült autentikus szövegek halmazaként értelmezzük. A korpuszban szereplő szövegek összeállítása tudatosan történt a tananyagkészítést mint elsődleges célt szem előtt tartva. A korpuszépítés tervezésekor kizárólag a témák meghatározására került sor, amely nem befolyásolta a beszélgetések során a gyermekek által használt nyelvi eszközöket, így a természetes nyelvhasználat jól megfigyelhető a korpuszban. A témák kiválasztása a gyermekek bevonásával történt, amelyről részletesen olvashatunk a III. 1 fejezetben.

## 5. A doktori értekezés szerkezete

A jelen doktori értekezés négy nagyobb egységből tevődik össze: az elméleti keret bemutatásából, a kutatás, a kutatási eredmények áttekintéséből, valamint a kutatási eredmények felhasználásának, vagyis a gyermekek számára készülő tananyag ismertetéséből.

A II. fejezetben kerül sor a jelen kutatás alapjául szolgáló elméleti kutatási keret bemutatására. A fejezet elején a kutatással kapcsolatos, releváns szakirodalmi ismertetések fókuszában a nyelvvelajátítás áll. A bevezető utáni első alfejezetben a nyelvvelajátítási elméletek közül a használatalapú nyelvvelajátítási modellek (Tomasello, 2003, 2015; Ghalebi – Sadighi, 2015) kerülnek kiemelésre. A második alfejezetben pedig egy használatalapú nyelvvelajátítási modellre, a lexikai jóslás elméletére (Hoey, 2003, 2004, 2005; Pace-Sigge, 2013) térünk ki részletesebben. A fejezet második részében azt a kérdést próbáljuk megválaszolni, hogy mitől válunk kompetens nyelvhasználókká. Ebben a részben a formulaszerű nyelvhasználathoz (Wray, 2002; Wray – Fitzpatrick, 2008; Siyanova-Chanturia et al., 2019; Pellicer-Sánchez et al., 2022) kapcsolódó szakirodalom, valamint a beszéd-folyékony-sággal (Segalowitz, 2010; Tavakoli, 2020) kapcsolatos kutatások kerülnek bemutatásra. A fejezet végén a korpusznyelvészettel, a korpuszok típusaival kapcsolatos szakirodalom áttekintésére kerül sor, amely előrevetíti a harmadik fejezetben bemutatásra kerülő témákat. A második fejezetben bemutatott öt alfejezet – használatalapú nyelvvelajátítási modellek, lexikai előfeszítés elmélete, formulaszerű nyelvvelajátítás, beszéd-folyékony-ság-kutatás, valamint a korpusznyelvészeti kutatások – mindegyike azzal zárul, hogy jelennek meg ezek az elméletek és kutatási irányzatok a nyelvoktatásban. A fejezet a korpuszinformált tananyagok jelentőségének bemutatásával zárul.

A III. fejezetben a jelen kutatáshoz szintén szorosan kapcsolódó, de inkább kutatómódszertani témakörök találhatók. A fejezet eleje a KorSzak Gyermekegyesületi Korpusz építésének céljának és folyamatának, a korpusz felépítésének, általános jellemzőinek áttekintését szolgálja, kiemelve, hogyan zajlottak a felvételek, milyen módon történt a beszélt nyelvi szövegek átírása, valamint annotálása. A második alfejezetben a kutatás módszertanának ismertetésére kerül sor, amely megábrázolja a Sketch Engine, korpusznyelvészeti eszközöket tartalmazó felületnek, valamint a Praat hangelemző szoftvernek a jelen kutatáshoz releváns funkcióinak ismertetését. A harmadik fejezet második és harmadik alfejezete a formulaszerű egységek beazonosításával foglalkozik,

bemutatva a fontosabb magyar nyelvvel kapcsolatos kutatások módszertanát (ld. Dóla, 2014, 2016, 2018), valamint néhány más agglutináló nyelvhez, a finn (ld. Lehtonen – Laine, 2003; Lehtonen et al., 2007; Soveri et al., 2007; Vartiainen et al., 2009) és török nyelvhez (ld. Durrant, 2013; Aksan – Aksan, 2014; Mersini, 2015; Mersini – Aksan, 2016) kötődő vizsgálatokét. Ezután kerül ismertetésre az a feltételrendszer, amely alapján a jelen kutatás során kiválasztásra kerültek a vizsgált lexikai egységek. Végül a harmadik fejezet végén rátérünk a kutatásetikai kérdésekre, amelyekben a korpuszalapú kutatásoknál felmerülő általános kutatásetikai kérdések mellett arra is kiemelten foglalkozunk, hogy a gyermek adatközlőkkel való vizsgálatok mire kell körültekintően odafigyelni.

A IV. fejezet a jelen doktori értekezésben bemutatott korpuszalapú kutatás eredményeinek ismertetése, amely a fent bemutatott hat kutatási kérdés alapján rendeződött. Az első alfejezetekben a leggyakoribb egyelemű lexémák, a folytatásban pedig a leggyakoribb két- vagy többelemből álló lexikai egységek használati mintázatainak, funkcióinak bemutatására kerül sor.

Végül az V. fejezetben arra térünk ki, hogy a kutatás során kapott eredményeket hogyan lehet használni egy gyermekeknek szóló tananyag elkészítéséhez. Vagyis hogyan lesz a korpuszból tananyag. Hogyan épül fel a készülőben lévő korpuszinformált tankönyv, valamint, hogy milyen módszertant választottunk a felépítéséhez (modellalapú – Szita–Pelcz, 2017; projektalapú – Capraro, 2013; Pieratt, 2019).

## **5.1 Az értekezésben idézett fordításokkal, rövidítésekkel és hivatkozásokkal kapcsolatos információk**

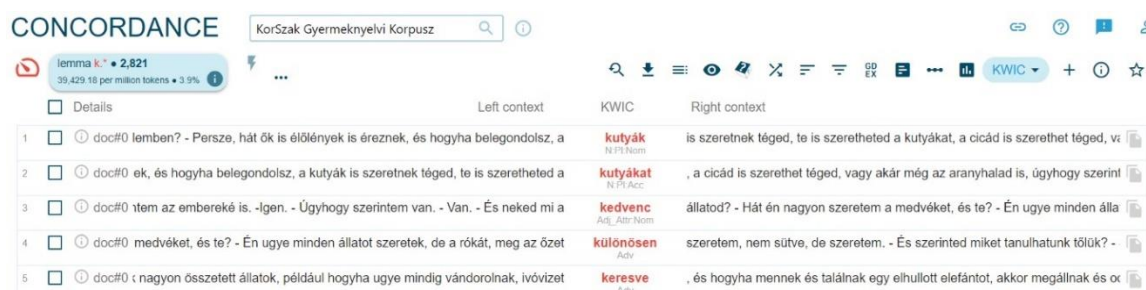
A jelen doktori értekezés fent bemutatott szerkezete mellett lényeges röviden ismertetni néhány olyan információt is, amelyek ismerete megkönnyítheti az olvasást. Az értekezésben használt hivatkozási stílus az APA 7. kiadásának (2020)<sup>2</sup> megfelelő, amely az alkalmazott nyelvészet területén elterjedt szabványos hivatkozási rendszer. Ennek alapján a felhasznált szakirodalomra való hivatkozások a szövegben történnek, a lábjegyzetben kizárólag a kiegészítő információk jelennek meg. Ebben az értekezésben néhány fordítás eredeti szövege, egy-egy fogalom magyar fordításának forrása, valamint a hiperhivatkozások találhatóak benne.

---

<sup>2</sup> <https://apastyle.apa.org/>

A szövegben található, eredetileg angol nyelvű idézetek saját fordításban jelennek meg. A fogalmakkal kapcsolatosan pedig azt kell megjegyeznünk, hogy amennyiben az adott fogalomnak létezett már korábban magyar nyelvű fordítása, az került használatra, ellenkező esetben itt is saját fordításról beszélhetünk. Az angol nyelvű eredeti fogalom szögletes zárójelben kerül közlésre az adott fogalom magyar fordítása után. Például lexikai előfeszítés [lexical priming]. Ezenkívül szögletes zárójelben, aposztróffal jelennek meg az angol nyelvhasználati példák magyar fordításai is. Például *salt and pepper* [’só és bors’].

Ahogy fent a *salt and pepper* esetében láthattuk, az értekezés folyószövegében dőlt betűvel szedve jelennek meg a nyelvhasználati példák, valamint emellett még a könyvek címei is. A korpuszból származó párbeszédekben, valamint a konkordanciasorokban vastag piros betűvel szedve jelennek meg a vizsgált kulcsszavak és kulcskifejezések [node, keyword]. Például az alábbihoz hasonlóan:



A kutatási eredmények bemutatásánál azok a nyelvhasználati példák, amelyek LEMMA formájában szerepelnek, arra utalnak, hogy az adott lemmából több olyan lexikai egység is származik, amely a korpuszban nagy gyakorisági mutatóval rendelkezik. Például az ÉN IS esetében a következő kételemű lexikai egységekre utal a lemma:

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| <b>ÉN IS</b>   | <b>188</b>      |                            |
| én is          | 144             | 76,60                      |
| nekem is       | 37              | 19,68                      |
| engem is       | 4               | 2,13                       |
| nálam is       | 2               | 1,06                       |
| velem is       | 1               | 0,53                       |

Végül ennél a résznél azt is meg kell jegyeznünk, hogy a korpuszban található használati mintázatok ismertetésekor a magyar emMorph (Novák, 2016) alapú morfológiai címkékészletet használtuk. Az értekezésben megjelenő címkék a következők<sup>3</sup>:

| <b>Morfológiai kód</b> | <b>Leírás</b>                  |
|------------------------|--------------------------------|
| <b>1Pl</b>             | többes szám, első személy      |
| <b>1Sg</b>             | egyes szám, első személy       |
| <b>Acc</b>             | "-t" tárgyesetrag              |
| <b>Adj</b>             | melléknév                      |
| <b>Adv</b>             | határozószó                    |
| <b>Inf</b>             | főnévi igenév                  |
| <b>Ins</b>             | "-val" eszközhatározói esetrag |
| <b>N</b>               | főnév                          |
| <b>Poss</b>            | birtokos paradigma             |
| <b>Pro</b>             | névmás                         |
| <b>Def.</b>            | határozott ragozás             |
| <b>NDef.</b>           | határozatlan ragozás           |
| <b>V</b>               | ige                            |

<sup>3</sup> A teljes emMorph kódlista a következő helyen található meg: [https://e-magyar.hu/hu/textmodules/emmorph\\_codelist](https://e-magyar.hu/hu/textmodules/emmorph_codelist)

## II. Elméleti kutatási keret

### 1. Bevezetés az elméleti kutatási keret meghatározásához

Ahogy a bevezető részben (I. fejezet) már említettük a jelen doktori értekezés fő célja a korpuszalapú kutatás módszertanának felhasználásával a KorSzak Gyermeknyelvi Korpuszban szereplő magyar anyanyelvű, 11–15 éves gyermekek természetes nyelvhasználatának megfigyelése és vizsgálata, valamint ennek a kutatásnak az eredményeinek felhasználása egy korpuszinformált magyar mint idegen nyelvi tananyag elkészítése érdekében.

A kutatás Hoey lexikai előfeszítés (2003, 2004, 2005) elméletét alapul véve, a gyermekek spontánbeszédében megjelenő formulaszerű lexikai egységekre (vö. Wray, 2002, 2019) koncentrálok. Ezek a lexikai egységek kiemelkedő szereppel bírnak az első nyelv elsajátításában, valamint hasznosak lehetnek a nyelvtanulók számára is.

Ebből kifolyólag a kutatásnak öt elméleti alappillére:

1. a nyelvsajátítással kapcsolatos elméletek közül a használattalapú nyelvelméleti modellek,
2. a fent említett lexikai előfeszítés elmélete,
3. a formulaszerű nyelvhasználat mint a fluensséghoz hozzájáruló egyik legjelentősebb tényező,
4. a beszédhelyességhez kapcsolódó legfontosabb elméletek,
5. valamint kutatómódszertani szempontból a korpusznyelvészeti megközelítés.

Az értekezés második fejezetében a kutatás alappilléreinek számító fent felsorolt elméleteket és megközelítéseket fejtem ki részletesen szakirodalmi áttekintéssel. Minden alfejezetben kitérek az adott elmélet vagy megközelítés nyelvoktatásban betöltött jelentőségére is, valamint kiemeltem a kapcsolódó kutatások releváns eredményeit.



## 2. A használat alapú nyelvelméleti modellek

A használat alapú nyelvelméleti modellek [usage-based language theories; item-based language acquisition; exemplar-based theories] röviden összefoglalva arra a kérdésre keresik a választ, hogyan sajátítjuk el a nyelvet, miközben azt használjuk (Ellis et al., 2016). A válasz megtalálásához azokat a kommunikatív és kognitív folyamatokat vizsgálják, amelyek a nyelvet alakítják (Slobin, 1997: 267). A használat alapú nyelvelméleti modelleket több területen is alkalmazzák, ilyenek az „empirikus és elméleti megközelítések: kognitív nyelvészet, konstrukciós grammatika, funkcionális nyelvészet, kognitív pszichológia, tanuláselmélet, pszicholingvisztika, statisztikai tanuláselmélet, gyermekkori nyelvsajátítás, idegtudomány, korpusznyelvészet, számítógépes nyelvészet, természetes nyelvi feldolgozás, emergencia és komplex rendszerelmélet, társalgási elemzés, dinamikus rendszerelmélet, szociolingvisztika és társadalmi tanuláselmélet” (Ellis et al., 2016: 24).

A használat alapú nyelvelméleti modellek nagyban hozzájárulnak a nyelvsajátítás megértéséhez (Dörnyei, 2009: 207). Az angol nyelvű elnevezések közül a usage-based language acquisition a legelterjedtebb, amely magát a nyelv használatát, amíg az item-based és az exemplar-based theory (vö. Hopper, 1987) inkább az itemeket és a példákat helyezik az elmélet fókuszába. Az elméletek szerint a helyes értelmezés céljából, amikor találkozunk egy nyelvi példával, a nyelvi rendszerünk összehasonlítja ezt a példát egy ugyanolyan vagy hozzá hasonló példával való korábbi találkozások emlékeivel (Tribushinina – Gillis, 2017; Tribushinina – Evers-Vermeul, 2017).

. A használat alapú modellek két kulcsfontosságú aspektusra épülnek. Az egyik, hogy a nyelv és a nyelvsajátítás általános kognitív és társadalmi folyamatokon, valamint a jelentésen alapul. A másik pedig, hogy a kommunikáció kihívása, hogy minden egyes beszélgetőpartner elméjében megteremtse a figyelem konvergenciáját a hasonló fogalmakra, vagyis a nyelv alkalmazásának a célja, hogy a beszélgetőpartnerek, ahogy Langacker nevezi, mentális érintkezéseket [mental contacts] hozzanak létre (Langacker, 1987 idézi Ortega et al., 2016: 1). A használat alapú elméletek szoros összefüggésben állnak a kognitív nyelvészettel, valamint a konstrukciós grammatikával, amelyben a nyelv alapegysége a konstrukció, vagyis egy jelentéssel teli és értelmes, releváns kommunikációs céllal használható egység (Tribushinina – Evers-Vermeul, 2017). A szintaktikai sémákat, az idiómákat, a morfológiai elemeket, a szófajokat és a lexikai egységeket mind a

konstrukció részeként értelmezik (vö. Langacker, 1987, Fillmore et al., 1988; Goldberg, 1995, 2007). A nyelvi szervezettséget és a nyelvi produktivitást a forma és a jelentés analógiájának eredményeként értelmezik (Ibbotson, 2013: 2).

A kommunikáció használatalapú megközelítésében két alapelv jelenik meg:

1. a jelentés egyenlő a használattal: ez a nyelvi kommunikáció funkcionális vagy szemantikai dimenziójának megközelítése, vagyis, hogy az emberek hogyan használják a nyelvi konvenciókat társadalmi célok elérésére;
2. a struktúra a használatból ered: a nyelvi kommunikáció strukturális vagy grammatikai dimenziójának megközelítése, vagyis hogyan keletkeznek a nyelv használatából a jelentésen alapuló grammatikai szerkezetek (Ghalebi – Sadighi, 2015: 191; Tomasello, 2015).

A használatalapú nyelvelméleti megközelítések szerint a nyelvi fejlődés annak a rengeteg asszociációnak a következménye, amely a nyelv használata közben megy végbe. Ezek alapján a nyelv használatából adódó példák a memóriában tárolódnak és az elérésük az ismétlések által válik gyorsabbá és hatékonyabbá (Tomasello, 2003, 2015; Ellis, 2002). A nyelvhasználati példák memorizálásának sikeressége attól függ, hogy milyen gyakran találkozunk velük. Ezek a találkozások, illetve a komplex nyelvi környezet biztosítja azt az információt, amelyből a gyermek – és a nyelvtanuló – kivonja a törvényszerűségeket, mintázatokat (Tomasello, 2013). A használatalapú nyelvelméleti modellek elvetik azt az elméletet, miszerint a nyelvsajátítás egy belső, velünk született nyelvsajátítási eszköz (vö. Chomsky, 1957) által történne. Szerintük ez inkább egy több tényezőből álló tanulási folyamatnak köszönhető, amely makro- és mikroszinten megy végbe. Vagyis az évszázadok alatt végbemenő nyelvváltoztatások során ugyanúgy, ahogy az első nyelv elsajátítása folyamán is. „Az egyén kreatív nyelvi kompetenciája a teljes nyelvhasználati történetének összes jelentéssel teli interakciójára vonatkozó emlékek együttműködéséből alakul ki” (Ellis et al., 2016: 24). Bár van köztük némi különbség, ezek az elméletek mind a nyelvsajátítás, a nyelvtanulás és nyelvfeldolgozás egy-egy valószínűség alapú modelljére [probability-based models] támaszkodnak (vö. Barlow – Kemmer, 2000; Kuperberg – Jaeger, 2016), nem pedig egy veleszületett absztrakt nyelvi szabályrendszerre, mint például az univerzális grammatika elméletére. Fontos még azt is megjegyezni, hogy az elméletek szerint a tanulási folyamat csak maximális kitettség [maximum exposure] és

gyakori nyelvhasználati lehetőségek biztosítása esetén lehet megfelelően hatékony (Dörnyei, 2009: 207).

## **2.1 Gyermek anyanyelv-elsajátítása a használatalapú elméletek alapján**

A jelen doktori értekezés fókuszában a tizenévesek korcsoportja áll, ugyanakkor a kutatás szempontjából lényeges bemutatni a használatalapú elméletek nyelvelsajátításhoz való viszonyát is. A használatalapú elméletek szerint a gyermekek korai nyelvi kompetenciája itemalapú (Babarczy et al., 2014: 472), amely azt jelenti, hogy a megnyilatkozásaik konkrét szavak és kifejezések köré szerveződnek, nem pedig rendszerszintű szintaktikai kategóriák vagy sémák köré, ahogy azt korábban gondolták, és amely valójában csak később, fokozatosan alakul ki náluk (Tomasello, 2003: 5–6). A korpuszokon történő kutatás eredményei alapján látható, hogy a gyermekek nyelvi input hatására nem aktiválnak automatikusan egy nyelvi feldolgozó rendszert, amely a grammatika fő szabályaival és paramétereivel van ellátva. Ehelyett a nyelvelsajátítás a konkrét nyelvi itemek, vagyis szavak és kifejezések elsajátításával kezdődik, miközben funkcionális kommunikációt folytatnak a körülöttük lévő személyekkel (Dörnyei, 2009: 208–209).

Tomasello (2003, 2013) szerint minden nyelvi szerkezetet azonos mechanizmusokon keresztül sajátítanak el a gyermekek. Ezek a mechanizmusok két fő kategóriába sorolhatóak: a szándékolvasás [intention-reading] és a mintakeresés [pattern-finding] csoportjába.

A nyelvelsajátításban pedig három folyamat játszik központi szerepet:

1. Az imitáció, amely az utánzásnál többet jelent, inkább a felnőttek által használt nyelv értelmes és funkcionálisan megfelelő megismétlésével egyenlő.
2. A mintakeresés, valamint a talált minták kategorizációja és sematizációja.
3. Végül pedig a konstrukciók kreatív kombinálása (Dörnyei, 2009: 209–210).

A gyermekek nyelvelsajátításában résztvevő, központi szerepet játszó folyamatok az első két évben fejlődnek ki. A csecsemők először észlelik a figyelmet (tekintetkövetés), ezután megtanulják manipulálni azt (irányító mutató). Később megértik a szándékot,

mások céljait, végül pedig kifejlődik náluk a közös célú társas koordináció képessége, vagyis a közös tevékenységben való részvétel, a kommunikációs cél elérése érdekében történő jelentésekről való egyezkedés [negotiation for meaning] (Tomasello, 2008).

A gyermekek tehát a használatlapú nyelvelméletek szerint először rutinszerűen, itemként sajátítják el a kifejezéseket és a hozzájuk tartozó jelentést, majd csak később szedik elemeikre és értelmezik őket. A használatlapú megközelítések szerint (vö. Langacker, 1987; Bybee, 1985, 1995; Croft, 2001; Tomasello, 2003, 2009, 2015) az első nyelv elsajátításának egyik központi eleme a formulaszerű nyelvhasználat, amelyre a II. 4 fejezetben fogok részletesebben kitérni. Ugyanakkor itt is lényeges megjegyezni, hogy a korpusznyelvészeti kutatások elterjedése óta már nem kétséges, hogy a formulaszerű nyelvhasználat jelentős részét képezi a gyermeknyelvnek és a dajkanyelvnek, vagyis a gyermekek felé irányuló nyelvnek [child-directed speech]. Több kutatást végeztek ebben a témában, amelyekből kiderül, hogy a gyermekek felé irányuló nyelv tele van formulaszerű egységekkel (vö. Bannard – Matthews, 2008; Theakston – Lieven, 2017). Ezeket a formulaszerű egységeket a gyermekek hallás után memorizálják, majd később próbálgatják és használják is. Bannard és Matthews (2008) kísérletében nemcsak a gyermekek felé irányuló nyelvet elemezte, hanem még arra is kitért, hogy a gyermekek általi formulaszerű nyelvhasználatot részletesebben megvizsgálja. Hároméves gyermekeket kértek meg arra, hogy bizonyos általuk ismert és ahhoz hasonló, de nem megszokott szókapcsolatokat ismétljenek. A gyermekek természetesen sokkal eredményesebbek és gyorsabbak voltak azon elemek ismétlésében, amelyeket már a mindennapi életükben többször is hallottak. A fentebb ismertetett kutatások mellett a használatlapú elméletek képviselői gyakran kutatják a grammatizáció folyamatát (a fogalom grammatikalizáció, grammaticizáció formában is használatos (vö. Dér, 2019), amely során általános értelemben véve egy lexikai elem fokozatosan grammatikai státuszt nyer és amely sok szempontból a makroszintű megfelelője annak a mikroszintű folyamatnak, amely alatt a gyermekek elsajátítják a nyelv szintaktikai szerkezetét (Dörnyei, 2009: 210). Tomasello (2003: 5) szerint a kommunikáció során a szavak szekvenciákká fonódnak össze, használati mintákat képezve, amelyek később nyelvtani szerkezetekké szilárdulnak meg. Ellis és Larsen-Freeman (2006: 567) amellet érvel, hogy a grammatizációs folyamat a nyelvi elemek gyakran előforduló szekvenciáinak automatizálását jelenti, amely szintaktikai konstrukciók megjelenését, majd a gyakori találkozások hatása miatti funkcióváltásokat eredményezhet. A gyakoriság azonban nemcsak a nyelv változása során, hanem a mikroszinten, vagyis a nyelvelsajátítás során is fontos szerepet játszik. A használatlapú elméletek kutatói szerint a gyermekek

nyelvi feldolgozó rendszerének mintakereső funkciója, vagyis a megjegyzett konstrukciókból történő szabályszerűségek kikövetkeztetése erősen függ a minták gyakoriságától (Ellis, 2002: 144). Ellis gondolatát Bybee (2008) is megerősíti, aki arról ír, hogy a pszicholingvisztika, a kognitív és funkcionális nyelvészet, a számítógépes nyelvészet, a korpuszalapú elemzések, valamint a diskurzuselemzés, mind arra a felismerésre jutottak, hogy a nyelvi tudás szilárdan a nyelvi tapasztalatokon alapul, és a használat gyakorisága a nyelv grammatikai tulajdonságainak alapvető meghatározója. Larsen-Freeman (2002: 281) ugyanakkor felhívja a figyelmet arra, hogy csupán a gyakorisággal nem lehet a gyermekek nyelvelsajátítását megmagyarázni (vö. Steinkrauss, 2017), valami más is kell hozzá, amelyet ő valószínűségi tendenciáknak [probabilistic tendencies] nevez. Ez a gondolat Bob, Hay és Jannedy *Probabilistic Linguistics* című könyvében alapul, amelyben kifejtik, hogy „a nyelvtudást nem kategorikus szabályok vagy megkötések minimális halmazaként kell felfogni, hanem gradien szabályok halmazaként, amelyet statisztikai eloszlás jellemez” (2003 idézi Dörnyei, 2009: 219). Ez azt jelenti, hogy a gyakoriság mellett a nyelvi minták, szerkezetek kiszámíthatóságának is jelentős szerepe van, és a korábbi elméletekkel ellentétben a grammatikai szabályok a használat valószínűségével társulnak. Így nem az a szerkezet rögzül, amely nyelvileg lehetséges, hanem az, amelyről azt gondoljuk, hogy nyelvileg valószínű, amellyel már többször találkoztunk abban a kontextusban, a már ismert nyelvi elemek társaságában (vö. II.3 fejezet: A lexikai előfeszítés elmélete).

## **2.2 Nyelvi észlelés és nyelvhasználat a használatalapú nyelvelméleti keretben**

A jelen kutatás szempontjából fontos megemlítenünk néhány jelentősebb használatalapú nyelvelméleti keretben történt pszicholingvisztikai és korpusz nyelvészeti kutatást. A fent említett valószínűségeknek ugyanis nemcsak a nyelvelsajátításban van meghatározó szerepük, hanem áthatják a teljes nyelvi rendszert (Dörnyei, 2009: 220). Így a nyelvi észlelést [language perception] is, amelynek során az észlelt nyelvi forma minden darabja lehet akár két- vagy többértelmű. A nyelvi észlelés összetett folyamatában az előzetes észlelés kiegészül az asszociációk révén kikövetkeztetett elképzeléseinkkel. Ahol az észlelés többértelmű is lehet, nem tudatos folyamatok által mérlegeljük a lehetőségeket, és a legvalószínűbbet választjuk ki közülük. A pszicholingvisztikai kutatások eredményei azt

mutatják, hogy a kompetens nyelvhasználók az ilyen kétértelműségeket a korábbi tapasztalataikból származó valószínűségi számítások gyors elvégzésével oldják meg (vö. MacDonald, 1994).

A magyar nyelvre vonatkozóan Gergely és Pléh (1994, 1995) végzett el kutatásokat a kétértelműséggel kapcsolatban. Azt vizsgálták, hogy a *dob* szó főnévként vagy igeiként aktiválódik inkább. Ehhez a szó toldalék nélküli (*dob*), kétértelmű toldalékot tartalmazó (*dobom*, amely lehet főnév és ige is), valamint egy egyértelmű toldalékkal rendelkező alakját (*dobsz*) használták. A vizsgálataikból kiderült, hogy attól függetlenül, hogy a *hegedű* szó irreleváns volt a kontextusban, a *dob*, *dobom* és még a *dobsz* alak után is gyorsabban ismerték fel a résztvevők, mint a releváns *hajít* igt. Tehát a vizsgálat során a *dob* szó főnévi jelentése is aktiválódott a kutatásban résztvevők mentális lexikonjában (Thuma, 2014: 219–222).

Egy másik magyar vizsgálat Thuma és Pléh (1995) modalitásközi priming kísérlete, amelyet harminc angol szakos magyar anyanyelvű hallgatóval végeztek el, és amelyben arra keresték a választ, hogyan működik a két nyelv a mentális lexikonban. A használt magyar mondatok utolsó szava vagy magyar vagy pedig angol nyelven volt kétértelmű (például *mosogató – sink* [ˈmosogató; sülyed]). Szignifikáns eredményeket azonban csak abban az esetben kaptak, amikor a mondatok utolsó szava főnév volt (Thuma – Pléh, 1995). A két bemutatott magyar kutatásban megfigyelhető jelenség a lexikai előfeszítéssel [lexical priming] magyarázható, amelyről a II. 3 fejezetben lesz részletesebben szó.

A kétértelműséggel kapcsolatos memóriavizsgálatok során (Anderson, 1989, 1990; Anderson – Milson, 1989; Schooler – Anderson, 1997) arra a megállapításra jutottak, hogy a kompetens nyelvhasználónak egy nyelvi elem kétértelműségekor a döntés meghozatalához három információra van szüksége: az adott nyelvi elem gyakoriságára, hogy milyen régen használták azt [recency], valamint a kontextusra. A gyakoriság jelentőségéről már többször szó esett, ezért térjünk rá a második információra, vagyis, hogy az adott nyelvi elem mennyire számít régen használatnak. Schooler és Anderson (1997) korpusznyelvészeti kutatásukban kimutatták, hogy létezik egy hatványfüggvény (log-log lineáris), amely a New York Times címlapján egy adott napon egy szó előfordulásának valószínűségét azzal hozza összefüggésbe, hogy mennyi idő telt el azóta, hogy a szó korábban előfordult az adott kontextusban. A nyelvi struktúrák a gyakoriságuk szerint rendezettek, de ezt a rendezettséget erősen befolyásolja, hogy mennyire használt vagy divatos az éppen kiválasztott nyelvi elem (Ellis et al., 2016: 38) (ld. II.3.1.2. Ismétléses előfeszítés).

A kétértelműség megoldásához szükséges harmadik információval, a kontextussal kapcsolatban Schooler (1993 idézi Ellis et al., 2016) megállapította, hogy egy adott szó nagyobb valószínűséggel fordul elő, ha más olyan szavak is jelen vannak körülötte, amelyek a múltban azonos kontextusban együtt fordultak elő vele. A kontextusnak egy tágabb értelmezését vette figyelembe, amely szerint kontextusnak nemcsak egy témakör vagy újságcikk nevezhető, hanem akár egy nyelvhasználót is annak tekinthetünk. „Bizonyos emberek, bizonyos dolgokról, bizonyos módon beszélnek” fogalmaz Ellis et al. (2016: 39). Ezzel kapcsolatban példaként megemlíti a Patrick Joula (2006) által írt számítógépes elemzőprogramot, amely egy a nevéből következtetve férfi szerző tollából származó krimiről (Robert Galbraith: *The Cuckoo's Calling*) a szöveg stílusának és a benne található szerkezeteknek az elemzéséből állapította meg, hogy valójában a Harry Potter sorozat írójének, J. K. Rowlingnak az álnéven született írása.

### **2.3 A használatalapú elméletek a nyelvoktatásban**

A fenti összefoglalóból láthattuk, hogy a használatalapú nyelvelméleti megközelítések szerint a nyelvi fejlődés a nyelv használata közben végbemenő nagymennyiségű asszociáció következménye. Ez nemcsak az anyanyelv elsajátítására vonatkozik, hanem a nyelvtanulásra is (Ellis – Wulff, 2019). Tehát ahhoz, hogy valóban megtanuljunk egy nyelvet, használnunk kell azt.

A nyelvoktatás szempontjából arra a megfigyelésre alapoznak, hogy még a nem egy helyen és időben lévő anyanyelvi beszélők is hasonló tartalmak kifejezésére, hasonló nyelvi elemeket használnak (O’Keeffe, 2007). A nyelvtanulás fontos részévé válik a megfigyelés, „a nagy mennyiségű autentikus nyelvi adat elemzése lesz a tanulás elsődleges forrása” (Pelcz, 2023: 15). Ezekből kiindulva három fogalmat kell még a megfigyelés és a használat mellett kiemelnünk. A hatékony nyelvtanulás első kulcsfogalma az ismétlés, mivel az ismétlés által a nyelvtanulók képesek lesznek gyorsabban és hatékonyabban tárolni és elérni a mentális lexikonjukban lévő nyelvhasználati példákat (Tomasello, 2003, 2015; Ellis, 2002). A második fogalom a maximális kitettség [maximum exposure], amely nyelvtanári szempontból arra utal, hogy a nyelvórákon megfelelő mennyiségű és minőségű célnyelvi inputot kell biztosítanunk a tanulók számára. Tanulóként pedig minél több alkalmat kell megragadnunk célnyelvi input szerzésére (vö. Mitchell et al., 2019: 128–167; Pérez-Parades et al., 2020). Az autentikus nyelvi anyag megfigyelésének fontossága mellett

ugyanis ez a nyelvelmélet rámutat arra is, hogy a tanulási folyamat során, azaz a nyelvórán minden találkozás egy-egy nyelvi elem beépítését szolgálja (Pelcz, 2023: 15). Végül a harmadik kulcsfogalom a gyakori nyelvhasználati lehetőségek biztosítása a tanulók számára, illetve nyelvtanulói nézőpontból ezeknek az alkalmaknak a megfelelő kihasználása (Dörnyei, 2009: 207). Ellis et al. (2016) szerint a nyelvtanulás során a tanulóknak a konstrukciók gyakoriságának megállapításából, a gyakoriságnak és kontextusának az összevetéséből, valamint a kommunikáció optimalizációjából álló folyamatot kell megtanulniuk, amelyet csak a gyakori nyelvhasználat tesz lehetővé számukra (38–39).



### 3. A lexikai előfeszítés elmélete

Michael Hoey lexikai előfeszítés [lexical priming] (2003, 2004, 2005) elméletével egy használatalapú nyelvelméleti modellt hozott létre, amellyel a célja az volt, hogy „összekösse a korpusznyelvészeti (lexikai aspektus) és a pszicholingvisztikai (priming aspektus) kutatásokat” (Pace-Sigge, 2013a: 149). Ebben a fejezetben az elmélet pszicholingvisztikai és korpusznyelvészeti előzményeinek rövid bemutatása után, magát a lexikai előfeszítés elméletet és annak eddigi kutatásait fogom részletesebben bemutatni.

#### 3.1 Pszicholingvisztikai előzmények

Ahogy a bevezető részben említésre került a priming alapvetően egy pszicholingvisztikai fogalom, amelyet Hoey az elmélete alapjaként használt fel és korpusznyelvészeti eszközökkel bizonyított. A priming fogalmának alapfeltevése, hogy az emberi agy nem véletlenszerűen fér hozzá a memóriához, mivel a kívánt információhoz hozzáférni sokkal könnyebb, ha az már egy másik információhoz kapcsolódik, illetve ez a kapcsolat annál erősebb, minél gyakrabban találkozunk az adott információval vagy egy ahhoz hasonlóval (Pace-Sigge, 2013a: 153). „Vagyis egy szó felismerése facilitálja a vele kapcsolatban lévő más szavak felismerését, egy modalitáson belül vagy modalitások között” (Pléh, 2014: 153).

A primingot, vagyis magyarul az előfeszítést (vö. Thuma – Pléh, 1995) a pszicholingvisztikai kutatásokban többféleképpen mutatják ki. Az ilyen kutatásokban használt tesztekben leggyakrabban a lexikai döntés [lexical decision task], a szóazonosítás [word identification task] és a szótőkiegészítés [stem completion task] feladat található (Vermekei, 2023a). A lexikai döntés típusú feladatokban a kísérletben résztvevő személyeknek gyorsan kell döntést hozniuk arról, hogy a bemutatott betűsor szónak tekinthető-e vagy sem (*kutya – gutya*). Ennél a feladatnál az előfeszítést az mutatja meg, ha az elsőhöz képest a betűsor második bemutatásakor a lexikális döntés meghozatalának látenciaideje csökken. A szóazonosítási feladattípus a szó felismeréséhez szükséges minimális kitettséget méri. Az előfeszítést ennél a vizsgálatnál az ismétlődő szavak alacsonyabb minimális kitettsége jelzi. A szótőkiegészítéssel feladattípusokban pedig a kísérletben résztvevő személyek kapnak egy hiányos szót (*bög\_\_*), amelyet az első eszükbe

jutó módon kell kiegészíteniük (például *bögre*). A szótókiegészítési feladatok során az előfeszítettséget a megoldások tendenciájából állapítják meg (Ellis et al., 2016: 39).

A primingnek több fajtáját is meghatározták. Létezik szemantikai, ismétléses, formaalapú, jelentésalapú, morfológiai és a maszkolt priming is (vö. Thuma, 2014: 164–194). A következő részben a két típushoz a szemantikai előfeszítéshez [semantic priming] és az ismétléses előfeszítéshez [repetition priming] kapcsolódó néhány fontosabb kutatás eredményeit tekintem át röviden, amelyek amellet, hogy gyakran vizsgált területek Hoey lexikai előfeszítés elméletét is megerősítik.

### 3.1.1 Szematikai előfeszítés

A szemantikai előfeszítés-vizsgálatokban azt elemzik, hogy a kísérletben résztvevő személyek milyen gyorsan ismernek fel egy adott szót miután kapnak egy hívóképet vagy egy hívószót, amit angolul prime-nak nevezünk. Meyer és Schvaneveldt (1971) kimutatta, hogy vannak olyan hívószók, amelyek lassítják a felismerés folyamatát, amíg mások felgyorsítják azt. Neely (1976, 1977), aki Meyer és Schvaneveldt kutatásaira építette saját vizsgálatait, *Semantic Priming of Lexical Memory* című munkájában írja le elsőként a priming és a lexika szavakat együtt (Pace-Sigge, 2013a: 160). Az 1976-os kísérletében résztvevők három különböző időtartamig (360 ms, 600 ms és 2000 ms) a célszó előtt szemantikailag kapcsolódó, nem kapcsolódó, valamint semleges szavakat láttak hívószóként. Az eredmények azt mutatták, hogy a kitettség időtartalmától függetlenül minden esetben a szemantikailag kapcsolódó szavak váltottak ki rövidebb reakcióidőt. Az „alany a priming szóval irányítja figyelmét azokra a szavakra, amelyek szemantikailag kapcsolódnak hozzá” írja könyvében (Neely, 1976: 652, idézi Pace-Sigge, 2013a: 160). Kísérletének eredményeiből három következtetést vont le, amelyekkel megerősítette Meyer és Schvaneveldt eredményeit is. A három következtetése szerint a hívószó:

1. megkönnyíti a lexikális döntéseket egy később bemutatott, szemantikailag kapcsolódó szóval kapcsolatban,
2. gátolja a lexikális döntéseket egy később bemutatott, szemantikailag nem kapcsolódó szóval kapcsolatban,
3. megkönnyíti a lexikai döntéseket egy később bemutatott álszóval [nonword] kapcsolatban (Neely, 1976: 654).

A következő mérföldköve a priming kutatásának Ratcliff és McKoon (1988) vizsgálata. Az úgynevezett összetett támpont felidézés<sup>4</sup> [compound cue retrieval] elméletükkel, amely szerint a priming azért következik be, mert a hívószó és a célszó egy összetett egységet alkot a rövidtávú memóriában, majd a rövidtávú memória interakcióba lép a hosszútávú memóriával, amely meghatározza a hívószó és a célszó familiaritását. Minél szorosabb az asszociációs kapcsolat a két szó között, annál erősebb ez a familiaritás és annál gyorsabban megy végbe a célszó felismerése (Kovács, 2000: 5). Egy későbbi kutatás során Trofimovich (2005) is arra a következtetésre jutott, hogy a priming jelensége kapcsolatban áll a hosszútávú memóriával. Első és második nyelvi kontextusban a beszélt nyelvben vizsgálta meg a szemantikai előfeszítést és észrevette, hogy a hallás utáni priming hatása hosszútávú, tehát ez a jelenség a hosszútávú memória részeként értelmezhető (Pace-Sigge, 2013a: 163). Visszatérve Ratcliff és McKoon (1988) kutatására egy korábbi kísérletükre utalva még azt is megjegyzik tanulmányukban, hogy a hívószó és a célszó négy szónál is távolabb lehetnek egymástól. Ez ellentétes Sinclair (1991) állításával, amelyről a II.3.2 fejezetben lesz bővebben szó, és amely szerint ez a távolság maximum négy szó lehet. Ugyanakkor meg kell jegyeznünk azt is, hogy mindkét kijelentés erősen függött az akkori technikai eszközöknek a kapacitásától. Ma már a modern korpuszeszközök segítségével az elemzett szó vagy kifejezés több ezer szomszédját is meg tudjuk vizsgálni.

Pszicholingvisztikai keretben ezenkívül sokan vizsgálták a jelentés és az asszociáció viszonyát. A szavak mentális lexikonban való asszociációjával kapcsolatban Shelton és Martin (1992) arra az eredményre jutott, hogy akkor gyorsabb a szófelismerés folyamata, amikor a hívószó és a célszó a kísérletben résztvevő alany mentális lexikonában asszociálódnak. McRae és Boisvert (1998) pedig az mutatta ki, hogy ha a szavak szorosan összefüggő jelentéssel bírnak, akkor asszociáció nélkül is gyorsan megtörténik a szemantikai előfeszítés. Ezt kiegészítve de Morney Davies (1998) kijelentette, hogy „még, ha két szó a szó szoros értelmében nem „rokon szemantikailag” (azaz nem ugyanabból a felettes kategóriából származnak), gyakori asszociációjuk a jelentés szintjén egyfajta kapcsolatot eredményez” (de Morney Davies, 1998: 394). Ezenkívül de Morney Davies Hoey lexikai előfeszítés elméletének két alapvető részét is előrevetíti. Egyrészt, hogy az egyes kifejezések a gyakori asszociáció eredményeképpen úgy vannak előfeszítve, hogy jelentsenek valamit. Másrészt pedig, hogy a szavak tulajdonságai közé tartozik az is, hogy mely szavakat vonzzák és melyeket taszítják, amelynek magyarázata, hogy az elménk

---

<sup>4</sup> A fogalom fordítását Kovács Ágnes tanulmányából (2000: 5) vettem át.

egymáshoz társítja őket (Pace-Sigge, 2013a: 162). Végül meg kell még említenünk Hernandez et al. (2001) kutatását, amelyben a lexikai priming és a mondat priming kétlépcsős modelljét (vö. Bedecker et al., 1995) próbálták megfigyelni lexikai döntési és keresztmodális megnevezési feladatok során. A kísérlet során nem találtak bizonyítékot arra, hogy a folyamat kétlépcsős lenne, így bebizonyították, hogy a priming egyetlenegy automatikus folyamat, nem bonthatjuk részekre.

### 3.1.2 Ismétléses előfeszítés

A második gyakran kutatott primingtípus – amely Hoey lexikai előfeszítés elméletét is megalapozta – az ismétléses előfeszítés. Erre a jelenségre vezethető vissza, hogy hajlamosak vagyunk újra felhasználni azokat a nyelvi szerkezeteket, amelyeket nemrég hallottunk vagy olvastunk (vö. II.2.2 fejezet). Ez a jelenség megfigyelhető a fonológia, fogalmi reprezentációk, lexikai választás és szintaxis területén is (vö. McDonough – Trofimovich, 2008). Bock 1986-os kutatásában nyelvi szerkezetek újrahasznosítását vizsgálta. A kísérletben résztvevő személyek egy mondatot olvastak el négyféle struktúrában: tranzitív – aktív (*George kicked the ball.*), tranzitív – passzív (*The ball was kicked by George.*), datívuszi szerkezet kettős [indirect, direct] tárggyal (*George gave the boy the ball.*) és datívuszi szerkezet prepozíciós változatban (*George gave the ball to the boy.*). A mondat elolvasása és elisméltése után, a résztvevőknek egy a témájában a mondathoz nem kapcsolódó képet kellett leírniuk. A kísérletben lévő feladatszekvencia utolsó eleménél, a képleírásnál az esetek nagy többségében azt a szerkezetet használták a résztvevők, amelyet közvetlenül előtte olvastak és ismételtek (Ellis et al., 2016: 40). Gries és Wulff (2005), akik német első nyelvű, angolul haladó szinten beszélő személyekkel végeztek hasonló kísérleteket vizsgálatukkal bebizonyították, hogy ez a jelenség nyelvtanulóknál is kimutatható. Branigan et al. (2007) pedig ezt kiegészítve azt találta, hogy az ismétléses előfeszítés nemcsak a monológokban, hanem a dialógusokban is megjelenik, „ami azt jelzi, hogy a megértés és a produkció osztozik a nyelvtani reprezentációkon, valamint, hogy a dialóguspartnerek összehangolják mind a nézőpontjaikat, mind a nyelvhasználatukat” (Ellis et al., 2016: 41). Garrod és Pickering (2009) szerint a párbeszéd különböző szinteken történő közös cselekvés. A legmagasabb szinten a beszélgetőpartnerek célja a mentális reprezentációik összehangolása. Ez az alacsonyabb szinteken végzett közös tevékenységen alapul: ilyen a már említett nyelvi szerkezetekről való döntések, de még a nem nyelvi folyamatok, a testtartás, a beszédtempó

vagy a hangmagasság összehangolása is (Ellis et al., 2016: 41). Az ismétléses előfeszítés ily módon a közös megértést segíti elő a dialógusok során.

### 3.2 Korpusznyelvészeti előzmények

A lexikai előfeszítés elméletét korpusznyelvészeti szempontból egészen az alkalmazott nyelvészet atyjának is nevezett Harold Palmerig lehet visszavezetni, aki az 1920-as években összegyűjtötte az angol nyelven leggyakrabban használt szavakat és kifejezéseket (1999) és megalkotta az először *Mechanism Grammar*-nek, majd később *Pattern Grammar*-nek, vagyis mintázatnyelvtannak<sup>5</sup> nevezett elméletet (Pace-Sigge, 2013a: 2), amelyet előbb Hornby (1954), majd Hunston és Francis (2000) dolgozott ki részletesebben és fejlesztett tovább. A mintázatnyelvtanról még a későbbiek folyamán részletesebben is szó esik (ld. II.3.2.1 fejezet).

Palmer után másodikként Firth (1957b) munkásságát kell megemlítenünk, aki egy új nyelvelméleti megközelítést alkotott. Firth szerint egy szó jelentése csak a kontextusában értelmezhető. A nyelv pedig mindig egy adott nyelvi közösség specifikus szokásaira épül, és ezek a szokások határozzák meg a nyelvhasználatot és annak jelentését. Továbbá azt is megállapítja, hogy egy nyelvész számára fontos a szógyakoriság és a szóeloszlás megfigyelése ezeknek a nyelvhasználati szokásoknak a megértésében. Az ő munkásságán alapulnak többek között a lexikai előfeszítés harmadik előzményének tekinthető Sinclair korpusznyelvészeti kutatásai is. Sinclair korpusznyelvészeti elemzése megerősítették, hogy a természetes nyelvhasználatban jelentős mértékben jelennek meg a szavak és nyelvi szerkezetek ismétlődő mintái. Sinclair (1991) a COBUILD [Collins Birmingham University International Language Database] projekt – amely során összeállította az angol nyelv első számítógép segítségével készült korpuszát – eredményeit idiómaelvében<sup>6</sup> [Idiom Principle] foglalta össze. Idiómaelve szerint a nyelvhasználónak nagyszámú, félig előre felépített szekvencia áll a rendelkezésére. Az általa idiómáknak nevezett előre felépített szekvenciák egyetlen választás eredményei, még akkor is, ha a szekvenciák elemezhetőnek tűnnek. Ez visszavezethető arra, hogy az emberek gyakran találkozhatnak hasonló helyzetekkel az életükben, amelyekre hasonló módon reagálnak, illetve annak is betudható, hogy szeretünk takarékoskodni és egyszerűen megoldani a problémákat. Ezenkívül a

---

<sup>5</sup> A fogalom fordítását Simon Gábortól (2012) vettem át.

<sup>6</sup> A fogalom fordítását Dóla Mónika tanulmányából (2014) vettem át.

spontánbeszéd gyorsasága is motiválhatja erre a nyelvhasználókat (1991: 110). Tehát Sinclair elmélete szerint a nyelvi tudás nagy mértékben nyelvi mintázatok [idiom patterns] ismeretéből épül fel. Ezeket a nyelvi mintázatok, szószerkezeteket könnyen megismerhetjük a korpuszokból, amelyek a természetesen megjelenő nyelv gyűjteményei [naturally occurring language].

Előzménynek tekinthető még Stubbs (1996, 2002) és Partington (1998) kollokációkkal kapcsolatos munkássága. Stubbs (1996, 2002) többek között arról ír, hogy a kollokációk, vagyis a gyakori szókapcsolatok elemzése segíthet a nyelvtanulóknak abban, hogy hatékonyabban sajátítsák el a célnyelvet. A kollokációk elemzése a nyelvi szintaktikai, szemantikai és pragmatikai jellemzőket is feltárhatja, ezáltal hozzájárulhat a nyelvészeti tudásunk bővítéséhez. Partington (1996, 2013) részletesen foglalkozik a kollokációkkal és azok szerepével a diszkurzusban. Hangsúlyozza a kollokációk fontosságát a szövegértésben, a nyelvhasználatban és az interakcióban. 2013-as könyvben számos példán keresztül mutatja be, hogyan lehet a kollokációkat nyelvi elemzés során használni, és hogyan lehet az eredményeket interpretálni a diszkurzus és a nyelvhasználat szempontjából. Szerinte a kollokációk megismerése a nyelvtanulás során hozzájárulnak a célnyelvi kifejezőeszközök megszerzéséhez és a nyelvi szerkezetek megértéséhez is. A kollokációk segítségével a nyelvtanulók megismerhetik azokat a szavakat, amelyek gyakran együtt fordulnak elő, és ezáltal hatékonyabban kommunikálhatnak anyanyelvi beszélőkkel.

### **3.2.1 Lexika, grammatika és szemantika kapcsolata a korpusznyelvészeti kutatások alapján**

A kognitív nyelvészeti megközelítés, amelyre a korpusznyelvészeti kutatások alapulnak a generatív grammatikával ellentétben nem próbálja a nyelvet olyan részekre bontani, mint a fonetika vagy a szemantika.

„A nyelvet az emberi megismerés részének tekinti, és a nyelv és a megismerés szisztematikusan összefonódik. [...] Egy nyelv elsajátítása magában foglalja a különböző összetettségi és sematizációs szintű konstrukciók ismeretét, valamint a kombinációjuk alapjául szolgáló valószínűségi (szemben a merev) tendenciák ismeretét” (Ellis et al., 2016: 31).

Sinclair (2004: 164) szerint a szavak kombinációi együtt alkotnak jelentést, nem pedig a különböző szójelentések adódnak össze egy-egy szerkezetben. Ennek a jelenségnek a leírására alkották meg a lexikogrammatika [lexico-grammar] fogalmát (Halliday, 1985). Sinclair COBUILD projektjéből is táplálkozott az előző részben (II.3.2 fejezet) már említett mintázatnyelvtan (Hunston – Francis, 2000). A mintázatok frazeológiai elemek, amelyek szavak és grammatikai struktúrák szintéziséből jöttek létre. Ezekből a mintázatokból megtudhatjuk, hogy a szavak jellemzően hogyan kapcsolódnak egymáshoz értelmes egységeket alkotva (Ellis et al., 2016: 31). Hunston és Francis (2000) korpuszelemzése – megerősítve Sinclair elméletét – két fontos eredményre jutottak. A többjelentésű szavak különböző jelentései más-más mintázatokkal együtt jelennek meg, illetve, hogy „bizonyos mintázatok hajlamosak lesznek olyan lexikai elemekhez társulni, amelyeknek egy bizonyos jelentésük van” (2000: 83). Ez Hoey lexikai előfeszítés elméletének (2005) is fontos részét képezi, amelyről a következő részben lesz bővebben szó.

### **3.3 A lexikai előfeszítés elmélete és legfontosabb kutatásai**

Az előző két részből (II.3.1 és II.3.2) láthattuk, hogy a pszicholingvisztikai és a korpusznyelvészeti kutatások szinte egymással párhuzamosan vontak le hasonló következtetéseket, erősítették meg egymás eredményeit. Hoey lexikai előfeszítés elméletével ezeket a területeket kívánta összehangolni. Bár ő 2005-ös könyvében inkább a korpusznyelvészeti szempontokat emelte ki, később Michael Pace-Sigge (2013b) a véleményem szerint szintén igen jelentős kutatásának elméleti keretében Hoey gondolatait további pszicholingvisztikai eredményekkel egészítette ki. A két kutatás együtt, egymást kiegészítve jelenti számomra a megbízható elméleti alapot a jelen doktori értekezésben bemutatott kutatáshoz.

Hoey lexikai előfeszítés (2005) fő gondolata azon a megfigyelésen alapul, amely szerint bizonyos szavak, illetve azok formái vagy vonzzák, vagy megpróbálják elkerülni egymást. Szerinte bármit hallunk, olvasunk, írunk le vagy mondunk ki azt erősíti meg bennünk, hogyan használunk bizonyos szavakat és hogyan nem (Pace-Sigge, 2013: 155). Kutatásában a korpusznyelvészet által leggyakrabban vizsgált három jelenséget figyelt meg:

1. Kollokációk (előzmények: Firth, 1957ab; Halliday, 1959; Sinclair, 1991; Stubbs,

1996; Partington, 1998), olyan szavak, amelyek gyakran megjelennek egymás társaságában (*döntést hoz*).

2. Kolligációk (előzmények: Firth, 1957ab; Halliday, 1959; Sinclair, 1991), vagyis grammatikai konstrukciókban szereplő szavak. Hoey angol példája a *that winter* [’azon a télen’] kifejezés, amely minden esetben múlt idejű mondatokban szerepel, amíg az *in winter* [’télen’] 50%-ban múlt idejű és 50%-ban jelen idejű mondatokban.
3. Szemantikai asszociációk (előzmény: Sinclair, 1991), amely kifejezéssel Hoey a korábbi szemantikai prozódia és a szemantikai preferencia fogalmát vonta össze (2005: 22–24). A szemantikai asszociációnál ismét a második pont példáit használta: az *azon a télen* kifejezés egy bizonyos eseményre utal, ezzel szemben a *télen* kifejezés általánosságra, időtlen igazságra.

Azzal kapcsolatban, hogy a szavak miért keresik bizonyos szavak társaságát és miért próbálnak elkerülni másokat, Hoey (2005: 7) megjegyzi, hogy a kollokációk egyik tulajdonsága, hogy szubverzívek. Ezzel arra utal, hogy a kollokációkat felépítő szavak gyakori együttes megjelenése azt erősíti meg bennünk, hogy az adott szókombináció egy természetes mintázat [natural pattern]. Amikor találkozunk egy bizonyos szókapcsolattal, az elménk automatikusan kapcsolatokat kezd kialakítani azzal. Az ismételt találkozások során a létező kapcsolatok vagy megerősödnek, vagy meggyengülnek (Hoey, 2005: 9) (vö. a II.5.3 fejezetben említett konnekcionalista elméletekkel).

A priming jelensége cselekvéshalmazokat vagy a nyelv esetében szókészleteket is a rendelkezésünkre bocsát (Pace-Sigge, 2013a: 151). A célszó előhívhatja például egy lexikai mező szavait: amikor meghalljuk vagy elolvassuk azt a szót, hogy *tulipán* sokkal gyorsabban ismerjük fel a vele azonos lexikai mezőhöz tartozó szavakat (*tavas*, *virág*, *madár*), mint az eltérőket (*orvos*). Ezen kívül Hoey szerint, amikor találkozunk egy szóval, szótaggal (lásd még III.3.1.1 fejezet: *fl* betűkombinációval kezdődő folyadékot jelentő angol szavak primingja) vagy szavak kombinációjával az előfordulásának kontextusát is memorizáljuk (vö. Halliday – Hasan, 1976: 274–292). Az ismételt találkozások során egyre több információt jegyzünk meg róla, ilyenek a kollokációi, tehát milyen szavak társaságában találkozhatunk vele, a kolligációi, vagyis, hogy milyen grammatikai mintákban szokott előfordulni, valamint a szemantikai és pragmatikai asszociációi is. Azt is kijelentette, hogy tudat alatt még a szavak jellegzetes diskurzusbeli megjelenéséről is információkat gyűjtünk, valamint azt is azonosítjuk, hogy mely műfajban, stílusban vagy



szituációban használjuk jellemzően az adott szót vagy szókombinációt (Hoey, 2005: 9–13).

**2. táblázat** A lexikai előfeszítés elméletének tíz pontja (Hoey, 2005: 13) alapján

**Minden szó előfeszített, hogy**

1. más szavakkal együtt forduljon elő. Ezek a kollokációi.
2. meghatározott szemantikai halmazokban forduljon elő. Ezek a szemantikai asszociációi.
3. egy adott pragmatikai funkcióval összefüggésben forduljon elő. Ezek a pragmatikai asszociációi.
4. bizonyos grammatikai pozíciókban és funkciókban (ne) forduljon elő. Ezek a kolligációi.
5. A szinonimák és hiponimák, vagyis az egy szemantikai mezőhöz tartozó szavak (tulipán, rózsza, szegfű = virágok) kollokációikban, szemantikai asszociációikban és kolligációik tekintetében különböznek egymástól<sup>7</sup>.
6. A többjelentésű szavak minden jelentésének kollokációi, szemantikai asszociációi és kolligációi különböznek egymástól.

**Minden szó előfeszített, hogy**

7. egy vagy több grammatikai szerepkörben forduljon elő. Ezek a grammatikai kategóriái.
8. a diskurzusban a kohézió bizonyos típusaiban (ne) vegyen részt. Ezek a szövegbeli kollokációi [textual collocations].
9. a diskurzusban bizonyos szemantikai kapcsolatokban forduljon elő. Ezek a szövegbeli szemantikai asszociációi [textual semantic associations].
10. a diskurzusban bizonyos pozíciókban (ne) forduljon elő. Ezek a szövegbeli kolligációi [textual colligations] (Hoey, 2005: 13).

Hoey szerint a nyelvhasználók mentális konkordanciasorokkal rendelkeznek, amelyeket a korpusznyelvészekhez hasonlóan használnak (2005: 14). A priming jelensége több szinten is megjelenik, nemcsak a szótagoknál, szavaknál, szókapcsolatoknál, hanem a

---

<sup>7</sup> vö. Meyer – Schvaneveldt (1976: 31): „Az azonban nem igaz, hogy a szoros jelentéskapcsolatok mindig megkönnyítik a szavak mentális feldolgozását. Egyes folyamatok valójában gátoltak, amikor két olyan szóval kell foglalkozniuk, amelyek jelentése rokon”.

hosszabb, több szókapcsolatból álló szekvenciáknál is jelen van. Ezt Krishnamurty után Hoey nesting-nek, vagyis beágyazódásnak nevezte (2005: 8). A beágyazódás akkor történik, amikor a priming eredményeként létrejött szókapcsolat egy újabb priming folyamat része lesz, amely során egy hosszabb szókapcsolatot, szekvenciát alkot, és az elsőként létrejött szókapcsolatot felépítő szavaktól független. A beágyazódás megkönnyíti a memória munkáját. A nyelvelsajátítással kapcsolatban Hoey megemlíti, hogy a gyermekek ezeknek a hosszabb szókapcsolatoknak az előfeszítését jegyzik meg előbb, majd később az azokat alkotó szavakét (vö. Tomasello, 2003; Götz, 2013; Babarczy et al., 2014).

„A szó (vagy szó szerkezet) elsajátítása és annak elsajátítása között, hogy milyen kollokációi vannak, nincs különbség, habár feltételezhetően a szó felismerésének elméletileg meg kellene előznie a szó visszatérő tulajdonságainak felismerését, az utóbbi folyamat elindulásához a szónak (legalább) kétszer elő kell fordulnia” (Hoey, 2005: 8).

A korábbi elméletekkel ellentétben Hoey megjegyzi, hogy a priming során a beszélő akaratlanul megismétli a nyelv egy aspektusát, amely megismétlése után a hallgatót is erre készíteti (vö. II. 3.2.1. Ismétléses előfeszítés a dialógusokban, Garrod – Pickering, 2009). A nyelv megismételt aspektusát viszont nem tekinthetjük teljes rendszernek, ahogy azt korábban gondolták (2005: 9). Hoey amellett érvel, hogy ebből következően a lexika sokkal rendezettebb, mint a grammatika. Ezzel utal Hopper (1987) emergens nyelvtanára [emergent grammar], amely szerint a grammatika nem más, mint a nyelvhasználati rutinok eredménye. Mivel minden egyes nyelvhasználó tapasztalatai mások, ezért minden egyes nyelvhasználó a grammatikájában is eltér másoktól. Ehhez hasonlóan az egyének priming tapasztalatai is eltérnek egymástól, de nem térhetnek el annyira, hogy a beszélgetőtársak ne értsék meg egymást. Ezért van szükség a primingok harmonizálására, amelyre például az oktatás szolgálhat (Hoey 2005: 11).

Hoey a primingnek két típusát különbözteti meg. A receptív és produktív primingot, amelyek közül az előbbi olyan kontextusban valósul meg, ahol nem vehetünk részt aktívan (tévénézés), az utóbbi pedig értelemszerűen ennek az ellentéte (diskurzus), amikor szükséges aktívnak lennünk. Ezenkívül megemlíti ő is a kontextus hatását, illetve, hogy kontextusnak számíthatnak maguk a nyelvhasználók is (vö. Ellis et al., 2016). Írók szövegeinek olvasása közben az olvasó a priming jelenségét tapasztalhatja. A szöveg

olvasása közben egy idő után meg tudja jósolni, hogy mely szavak következnek az éppen olvasott szerkezetben, de általában ezek csak ideiglenes priming hatásnak minősülnek, nincs minden esetben hosszútávú hatásuk (Hoey, 2005: 11–12). A priming jelensége megjelenik a humorban és az irodalmi kreativitásban is, amely abból adódik, hogy a szerző szándékosan eltér a domináns lexikai elemtől ezzel úgynevezett priming konfliktust okozva, amely miatt a szerkezet érdekessé válik az olvasó számára (Hoey, 2005: 169–177).

### 3.3.1 A lexikai előfeszítés a beszélt nyelvben

A lexikai előfeszítés elméletet Hoey egy általa épített főleg írott nyelvet tartalmazó korpuszon vizsgálta meg. A korpuszt a Guardian című újság tíz évet felölelő cikkeiből (95 millió szó), valamint a British National Corpus írott (3 millió szó) és beszélt (230.000 szó) nyelvi részéből állította össze. Michael Pace-Sigge (2013a: 160) szerint az elmélet talán jobban bizonyítható, a priming jelensége pedig jobban megfigyelhető egy beszélt nyelvi korpuszt vizsgálva, amelyet a beszélt nyelv sajátosságaival indokol:

1. a diskurzus résztvevői spontánbeszéd közben időhiány miatt nem gondolják annyira végig a megnyilatkozásaikat, mint írás közben,
2. valamint, hogy a témák, a regiszterek a beszélgetések közben, a szituációhoz igazodva változnak.

A diskurzusban résztvevők a fent említett két ok miatt fokozottabban támaszkodnak az előre felépített egységekre, a már megszokott gyakran használt lexikai egységekre megkönnyítve ezzel a kommunikációt.

Pace-Sigge (2013b: 4) állításának indoklásaként még idézi Stubbs (1996: 72) eredményeit, amikor azt állítja, hogy a beszélt nyelvi szöveg lexikai sűrűsége kevesebb, mint az írott szövegé, valamint, hogy Wray (2002: 46) és O’Keeffe et al. (2007: 159) kijelentették, hogy a beszélt nyelvi szövegekben nagy arányban jelennek meg formulaszerű egységek. Kutatását emiatt Pace-Sigge egy beszélt nyelvi korpuszon végezte, a Liverpool English Corpus-on (SCO) vizsgálta meg a lexikai előfeszítés elméletét. Az SCO korpuszt alkotó informális beszélgetések hanganyagait 2002 és 2005 között vették fel, összesen ötven liverpooli nyelvjárásban beszélő adatközlővel, akik 8 és 80 éves kor közöttiek voltak. A korpusz összesen 119.079 szót tartalmaz (Pace-Sigge, 2013b: 61–62). Pace-Sigge a liverpooli nyelvjárást a British National Corpus beszélt nyelvi szövegeivel összehasonlítva vizsgálta. Az eredményei pedig megmutatták, hogy milyen eltérések találhatók bennük a

lexikai elemek mintázatos csoportosításával kapcsolatban (ilyen a *yeah* és *yes*, vagy *just* és *like* használata). Bár Lester (2015: 356–358) szerint a könyv nehezen olvasható és a módszertanában is vannak megkérdőjelezhető aspektusok, amivel egyetértek, véleményem szerint (és ezt Lester is kiemeli) Hoey munkáját jól kiegészíti. Elsősorban e két mű feldolgozását követően kaptam világos képet arról, hogyan függenek össze a korpusznyelvészeti, pszicholingvisztikai, valamint kognitív nyelvészeti kutatások.

### 3.3.2 Lehet-e vizsgálni a primingot korpusznyelvészeti eszközökkel?

Hoey előtt a priming jelenségét nem vizsgálták korpusznyelvészeti eszközökkel, mert a pszicholingvisztikai keretben dolgozó kutatók a kísérleti kutatásokat tartották erre megfelelő eszköznek. Azóta és Pace-Sigge (2013a és 2013b) beszélt nyelvi korpuszon történt kutatása előtt, Gries (2005) és Ellis et al. (2008) készített vizsgálatokat a témával kapcsolatban. Gries (2005: 365) vizsgálata azért jelentős, mert az eredményeit összevetette a kísérleti kutatásokban kapott eredményekkel és kijelentette, hogy azok nagyon hasonlóak<sup>8</sup>. Tehát mindkét megközelítés, a kísérleti kutatás és a korpuszalapú kutatás is igen jól használható a priming vizsgálatára. Gries kutatását Ellis et al. (2008) megismételte és kiterjesztette a nem anyanyelvi beszélőkre is, amelynek eredményeire a lexikai előfeszítés nyelvoktatásban betöltött szerepéről szóló részben (II.3.4 fejezet) fogok részletesebben kitérni.

## 3.4 A lexikai előfeszítés jelentősége a nyelvoktatásban

A lexikai előfeszítés folyamata akkor megy végbe, ha a nyelvhasználó többször találkozik az adott szóval vagy szerkezettel, és ezeknek a találkozásoknak a tapasztalatait felhasználva tud általánosítani. Nem csak a gyakoriság a fontosabb az ilyen találkozások során, hanem a figyelem ráirányítása az adott szerkezetre is elengedhetetlen. A nyelvtanulóknak tehát időre van szükségük ahhoz, hogy a második nyelvükön is kialakuljon ez a jelenség, ezért lényeges, hogy minél többször találkozzanak vele és vegyék is észre. A nyelvoktatásban

---

<sup>8</sup> Részlet a kutatás szövegéből:

„A jelen adatokban a előfeszített szerkezet és a nem előfeszített szerkezetek aránya 1,5 és 1,9 a prepozicionális dativumok és a ditranzitívumok esetében. Összehasonlításképpen Bock (1986: 364) klasszikus tanulmányában (...) a megfelelő arányok 1,5 és 2,1 a prepozicionális dativusok és ditranzitívusok esetében; az ő és az én arányaim közötti különbségek nyilvánvalóan elhanyagolhatóak. (...)” (Gries, 2005: 373).

használt tananyagok és maguk a nyelvtanárok is elősegíthetik ezt a folyamatot (Hoey, 2005: 185–186). Hoey azt tanácsolja a nyelvtanároknak, hogy lehetőség szerint autentikus adatokat használjanak, amikor a céljuk a lexikai priming kialakulásának elősegítése, amelyet a nyelvoktatás hatékonysága érdekében érdemes a céljaik közé sorolni. A módszertannal kapcsolatban Hoey követendő példaként említi meg Lewis (1993) lexikai megközelítését, Willis (2003) könyvét a lexikai és grammatikai mintázatokkal kapcsolatban, valamint Woolard (2004) kurzuskönyvét, amelyben a szerző célja a leggyakoribb kollokációk megismertetése a nyelvtanulókkal. A hasznos módszertani példák mellett arra is kitér, hogy vannak olyan példák is, amelyek hátráltatják vagy megakadályozzák a priming jelenségének kialakulását. Ilyen lehet, ha egy kurzuskönyv túlhangsúlyoz egy adott szerkezetet, illetve a nyelvtani szabályok olyan módon történő bemutatása, amely nem kapcsolódik ahhoz a tipikus lexikához, amellyel használjuk az adott szabályt. Hoey azt is kiemeli, hogy fontos mind az írott, mind a beszélt nyelven történő priming folyamatának elősegítése, hiszen a nyelvtanulóknak nemcsak a betűkombinációkat, hanem a hozzájuk kapcsolódó hangokat is külön meg kell ismerniük. Megemlíti Krashen (1985) input hipotézisét is, amely szerint a nyelvtanulóknak a nyelvelsajátításhoz nagy nyelvi inputra van szükségük, de ami még ennél is fontosabb, hogy az érthető input [comprehensive input] legyen. Vagyis csak egy lépéssel legyen a nyelvtanuló aktuális nyelvi szintje fölött. Hoey ezzel kapcsolatosan azt is megjegyzi, hogy új asszociációs kapcsolatok is csak ebben az esetben alakulnak ki. Hiszen, ha az input nyelvi szintje túl alacsony, a nyelvtanuló „egyszerűen kikapcsol”. Ha pontosan azon a nyelvi szinten van, akkor az hasznos a már ismert primingok megerősítésében, de ebben az esetben új kapcsolatok nem alakulnak ki (Hoey, 2005: 187).

Ellis et al. (2008) kutatásában kitér még egy aspektusra, amelyet fontos megjegyezni a nyelvtanulókkal kapcsolatban. Szerintük a priming eltérő módon működik a felnőtt anyanyelvi és a nem anyanyelvi beszélők esetében. Az anyanyelvi beszélőkkel ellentétben, a nyelvtanulókra nagyobb hatással van a gyakoriság, mert nekik szükségük van a többszöri találkozásokra ahhoz, hogy rájőjjenek a szókapcsolatok formulaszerűségére. Ez azt jelenti, hogy ők minden esetben „a magas gyakorisági mutatókkal rendelkező mintákat dolgozzák fel folyékonyabban” (Ellis et al., 2008).

## 4. A formulaszerű nyelvhasználat

Számos kutatás kimutatta, hogy a spontán beszédfeldolgozáshoz kapcsolódó kognitív terhelés miatt a megnyilatkozások „korlátozott és ismétlődő repertoárral” rendelkeznek (Biber et al., 1999: 1049), valamint formulaszerű, előre megszerkesztett [pre-fabricated] lexikai egységekből állnak (Wray, 2002). Azt is megállapították, hogy az anyanyelvi beszélők diskurzusában magas a formulaszerű lexikai elemek aránya (vö. Altenberg, 1998; Biber et al., 1999; Pawley – Syder, 1998). Ezt bizonyítja Altenberg kutatása, aki a London-LundCorpus, angol beszélt nyelvi korpusznak a vizsgálatával arra a következtetésre jutott, hogy a rutinná vált, részben vagy teljesen előre elkészített struktúrák a megnyilatkozásoktól a diskurzusig, a nyelvi szerveződés minden szintjére hatással vannak. Az általa talált struktúrák, amelyek leginkább többé-kevésbé konvencionizált kifejezéseként írhatók le, a teljesen lexikalizált egységek és a szabad szerkezetek határán helyezkednek el. Tanulmányában azt is megbecsülte, hogy a korpuszban szereplő szavak több mint nyolcvan százaléka ilyen szókombinációk részét képezi (Altenberg, 1998). Egy másik korpusznyelvészeti vizsgálatban Erman és Warren (2000) a London-LundCorpus mellett összesen tizenkilenc mintát vizsgált meg a Lancaster-Oslo-Bergen írott nyelvi korpuszból. Számításaik alapján a formulaszerű nyelvhasználat a kompetens nyelvhasználók beszédének és írásának körülbelül ötvenöt százalékát tette ki. Biber et al. (1999) kutatásában azt találta, hogy a többszavas kifejezések a beszélt nyelv huszonnyolc százalékát, az írott diskurzus húsz százalékát tették ki. Lewis pedig arra a következtetésre jutott, hogy „a nyelv olyan egységekből [chunks] áll, amelyek kombinálva folyamatos, koherens szöveget hoznak létre” (Lewis, 1997: 7).

Az említett kutatások eredményeiből láthatjuk, hogy a formulaszerű egységek a beszélt és írott nyelv jelentős részét képezik. Götz *Fluency in Native and Nonnative English Speech* (2013) című könyvében szintén a lexikai egységek szükségességéről ír, amelyeket többszavas szekvenciáknak [multiword sequences] nevez. Arra a következtetésre jutott, hogy ezeket a többszavas szekvenciákat gyermekkorban rutinszerűen sajátítjuk el (vö. II.2.1 fejezet: Gyermek anyanyelv-elsajátítása a használatalapú elméletek alapján), de később felnőttkorunkban fejlődnek tovább, egészülnek ki, illetve ekkor értelmezzük őket. A többszavas szekvenciák használata beszéd közben nem igényel nagy figyelmet, illetve a megnyilatkozások kiterjesztése sem követel meg kognitív erőfeszítést. Wray (2002) szintén azt állítja a kollokációkkal kapcsolatban, hogy formulaszerű egységekként az első nyelv

elsajátításában fontos szerepet játszanak. A II. 2 fejezetben említett első nyelv elsajátítását vizsgáló kutatások használatalapú nyelvelméleti modellek kidolgozása során is erre a következtetésre jutottak. Tehát elmondható, hogy a formulaszerű egységeket először elemzés nélkül memorizálják a gyermekek, és csak később kezdik analizálni őket (Tomasello, 2003: 305–307).

A kutatások eredményeit figyelembe véve megállapítható, hogy a kompetens nyelvhasználók jelentős számú formulaszerű lexikai elemet ismernek és használnak aktívan. Valamint arra is következtethetünk, hogy a formulaszerű egységek használata valóban lényeges része a kommunikációnak. A fejezet további részeiben a formulaszerű nyelvhasználat jellemzőit, az egységek beazonosítási lehetőségeit, valamint a nyelvtanításban való jelentőségét mutatom be.

#### **4.1 A formulaszerű egységek jellemzői**

Ahogy az előző részben említésre került elmondható, hogy a formulaszerű nyelvhasználat nagyban hozzájárul a beszédprodukciónak hatékonyságához, természetességéhez és folyékonyságához, hiszen az anyanyelvi beszélők is nagymértékben támaszkodnak ezekre a szerkezetekre.

Wray szerint a formulaszerű szekvencia kifejezés minden olyan többszavas láncra utal, amelyet „a nyelvhasználó úgy érzékel, hogy egyetlen lexikai egységként azonosítható vagy hasznosítható” (2019: 267). A formulaszerű nyelvhasználatot részletesebben a következőképpen határozza meg:

„[A formulaszerű szekvenciák] szavak vagy más elemek folytonos vagy nem folytonos sorozata, amely előre gyártott, vagy annak tűnik, azaz a használat időpontjában nem generálás vagy grammatikai elemzés által keletkezik, hanem inkább egészben tárolódik és hívható le a memóriából” (Wray, 2002: 9).

3. táblázat A formulaszerű nyelvhasználat aspektusai (Wray, 2002:9) alapján

| A formulaszerű nyelvhasználat aspektusai |   |                                       |     |                               |                          |     |   |  |
|--|---|---------------------------------------|-----|-------------------------------|--------------------------|-----|---|--|
| Sz.                                      | Magyar                                  | Angol                                 | Sz. | Magyar                        | Angol                    | Sz. | Magyar  | Angol                                      |
| 1.                                       | állandósult kifejezések                 | [set phrases]                         | 20. | gambitok                      | [gambits]                | 39. | nem kompozíciós                                 | [noncompositional]                         |
| 2.                                       | amalgámok                               | [amalgams]                            | 21. | gestalt                       | [gestalt]                | 40. | nem komputációs                                 | [noncomputational]                         |
| 3.                                       | analizálatlan beszédegységek            | [unanalyzed chunks of speech]         | 22. | holofrázisok                  | [holophrases]            | 41. | nem produktív                                   | [nonproductive]                            |
| 4.                                       | analizálatlan többszavas beszédegységek | [unanalyzed multiword chunks – units] | 23. | idiómák                       | [idioms]                 | 42. | nem propozicionális                             | [nonpropositional]                         |
| 5.                                       | automata                                | [automatic]                           | 24. | idiomatikus                   | [idiomatic]              | 43. | összetételek                                    | [composites]                               |
| 6.                                       | befagyasztott kifejezések               | [frozen phrases]                      | 25. | ismétlődő megnyilatkozások    | [recurring utterances]   | 44. | összetett lexémák                               | [complex lexemes]                          |
| 7.                                       | befagyasztott metaforák                 | [frozen metaphors]                    | 26. | kész kifejezések              | [ready-made expressions] | 45. | petrifikációk                                   | [petrifications]                           |
| 8.                                       | bemagolás                               | [rote]                                | 27. | kész megnyilatkozások         | [ready-made utterances]  | 46. | praxonok  | [praxons]                                  |
| 9.                                       | egységek                                | [chunks]                              | 28. | készletezett megnyilatkozások | [stock utterances]       | 47. | rögzített kifejezések                           | [fixed expressions]                        |
| 10.                                      | előre felépített mintázatok             | [prefabricated patterns]              | 29. | klisék                        | [clichés]                | 48. | rögzített kifejezések az idiómákat is beleértve | [fixed expressions including idioms - FEI] |
| 11.                                      | előre felépített rutinok                | [prefabricated routines]              | 30. | kollokációk                   | [collocations]           | 49. | rutinszerű formulák                             | [routine formulae]                         |
| 12.                                      | előre kódolt konvencionizált rutinok    | [precoded conventionalized routines]  | 31. | konstrukciók                  | [constructions]          | 50. | sémák   | [schemata]                                 |



| A formulaszerű nyelvhasználat aspektusai |  |  |     |                           |                              |     |   |  |
|--|--|--|-----|---------------------------|------------------------------|-----|---|--|
| Sz.                                      | Magyar   | Angol  | Sz. | Magyar                    | Angol                        | Sz. | Magyar  | Angol  |
| 13.                                      | előre összeállított beszéd   | [preassembled speech]  | 32. | konvencionizált formák    | [conventionalized forms]     | 51. | stabil és ismerős kifejezések speciális alszavakkal | [stable and familiar expressions with specialized subsenses] |
| 14.                                      | félkész mondatok, amelyek egyetlen választási lehetőséget jelentenek | [semi-preconstructed phrases that constitute single choices] | 33. | koordináták               | [co-ordinate]                | 52. | szabálytalan  | [irregular]  |
| 15.                                      | formulák   | [formulas/formulae]  | 34. | lexikai szimplex          | [lexical simplex]            | 53. | szintetikus   | [synthetic]  |
| 16.                                      | formulaszerű beszéd  | [formulaic speech]   | 35. | lexikálizált kifejezések  | [lexicalized phrases]        | 54. | sztereotípiák                                       | [stereotypes]  |
| 17.                                      | formulaszerű nyelv   | [formulaic language]   | 36. | lexikalizált mondatrészek | [lexicalized sentence stems] | 55. | sztereotipizált kifejezések                         | [stereotyped phrases]  |
| 18.                                      | fossilizálódott formák   | [fossilized forms]   | 37. | lisztémák                 | [listemes]                   | 56. | többszavas elemek/egységek                          | [multi-word items/units]                                     |
| 19.                                      | frazémák   | [phrasemes]  | 38. | mondatépítő elemek        | [sentence builders]          | 57. | többszavas lexikai jelenségek                       | [multi-word lexical phenomena]                               |

A jelenség leírására számos kifejezés található a szakirodalomban (vö. Nelson, 2018: 433). Wray (2002) könyvében több mint ötven olyan fogalmat gyűjtött össze (3. táblázat), amely különböző nézőpontokból közelíti meg a formulaszerű egységeket. A kifejezések között vannak, amelyek a jelentést (idiómák) és a szerkezetet (lexikalizált mondatrészek), mások pedig rögzítettségüket (többszavas egységek) hangsúlyozzák (Hunston, 2022: 102). Bár a felsorolt kifejezések között van némi különbség, mégis mindegyik ugyanazt a jelenséget próbálja leírni: olyan szavak csoportjait, amelyek gyakran fordulnak elő együtt, egymás társaságában (vö. Firth, 1957ab; Sinclair, 1991; Hoey, 2005).

A formulaszerű szekvenciák mellett az angol szakirodalomban főként a lexikai egységek [lexical chunks], a lexikai kötegek [lexical bundles] és a többszavas kifejezések [multi-word expressions] használatosak. Siyanova-Chanturia és Pellicer-Sánchez (2019: 2) szerint a legmegfelelőbb a formulaszerű nyelvhasználat kifejezést használni, mert ez a kifejezés egyszavas egységeket is magában foglal. Ebbe a kifejezésbe bele tartozhatnak a felkiáltások (Hurrá!) vagy különféle beszédformulák (Ja!) is. Ugyanakkor, ha csak a többszavas kifejezések [multi-word expressions] fogalmat használnánk, akkor a fent említettek kikerülnének a definícióból, pedig a többszavas egységekhez hasonlóan működnek. A formulaszerű nyelvhasználat kifejezés használata a mi szempontunkból is megfelelőbb, hiszen a magyar nyelv agglutináló mivoltából adódóan az indoeurópai nyelvekben több szóval kifejezett jelentéseket, sokszor egyetlen szó használatával képes átadni.

Ahogy Wray felsorolásából láthatjuk, a formulaszerű nyelvhasználat fogalma alá rendkívül sok hasonló, mégis néhány apróságban eltérő jelenség került besorolásra (Nelson, 2018: 433). A felsorolt elemeknek két jellemzője közös. Egyrészt visszatérők, azaz a természetes nyelvhasználatban gyakrabban megjelennek, mint az újonnan készített szerkezetek. Másrészt pedig gyorsabban feldolgozhatóak, mint az új szerkezetek (Carrol – Conklin, 2020: 96).

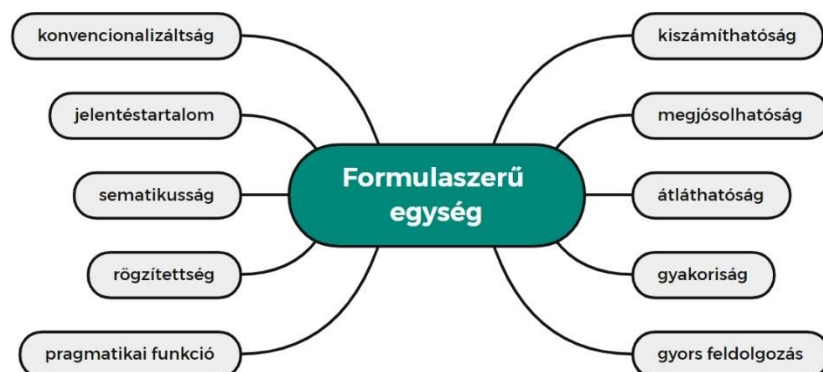
Ugyanakkor a formulaszerűnek nevezett elemek több fontos, karakterizáló tényezőben eltérnek egymástól (Carrol – Conkin, 2020: 96). Ilyen a rögzítettségük, illetve konvencionalizáltságuk foka, sematikusságuk, vagyis mennyire variálhatóak, a szemantikai egységük, a kompozíciós fokuk, valamint, hogy milyen pragmatikai funkciót látnak el (vö. Buerki, 2016). Carrol és Conkin (2020) tanulmányában megemlíti négy kutatást a formulaszerű nyelvhasználat jellemzőit illetően, amelyeket a jelen kutatás szempontjából is érdemes röviden összefoglalni.

- Az első Columbus (2010) kutatása, aki különböző formulaszerű egységeket: idiómák, lexikai kötegek, valamint kollokációk olvasási idejét hasonlította össze egymással. Az eredményei arra utaltak, hogy a formulaszerű egységeket sokkal gyorsabban olvasták el

a kutatásban résztvevők, mint a nem formulaszerűeket, és azok közül is a leggyorsabban elolvasottak az idiómákat találta, amely a meghatározása szerint átvitt értelmű kifejezések (például angolul *Break a leg*. [’Kéz és lábtörést!’]). Később a kutatás kiterjesztésében (2013) megállapította, hogy a formulaszerű egységek jellegzetes tulajdonsága a gyakoriság, az ismerősség [familiarity], valamint az érzékelt szemantikai átláthatóság [perceived transparency].

- Jolsvai, McCauley és Christiansen (2013) vizsgálatuk során megállapították, hogy a formulaszerű szekvenciák feldolgozását még a gyakoriságuknál is jobban befolyásolta a jelentéstartalmuk.
- Gyllstad és Wolter (2016) kutatását is érdemes megemlíteni, amely során Columbushoz (2013) hasonlóan arra az eredményre jutottak frázisos döntési feladatokat alkalmazva [phrasal decision tasks], hogy a gyakoriság mellett a szemantikai átláthatóság [transparency] is gyorsabb feldolgozási sebességhez vezetett.
- Az utolsó kutatás pedig Carrol és Conklin (2020) saját vizsgálata, amelyben a formulaszerű egységek közül az idiómákat, binominális elemeket, vagyis két szót egy és kötőszóval összekötött szekvenciákat (például angolul *salt and pepper* [’só és bors’]), valamint a kollokációkat hasonlították össze. Az eredményeik alapján a gyakoriság mellett a kiszámíthatóságot, megjósolhatóságot [predictability] is a formulaszerű egységek jellemzői közé sorolták. Szignifikáns különbséget nem találtak az közöttük, de azt meg tudták erősíteni, hogy minden formulaszerű egységet gyorsabban dolgozták fel a kutatásban résztvevők, mint a nem formulaszerűeket.

**1. ábra** A formulaszerű egységek jellemzői



A formulaszerű nyelvhasználat altípusainak néhány meghatározása fő tulajdonságként

említi meg, hogy ezek a nyelvi elemek a kognitív terhelést csökkentve holisztikusan tárolódnak és hívódnak elő a memóriából (Wray, 2002). A holisztikus tárolás és előhívás kérdése azonban még nem került bizonyításra (Siyanova-Chanturia – Pellicer-Sánchez, 2019: 5), ezért nem ért vele egyet minden szerző. Annál is inkább nem, mert a formulaszerű szekvenciák sokszor nem egymás szomszédaiként jelennek meg. Sinclair (1991) is kijelentette, hogy a kollokációk nem mindig találhatóak meg közvetlenül egymás mellett. A szövegben többnyire maximum a vizsgált szó bármely oldalán, tehát előtte vagy utána öt szóval arrébb jelenhetnek meg. Ahogy a szemantikai előfeszítésről szóló fejezetben említettük, ez Ratcliff és McKoon (1988) szerint akár több szó is lehet (ld. II.3.1 fejezet). Egy másik kutatásban Vilkaite (2016) szemmozgáskövető vizsgálatával kimutatta, hogy a főnév és ige kollokációkat (*információt ad*) az anyanyelvi beszélők gyorsabban olvassák el, mint a kontroll mondatokban szereplő egyéb kifejezéseket (*információt összehasonlít*). Ez akkor is igaz, ha a főnév és ige kollokáció egymás mellett található, de akkor is, ha három szó választja el őket (*az információ egy kis részét adta át*). Sinclair és Vilkaite vizsgálatai a kollokációkat illetően a holisztikus hipotézis ellen értelmezhetőek, ugyanakkor alátámasztják a lexikai előfeszítés (Hoey, 2005) valamint a valószínűség-alapú modellek [probability-based models] (vö. Barlow – Kemmer, 2000; Kuperberg – Jaeger, 2016) elméletét, a pszicholingvisztikai priming vizsgálatok (vö. Myles – Cordier, 2017) eredményeit, valamint a Cheng, Greaves, Sinclair és Warren (2009) által leírt congruence fogalmát is (Carroll – Conkin, 2020: 100), amely a szavak együttes előfordulását jelenti, függetlenül attól, hogy hol helyezkednek el.

## **4.2 A formulaszerű nyelvhasználat jelentősége a nyelvoktatásban**

Ahogy a fejezet elején említettem, a formulaszerű nyelvhasználat elengedhetetlen része a kompetens nyelvhasználók kommunikációjának, ezért a nyelvtanulók számára is fontos eszköz. A közelmúltban számos tanulmányt szenteltek a formulaszerű nyelvhasználatnak. A formulaszerű egységek elsajátítása hatékonynak bizonyult a második nyelven történő kommunikáció képességének fejlesztésében, és hozzájárul a beszéd folyékonyágához (Wray-Fitzpatrick, 2008) is. Segalowitz (2010: 126) azt állítja, hogy „a formulaszerű szekvenciák helyes használatának képessége hozzájárul a beszéd anyanyelvihez hasonló természetességéhez, és az üzenetfeldolgozási terhelés modulálásához könnyebbé és hatékonyabbá téve ezáltal a kommunikációt”.

Azok a nyelvtanulók, akik nem tudják megfelelően használni ezeket a lexikai egységeket, nem fognak profitálni a formulaszerű egységekből, és ez a hiány növelheti

beszédfeldolgozási terhüket, amely veszélyeztetheti a beszédfolyékonyságukat. Hill (2000) azt is állítja, hogy azok a tanulók, akiknek hiányzik a kollokációs kompetenciájuk, azok hosszabb, terjedősebb kifejezéseket használnak sok grammatikai hibával és rendszeresen küzdenek az önmaguk megértésével.

- (1) őszintén szólva
- (2) jó emlékem van a N\_Ins + kapcsolatban

(KorSzak Gyermeeknyelvi Korpusz, 2020)

Nyelvtanulói szempontból fontos megjegyezni azt is, hogy amellet, hogy természetesebbé teszik a beszédünket például a fenti első példához hasonló rögzült formulaszerű egységek, a második példához hasonló részben rögzült formulaszerű szekvenciák ezenkívül nagy kommunikációs értékkel is bírnak. Hiszen csak egyszer kell megtanulnunk őket, de többféleképpen, variációkban, időnként többféle kontextusban is használhatóak (Vermeki, 2021bc).

Lewis (1993) lexikai megközelítése [Lexical Approach] ezeken az gondolaton alapul. A formulaszerű nyelvhasználat azonban már az audiolingvális módszer elterjedése óta jelentős szerepet játszik a nyelvoktatásban (Tavakoli, 2020: 32). Ugyanakkor még ennél is korábban, az audiolingvális módszer elterjedése előtt, a korábban említett (vö. II. 3.2 fejezet) Harold Palmer kijelentette az angol társalgási nyelvről, hogy fejlesztésének alapvető irányelve lehet a gyakori, hasznos szavak csoportjainak memorizálása (1999: 185).

A formulaszerű szekvenciák, a kompetens nyelvhasználókhöz hasonló kifejezések aktív használata nemcsak a beszéd folyékonyságát, hanem a grammatikailag pontos nyelvhasználatot is támogatja (Kirk, 2014: 105). Ez annak köszönhető, hogy a gyakran ismétlődő lexikai egységek felismerése lehetővé teszi a nyelvi input gyorsabb feldolgozását. Kimutatták, hogy a nyelvtanulók gyorsabban felismerik a vizsgált szavakat, ha olyan szavak társaságában hallják őket, amelyekkel általában megjelennek. Hasonlóképpen, szemmozgáskövető vizsgálatok (Siyanova-Chanturia et al., 2011; Pellicer-Sánchez et al., 2022) azt is kimutatták, hogy a kísérletben résztvevők gyorsabban el tudták olvasni az ismerős lexikai egységekből álló bekezdéseket.

Az osztálytermi környezetben a formulaszerű egységek forrásaként egyrészt a tanár, másrészt a tananyagok szolgálhatnak (Meunier, 2012: 113). A nyelvtanulók a formulaszerű egységek a gyakori nyelvhasználat által tudják memorizálni. Ezt erősítik a konnekcionalista

elméletek (ld. II.5.3 fejezet), amelyek szerint folytonos ismétlés által az idegpályák megerősödnek, amikor azonban nincs ismétlés, ezek az idegpályák elgyengülnek, ami felejtéshez vezet (Ellis et al., 2016). Az, hogy melyik formulaszerű egységek, szekvencia lexikalizálódik, a gyakoriság típusától is függ. Bybee (1995 idézi Lieven – Tomasello, 2008) kétfajta gyakoriságot különböztetett meg egymástól: a tokengyakoriság [token frequency], amely a tényleges szavak előfordulásának számát jelöli, valamint a típusgyakoriság [type frequency], amely alatt egy adott nyelvi mintában előforduló különböző lexikai elemek számát értjük. Magas tokengyakoriság esetén az adott lexikai elemhez való hozzáférés erőssége nagyon erős, így jegyezzük meg például a kollokációkat. Bybee példája erre a *take a break*, vagyis a *szünetet tart* kifejezés. Magas típusgyakoriság esetén az adott nyelvi jelenség több lexikai elem is megjelenik, ami könnyebben eredményez általánosítható mintát (Lieven – Tomasello, 2008: 174–179). Bybee (2008: 218) szerint a típusgyakoriság sokszor morfológiai mintázatokra vonatkozik, erre a szabályos és szabálytalan angol múlt idejű igéket hozza példaként. Az olyan múlt idejű alakok, mint a *go – went* [’megy – ment’] típusfrekvenciája alacsonyabb, mint a szabályos *-ed* végződésű igéké. Így az utóbbiból gyakrabban általánosíthatunk, ezért hallhatjuk angol anyanyelvű kisgyermektől, illetve nyelvtanulóktól sokszor a *go* ige egyszerű múlt idejű alakja a *went* helyett a *\*goed* túláltalánosítást. Magyar példa lehet erre a tárgyrag (*\*pénzet*), vagy a *-zik* igeképző túlhasználata (*\*hózik*).

Összefoglalva tehát elmondhatjuk, hogy a formulaszerű egységek alapvető fontosságúak a nyelvfeldolgozás és a nyelvi produkció szempontjából is. Ezek a formulaszerű egységek gyakran fordulnak elő a kompetens nyelvhasználók nyelvében, megfelelő használatukkal a természeteshez hasonló lesz a tanulók nyelvhasználata. Használatuk fejleszti a produktív készségeket (beszéd- és íráskészség), mert a formulaszerű egységek memóriából való gyors felidézésével felszabaduló idő és energia más területeken is hasznosítható. A nyelvtanulók beszéde folyékonyabbá válik ezeknek a szekvenciáknak a használatával. Ugyanakkor fejleszti a receptív készségeket (olvasás, hallás utáni szövegértés) is, mivel a gyakran ismétlődő lexikai egységek felismerése lehetővé teszi a nyelvi input gyorsabb feldolgozását. Ha olvasás vagy egy szöveg hallgatása közben rendelkezésünkre állnak formulaszerű egységek gyűjteményei, akkor a kompetens nyelvhasználókhöz hasonlóan, gyorsabban és hatékonyabban értjük meg a szöveg jelentését.

## 5. A beszédfolyékonyság

Az eddig bemutatott elméletek (a használatalapú nyelvelméleti modellek, a lexikai előfeszítés jelensége, a formulaszerű nyelvhasznát) mind szoros összefüggésben állnak a fluensség fogalmával. Ahogy a bevezetőben említettem nyelvtanulóként legfőbb célunk a természetes, vagyis az anyanyelvi beszélők megnyilatkozásaihoz hasonló folyékony nyelvhasználat elsajátítása. A fluensség fogalma mind a négy nyelvi alapkészségre (olvasás, írás, beszédértés, valamint beszédképesség) vonatkozhat, mégis mivel a (poszt)kommunikatív nyelvoktatásban a hangsúly eltolódott a szóbeli kommunikációra, a legtöbbször a beszédképességgel hozzuk összefüggésbe. Mivel ez a kutatás egy beszélt nyelvi korpuszon történik, ezért ebben az értekezésben is ez kerül kiemelésre.

### 5.1 A beszédfolyékonyság meghatározása

Komplex, nehezen megfogható jelenség révén, nem könnyű definiálni, hogy mit jelent az, amikor valaki folyékonyan beszél egy nyelvet, legyen az az első nyelve vagy egy második nyelv. A beszédfolyékonyságot nem lehet önmagában értelmezni, számos összetevőből áll, és számos faktor van rá hatással. Magyarázataira az 1970-es évektől kezdve születtek különböző értelmezések és modellek. A legegyszerűbb definíció szerint, aki folyékonyan beszél, az érthetően, pontosan fejezi ki magát és nem akad meg vagy ugrik vissza a beszédében (Byrne, 1976). Fillmore (1979) ehhez még hozzáteszi a koherensséget, a kreativitást, az ötletességet és a megfelelés kritériumait. Az ezen a területen legtöbbet idézett Lennon (2000: 26) szerint emellett fontos az is, hogy a gondolatait világosan át tudja adni, illetve attól függ, hogy mennyire érezzük folyékonynak valakinek a beszédét, hogy a kognitív folyamatai és a pillanatnyi nyelvtudása segítségével milyen hatékonyan tud kommunikálni (Tavakoli, 2020:6).

Tavakoli (2020) szerint a fluensség fogalmának meghatározását több tényező is nehezíti. Egyrészt az, hogy a különböző tudományágak, a fent említett megközelítések nem mindig állnak kapcsolatban egymással. Megjegyzi, hogy a beszédfolyékonyságot eddig főleg kognitív nyelvészeti szempontból vizsgálták, és nem igazán vették figyelembe a pszicholingvisztikai kutatások eredményeit, valamint a szociológiai és pragmatikai jellemzőket sem. Ugyanakkor hasonlóan ehhez nincs még szoros kapcsolat például a diskurzuselemzés és a másodiknyelv-elsajátítás [Second Language Acquisition] vizsgálata között sem (Tavakoli, 2020: 17). Tavakoli véleménye szerint az is nehezíti a fluensség fogalmának meghatározását,

hogy a tudományos kutatások és a gyakorlat, tehát például a beszédfolyékonyság fejlesztése vagy mérése és értékelése sem kötődik szorosan egymáshoz (2020: 17–18).

A beszédfolyékonyság kutatásának területén készült részletes, áttekintő munkájában Koponen és Rigggenback (2000: 5–24) arról számol be, hogy szinte lehetetlen egyetlen definíciót találni a fluensség fogalmára. A *Perspectives on Fluency* című könyv az egyik első olyan mű, amely rendszerbe foglalta a második nyelvre vonatkozó beszédfolyékonysággal kapcsolatos addig megjelenő kutatásokat, valamint azokat a nézőpontokat, amelyeket azok alapjaiként szolgáltak. Az első nézőpont szerint a fluensség mérhető időbeli változók halmaza, amely magában foglalja a beszédtempót, a megszakítás nélküli beszéd időtartamát, valamint a szünetek helyét és hosszát. Ilyen Ellis és Barkhuizen meghatározása, akik úgy definiálják a folyékonyságot, mint „a nyelv valós idejű előállítás indokolatlan szünet vagy habozás nélkül” (2005: 139). Habár Segalowitz (2010: 39) azt állítja, hogy „nem lehetséges globálisan folyékonynak jellemezni egy nyelvtanuló második nyelven történő megnyilatkozásait valamilyen egydimenziós, abszolút módon”, a nyelvészek számos időbeli kritériumot használnak ennek mérésére (4. táblázat). Ilyen a beszédsebesség, amelyet általában a percenként elhangzó szótagok vagy szavak számában mérnek. Emellett a szünetek gyakorisága, a szünetek helye és hossza, vagy a szünet nélkül elhangzó szótagok átlagos száma (Thornbury, 2022: 222) szokott a mérések fókuszában lenni.

**4. táblázat** A beszédfolyékonyság mérhető egységei Kormos (2006: 163) alapján

| Mérték, egység                         | Meghatározás  |
|--|---|
| Beszédarány (szótag / perc)            | Az egy perc alatt elhangzó összes szótag összege és az összes szünet összege elosztva a teljes beszédidővel.                |
| Artikulációs arány (szótag / perc)     | Az egy perc alatt elhangzott összes szótag összege elosztva a teljes beszédidővel (szünetek nélkül).                        |
| Fonáció-idő arány (százalékos arány)   | A tényleges beszédidő és a beszéd teljes idejének százalékos aránya.  |
| Egy egység átlagos hossza (szótagszám) | Átlagos szótagszám a szünetek között.   |
| Csendes szünetek/ perc                 | Az egy perc alatti összes szünetszám elosztva a teljes beszédidővel.  |
| A szünetek hossza (másodperc)          | Az összes szünet hossza.  |
| Kitöltött szünetek / perc              | Az egy perc alatt összes olyan szünet, amely nem csendes (pl. töltelékszak, hümmögés, őzés) elosztva a teljes beszédidővel. |
| Diszfluenciák / perc                   | Az egy perc alatti összes diszfluencia (ismétlések, újrakezdések, javítások) elosztva a teljes beszédidővel.                |



|                            |  |
|----------------------------|--|
| Ütem, sebesség             | A hangsúlyos szavak száma percenként.        |
| Hangsúlyos szó: összes szó | A hangsúlyos szavak és az összes szó aránya. |

Más kutatók a folyékonyt a beszéd mögöttes folyamatainak automatizálásaként értelmezték. Egy harmadik nézet szerint pedig a folyékonyt a magas szintű nyelvtudás összetevőjeként tekintjük. Az adott beszédmintát annak hallgatói értékelnek a beszédprodukciónak elhangzásának pillanatában. Az utolsó nézőpont pedig a pedagógiai, amely a folyékony fogalmát szembe állítja a pontos, grammatikailag helyes nyelvhasználattal (Kirk, 2014: 102).

A Koponen és Riggenback (2000: 5–24) által összefoglalt négy megközelítés közül az első és talán a legtöbbet kutatott terület, a pedagógiai. A *Teaching Oral English* című könyvében Byrne, aki az egyik első meghatározását alkotta meg a fogalomnak azt írja, hogy a beszédfolyékonyt a „képesség arra, hogy valaki érthetően, pontosan és hezitálás nélkül fejezze ki magát” (Byrne, 1976: 9). Byrne-t követve Fillmore (1979) is megpróbálta definiálni a fogalmat. Szerinte a folyékonyt a nyelvi produkcióban jelenik meg, és minden esetben a kompetens nyelvhasználókhoz viszonyítva. Sokrétű meghatározásában, amelyet az első nyelven történő megnyilatkozásokra hozott létre, többek között megjelenik a szociolingvisztikai tudás alkalmazás és a nyelvi kreativitás is, amely segítségével hosszán, koherensen és szemantikailag gazdagon tudjuk kifejezni gondolatainkat. Négyféle képességet határozott meg, amely hozzájárul a folyékony nyelvhasználathoz, amelyek a következők:

1. A képesség arra, hogy meghatározó szünetek nélkül kitöltsünk egy hosszabb időt beszéddel. Tehát a beszéd mennyisége nagyobb nyomatékot kap, mint a beszéd minősége, hiszen az időkitöltésére helyezi a hangsúlyt (Kirk, 2014:101).
2. Koherens, értelmes, szemantikailag gazdag mondatok produkciójának képessége. Itt a beszéd komplexitására és minőségére helyezi a hangsúlyt (Kirk, 2014:101).
3. Egy széles skálájú kontextusban történő megfelelő beszédprodukciónak képessége. Tehát nem maradunk szótlannak, amikor egy idegen emberrel találkozunk, vagy egy váratlan szituációban találjuk magunkat.
4. Kreatív és képzeletgazdag nyelvhasználat képessége. Például viccek, anekdoták mesélése.<sup>9</sup>

---

<sup>9</sup> Az eredeti szöveg:

„the ability to talk at length with few pauses; the ability to fill time with talk; the ability to talk in coherent and semantically dense sentences; the ability to have appropriate things to say in a wide range of contexts; and the ability to be creative and imaginative in the language use” (Fillmore, 1979: 51).

A folyékonyan beszélő emberrel találkozáskor „az a benyomásunk támad [...], hogy nagyon gyorsan előre megtervezi a helyzetre adott válaszok alternatív módjait, és kiválasztja közülük azt, amelyik a lehangzatosabbnak vagy legintelligensebbnek tűnik számára” (Fillmore, 1979: 93). Ezzel szemben a nem folyékonyan beszélő személynek a beszéde akadozó, lassú, beszéd közben pedig időnként meg-megáll.

Az ehhez hasonló, mindenre kiterjedő, vagyis holisztikus meghatározások alapján készültek a második nyelvre vonatkozó definíciók (Tavakoli 2020: 4–5). Ilyen például Brumfité, aki meghatározásának fókuszába a természetes nyelvhasználatot helyezi. A következőképpen definiálja a fleunsség jelentését: „természetes nyelvhasználat, attól függetlenül, hogy anyanyelvi beszélőhöz hasonlító megértést és produkciót eredményez-e” (Brumfit, 1984: 56). A természetes nyelvhasználatot egy olyan tevékenység meghatározásán keresztül mutatja be, amely az addigi hagyományos osztálytermi tevékenységekkel szemben a kompetens nyelvhasználó mindennapi szituációira hasonlít. Az ilyen tevékenységnél elengedhetetlen, hogy a beszélő birtokolja a nyelvet, hogy a hallgató ne egy közvetítő segítségével értse meg az üzenetet, és hogy a beszélgetés tartalmát a résztvevők határozzák meg. Lehet benne improvizáció, parafrázis, önmaguk javítása is csak úgy, mint egy hétköznapi beszélgetésben (Brumfit, 1984: 56–57).

Az eddig említett definíciók elsősorban a megnyilatkozások, tehát a beszélő, nyelvtanuló nyelvprodukciónak gyorsaságát, minőségét veszi figyelembe. A későbbi meghatározásokban már jobban fókuszba helyeződik az üzenet befogadójának, a kompetens nyelvhasználónak a szempontja is (Ejzenberg, 2000; Hasselgreen, 2005 idézi Kirk, 2014: 102). Az ezen a területen legtöbbször idézett meghatározás Lennoné, aki a korábbi, általában tágan értelmezett, holisztikus definíciót két részre osztotta. Szerinte az egyik értelmezése a beszédfolyékonyságnak egy általános meghatározása a szóbeli, idegen nyelvi jártasság, amelyet általában a tanuló beszédének egy kompetens nyelvhasználó általi értékeléséből ismerhetünk meg. A másik meghatározás pedig egy már jóval mérhetőbb és objektívebb szempontokat tartalmazó definíció (Tavakoli, 2020: 6). Lennon szerint a folyékony nyelvhasználat „a gondolatok vagy kommunikációs szándékok gyors, zökkenőmentes, pontos, világos és hatékony fordítása célnyelvre, bizonyos időbeli korlátok mellett” (Lennon, 2000: 26). A beszédfolyékonyságot tehát abból a szempontból nézi, hogy a beszélő a kognitív folyamatai és a pillanatnyi nyelvtudása segítségével milyen hatékonyan tud kommunikálni (Tavakoli, 2020: 6). Ez a gondolat a beszédfolyékonyság mai megközelítéseinél nagy jelentőséggel bír.

## 5.2 A beszédfolyékonyság kutatásának mai megközelítései

### 5.2.1 A Segalowitz-féle hármas modell

A másodiknyelv-elsajátítással kognitív keretben foglalkozó kutatók a beszédfolyékonyságot a nyelvi produkció során történő automatizmusokkal teszik egyenlővé. Az automatikus folyamatok kifejlődését a nyelvtanulók ismétlés által érik el, amely a begyakorolt nyelvi szerkezetek minőségi javulásához is vezet, valamint lehetővé teszi a beszélők számára, hogy azt az erőforrást, amelyet a szerkezetek létrehozására fordítanak az interakció más, magasabb rendű jellemzőihez rendeljék (vö. II.4 fejezet). Ilyen lehet az odaillő, megfelelő kifejezések kiválasztása az adott szituációban (Thornbury, 2022: 222). Kirk (2014) még azt is megjegyzi, hogy konnekcionalista elméletek szerint az automatizálást fejleszteni rendszeres ismétlés által lehet, amely azzal magyarázható, hogy az ismételt találkozások által erősítjük az egyes emlékekhez fűződő idegi kapcsolatokat. Ha nincs elég ismétlés, az idegi kapcsolatok elgyengülnek, amely felejtéshez vezet (Kirk, 2014: 105). Hasonlókról ír Hoey (2005) is a lexikai előfeszítés elméletéről szóló könyvében, amely az II.3 fejezetben került bemutatásra. Segalowitz 2010-es könyvében a kognitív keret mellett, szociális perspektívával vizsgálja a második nyelven történő beszédfolyékonyságot, amelynek három típusát határozza meg. Az első típus a kognitív beszédfolyékonyság [cognitive fluency], amelyet a beszédprodukció háttérben zajló folyamatok hatékonysága határoz meg, tehát a beszédfolyamaton belüli automatizálás foka. Ez fontos szerepet játszik a második, a megnyilatkozáshoz köthető fluensségben [utterance fluency], amely a megnyilatkozások mérhető változóival kapcsolatos, tehát például a beszédtempóval és a szünetekkel. Skehan (2003) szerint ennek a fluensségnek a mérése során az eredményeinek a validitásához elengedhetetlen a csendes szakaszoknak [breakdown fluency], a cseréknek, javításoknak, újrafogalmazásoknak [repair fluency], a beszédsebességnek [speech rate], vagyis a szavak, szótagok percenkénti számának, valamint az automatizációnak [automatisation] analizálása, amelyet a szünetnélküli egységek hosszának mérésével elemezhetünk (Skehan, 2003: 8). Visszatérve a Segalowitz által meghatározott fluensségfajtákhoz, a harmadik típus az észlelt fluensség [perceived fluency], amely a hallgató benyomása a beszélő beszédprodukciójának hatékonyságáról. Ezt a hármas modellt számos kutatási területen használják, például a másodiknyelv-elsajátítás kutatásában, illetve más kognitív tudományterületeken. A hármas modell azért is fontos, mert nemcsak a beszélőre fókuszál, hanem Lenonhoz hasonlóan a hallgató szerepét is hangsúlyozza. Segalowitz kiemelt néhány kutatást (vö. Kormos – Johnson, 2016), amelyből kiderül, hogy a beszédfolyékonyság

érzékelése nagy mértékben korrelált a nyelvi jártasság átfogó objektív értékelésével, valamint a beszélő saját önértékelésével is (Tavakoli, 2020: 26–27).

**5. táblázat** A Segalowitz-féle hármas modell és annak meghatározó tényezői

| <b>A beszédflowékonyág típusa:</b>                      | <b>Meghatározza, befolyásolja:</b>       |
|---|--|
| <b>1. Kognitív beszédflowékonyág</b>                    | az automatizálás foka                    |
| <b>2. Megnyilatkozásokhoz köthető beszédflowékonyág</b> | a csendes szakaszok                      |
|   | a cserék, javítások, újrafogalmazások    |
|   | a beszédsebesség                         |
|   | az automatizáció                         |
| <b>3. Észlelt beszédflowékonyág</b>                     | a helyes nyelvhasználat                  |
|   | az idiomatikusság                        |
|   | az akcentus                              |
|   | az interakciós készségek                 |
|   | a stratégiai kompetencia                 |
|   | a paralingvisztika (gesztusok, tekintet) |

Thornbury (2022) állítása szerint több tényező befolyásolhatja Segalowitz harmadik típusú beszédflowékonyágát, vagyis az észlelt fluensséget (ld. 5. táblázat), amelyek a következők:

1. Helyes nyelvhasználat [accuracy]: létezik egy olyan határ, az úgynevezett túrés küszöb, amely alatt a hallgató a beszélő által előállított nem szabványos formák számát és gyakoriságát nem veszi figyelembe abban az esetben, ha az üzenetet megérti. Chambers-t (1997: 540) idézve Thornbury megjegyzi, hogy „a flowékonyág megítélése valójában magában foglalja a nyelvi helyességet valamilyen módon”, ami azt jelenti, hogy a Chambers általa „sima és folyamatos beszédnek” definiált fluensség csak egy bizonyos határig alkalmazható.
2. Idiomatikusság [idiomacity]: vagyis a szokatlan és nem tipikus kifejezések használatával ellentétben, a beszélő nyelvhasználat a milyen mértékben tükrözi az anyanyelvűekhez hasonló nyelvhasználatot.
3. Akcentus [accent]: azaz a beszélő akcentusa mennyire befolyásolja a megértést.
4. Interakciós készségek [interaction skills]: a beszélő képessége a többi beszélővel való együttműködésre. Hogyan tudnak bekapcsolódni egy beszélgetésbe, hogyan adják át a

szót, de ide tartozik a pragmatikus fluensségnek nevezett tényező is, amely egy olyan „párbeszédre jellemző jelenség, amely egyesíti a megnyilatkozások pragmatikus helyénvalóságát és a zökkenőmentes folytonosságot a folyamatban lévő beszélgetésben” (House, 1996: 228 idézi Thornbury, 2022: 223).

5. Stratégiai kompetencia [strategic competence]: vagyis a kommunikációs stratégiák alkalmazásának képessége, a második nyelv hiányos ismeretéből adódó félreértések elkerülésének módjai, elkerülő stratégiák, vagy a kommunikációs hibák javításának képessége.
6. Paralingvisztika [paralinguistics]: azaz a gesztusok, a tekintet és a proxémia, vagyis a beszélők közötti távolság megfelelő alkalmazása (Thornbury, 2022: 223).

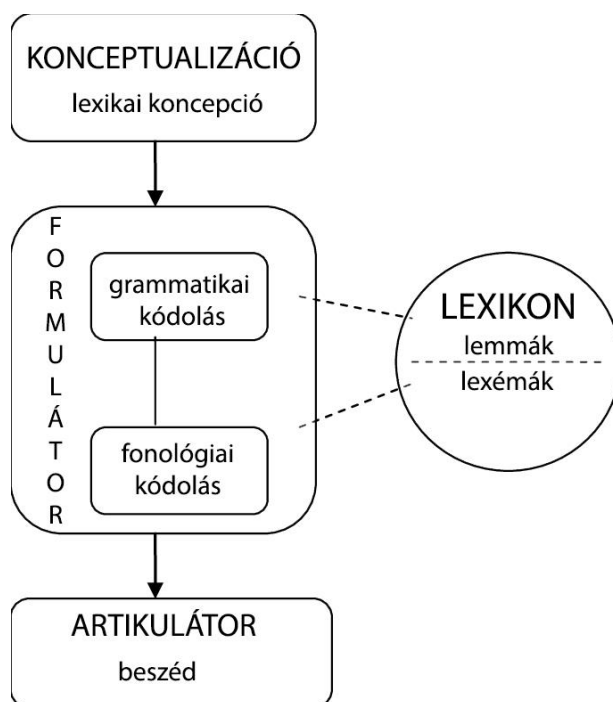
Segalowitz hármas modelljével kapcsolatban még azt is lényeges megjegyeznünk, hogy a beszédhelyesség három típusa mellett figyelembe veszi a vizsgálati alanyul szolgáló nyelvtanulók társadalmi tapasztalatait és motivációit is. Ezzel kapcsolatban azt is kijelenti, hogy különböző beszélők, különböző feladatokra, korábbi tapasztalataikra támaszkodva és motivációjuktól függően máshogy reagálnak. Az sem egyforma az idegen nyelven történő kommunikáció közben, hogy ki milyen stratégiát választ a hiányosságai miatti nehézségek leküzdésére. Valamint azt is megjegyzi, hogy a beszélőknek saját beszédmintázatuk van az első nyelvükön, amely befolyásolja a második nyelven történő kommunikáció közbeni beszédsebességet, a szünetek tartását és a javítási szokásaikat is (Tavakoli, 2020: 27). Ezeket az általa felsorolt tényezőket érdemes figyelembe venni a beszédhelyesség kutatásánál.

### **5.2.2 Beszédproduktions modellek**

A Segalowitz-féle első beszédhelyességi típus, a kognitív beszédhelyesség jobb megértéséhez szükséges, hogy megismerjük a legfontosabb beszédproduktions modelleket. Az első, Levelt (1989, 1999) által megalkotott modell három fő részből áll:

1. konceptualizáló rész, a jelentéstartalom meghatározása [conceptualisation],
2. formulátor rész, szabályhoz rendelés [grammatical, lexical formulation],
3. artikulátor rész [articulation].

2. ábra Levelt beszédprodukción-modellje Levelt (1989: 9) alapján

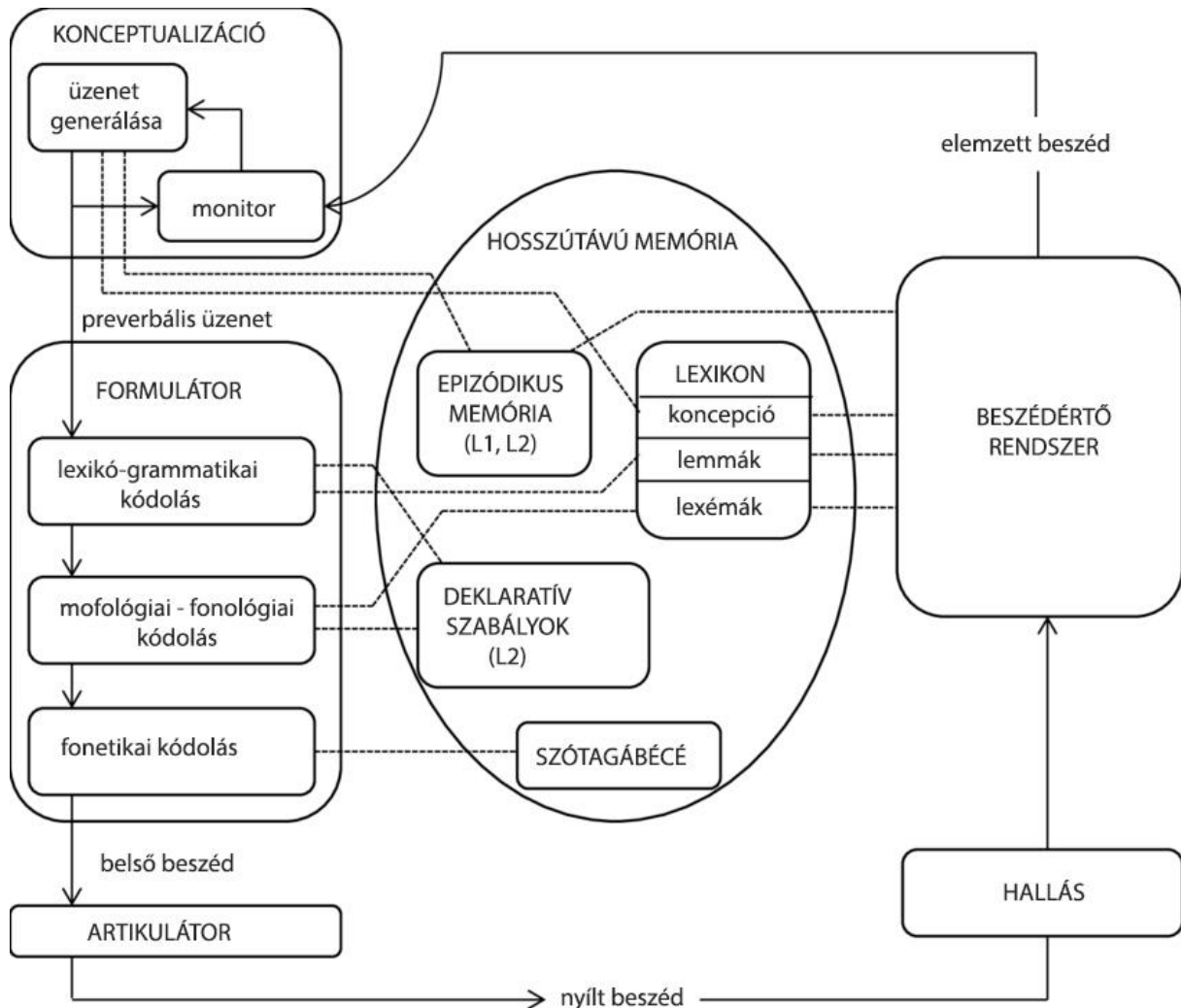


Levelt modelljében a beszédprodukción egy egyirányú folyamat, amely három különálló, automatikusan végbemenő lépésben történik. A konceptualizáló rész a kommunikációs szándékot preverbális üzenetté alakítja át. Ez a makro tervezési szakasz még a nyelvtől független rész (Tavakoli, 2020: 21). Tartalmazza az összes szükséges információt, amely a formulátor résznek (mikro tervezési szakasz) elengedhetetlen a szükséges lexikai elemek kiválasztásához és a nyelvtani szabályok alkalmazásához. Egy felületi struktúrát vagy belső beszédet hoz létre. Ezt használhatja fel az artikulátor a tényleges beszéd előállításához (Kirk, 2014:103). A három részben lezajló eljárás a kompetens nyelvhasználóknál gyorsan, parallel történik és nem igényel nagy kognitív erőfeszítést. Abban a pillanatban, ahogy egy információ egység elhagyja a konceptualizálás lépését és tovább halad a szabály rendeléshez, az előző lépés máris elkezd dolgozni a következő jelentéstartalom meghatározásán. Ha ez a folyamat nem automatikusan történik, például egy nyelvtanuló esetében, akkor a három lépés észrevehetően lassabban és sok befektetett kognitív munkával megy végbe (Tavakoli, 2020: 10–11). A modellben található egy folytonos önellenőrző rész [monitor loop] is, így lehetséges, hogy a megnyilatkozó a tényleges beszéd előtt megváltoztassa a tervét. A pontatlanságok felismerése azonban diszfluenciát, önmagunk javítását és a megnyilatkozás újakezdését eredményezheti (Kirk, 2014:103).

Ennek a modellnek több hiányossága is van: nem jelenik meg benne a diskurzus, csak az egyénre koncentrál, de nem veszi figyelembe a kompetens nyelvhasználók beszédének

különbségeit sem, valamint nem jelenik meg benne a második nyelven történő beszédprodukciónak leírása sem (Tavakoli, 2020: 11). Levelt eredeti modelljét Bot (1992) dolgozta át először a második nyelven történő megnyilatkozásokra, majd ezt Bock és Levelt (1994) kiegészítette, amelyet később Kormos (2006) dolgozott át még részletesebben. Ma már erre az utóbbira hivatkoznak a legtöbbit.

**3. ábra** Kormos beszédprodukciónak modellje Kormos (2006: 168) alapján



Részletes kétnyelvű beszédprodukciónak modelljében Kormos (3. ábra) megjeleníti a hosszútávú és epizodikus memóriatípusokat is, valamint a grammatikai szabályok tudásának is külön részt szán (Kirk, 2014:103). Kormos modelljén keresztül bemutatja a második nyelven történő beszéd folyékony fejlődésének folyamatát, hogy a nyelvtanulók hogyan próbálják használni a fejlődésben lévő grammatikai, lexikai és fonológiai ismereteiket a sokkal könnyebben elérhető első nyelvi ismereteikkel szemben. Kormos szerint a tudásuk nagyrésze expliciten tanított és deklaratív, nem pedig implicit procedurális tudás. Ehhez a tudáshoz még

nincs automatikus hozzáférésük, amely meglassítja és még több kognitív figyelmet igénylővé teszi a folyamatot. Diszfluencia így minden szakaszban megjelenhet, de főként a formulátor részénél, ahol a lexikai-grammatikai jelenségekhez másfajta nyelvi feldolgozásra van szükség (Tavakoli, 2020: 11–14). Ebben a modellben a deklaratív grammatikai tudás nem ugyanaz, mint a procedurális. A kognitív beszédhelykonyság nem a nyelvtani szabályok felgyorsított használatából fakad, hanem inkább a beszédprodukción automatizálásából (Hulstijn, 2007 idézi Kirk, 2014:103).

### 5.2.3 Automatizálás

A beszédhelykonysághoz hasonlóan az automatizálásnak sincs egy mindenki által elfogadott meghatározása. Az ismert definíciók főleg a különböző képességfejlesztést, a memóriát, a figyelmet, valamint az információfeldolgozást vizsgáló kutatásokból származnak (Tavakoli, 2020: 24). Schmidt szerint az automatikus beszédhelykonyság ismertetőjegyei a következők:

1. gyors és hatékony,
2. könnyed,
3. nem korlátozza a rövid távú memória kapacitását,
4. nem áll önkéntes ellenőrzés alatt,
5. nehezen módosítható vagy gátolható,
6. nem érhető el az önvizsgálat számára (Schmidt, 1992: 360).

Dörnyei (2009) összefoglalja az évek során született definíciókat, amelyek közül léteznek szűkebb és tágabb értelműek is. A tág értelmű meghatározások általában az automatikus viselkedés három aspektusának egyikére koncentrálnak: Logan (2005) szerint az automatizálás a reakcióidő felgyorsítása. Lee (2004) azt állítja, hogy az automatizálás a procedurális memória megfelelő használatának elsajátítása. A legtöbb meghatározásban megjelenik a memória, azon belül pedig a hosszútávú memóriának két típusa. Az első típusú a már említett procedurális, begyakorolt tudás, amely a legtöbb rutin alapja, nem igényel tudatos figyelmet. A másik típus a deklaratív, memorizált tudás, amely tudatos és explicit módon tárolt információ. Az ilyen ismeretek előhívásához több figyelemre van szükség (Tavakoli, 2020: 24). Végül a harmadik aspektus, amelyre az automatizálás meghatározásai összpontosítanak a szándékos kontrol hiánya egy kognitív tevékenység végrehajtása során (Segalowitz – Hulstijn, 2005; Dörnyei, 2009: 270–274).

Összefoglalva a különböző definíciókat elmondhatjuk, hogy az automatizálás fogalmának



meghatározásai között szerepel kritériumként a megállíthatatlanság [ballistic], hiszen automatikusan történik a folyamat, a nem tudatos [unconscious] jelző, vagyis figyelmet nem igénylő folyamatról van szó, valamint a stabilitás [stable], tehát az interferenciával szembeni ellenállás is (Dörnyei, 2009: 270–274).

**6. táblázat** Az automatizálás értelmezéseinek összefoglalása Dörnyei (2009: 270–274) alapján

|                                    |   |
|------------------------------------|---|
| <b>Kritériumok:</b>                | megállíthatatlan, nem tudatos, stabil   |
| <b>Tág értelmű meghatározások:</b> | (1) a reakcióidő felgyorsítása<br>(2) a procedurális memória elsajátítása<br>(3) a szándékos kontrol hiánya egy kognitív tevékenység végrehajtása során |

Az automatikus folyamatok fejlesztésével kapcsolatban Hulstijn (2002) azt állítja, hogy csak a beszéd tempójának gyorsítása nem elegendő hozzá. Szerinte a valódi automatizálásnál a beszédfeldolgozásának nemcsak mennyiségbeli, hanem minőségbeli javulására is szükség van (idézi Dörnyei, 2009: 271–272). A második nyelven történő automatizálás vizsgálatának meghatározó alakja, a már fentebb említett Segalowitz, a beszédfolyékonyságra megalkotott hármas modelljének kidolgozása közben az automatizálás két aspektusával is foglalkozott (2010). Az egyik a feldolgozás sebessége, a másik pedig a feldolgozás közbeni stabilitás kérdése. A vizsgálataihoz grammatikai információk feldolgozásával foglalkozó kutatások eredményeit használta. Mindkét esetben kiemelte, nem gondolja, hogy reprezentatív eredményeket kapott, de a kutatásokból az látszott, hogy ismétléssel és gyakorlással a nyelvtanulók kognitív terhelése csökkent, így egyre jobb eredmények születtek. Ugyanakkor Hulstijnhoz hasonlóan szerinte sem szabad az automatizálás értelmezését ennyire, csupán a gyakorlásra leszűkíteni, hiszen a megnyilatkozásszintű gyakorlás nem feltétlenül támogatja a mögöttes kognitív folyamatok minőségi javulását (Tavakoli, 2020: 25–26).

### 5.3 A beszédfolyékonyság kurrens meghatározása

Az előző részekből (II.5.1–II.5.3 fejezet) láthattuk, hogyan alakult a beszédfolyékonyság meghatározásának változása az évek során, valamint, hogy milyen aspektusokat tartottak fontosnak a fluensség kérdésével foglalkozó kutatók. Tavakoli (2020) két okból gondolja pontatlannak a korábbi meghatározásokat. Egyrészt azért, mert a beszédfolyékonyságnak a gyorsaság, illetve automatikusság alapú meghatározása olyan elemeket hagy figyelmen kívül, mint a kultúra vagy a kontextus, amelyben a beszélgetés zajlik, ezek pedig kulcsfontosságú

tényezők. Másrészt pedig azt állítja, hogy a korábbi modelleknél még részletesebbre lenne szükség ennek a komplex fogalomnak a meghatározásához. A Lennon-féle kétrészes meghatározással és a Segalowitz-féle hármassal teljes mértékben egyetért, ugyanakkor szerinte ezek még mindig kevésnek bizonyulnak a beszédhelyesség értelmezéséhez. Emiatt ő és Hunter 2018-ban egy négyzetes piramis modellt állított fel a fogalom meghatározására:

4. ábra Tavakoli és Hunter piramis modellje Tavakoli (2020: 7-8) alapján



A négyzetes piramismodell szerint a beszédhelyesség:

- 1) nagyon szűk értelemben: beszédsebesség, szünet, javítás
- 2) szűk értelemben: könnyedség, áramlás és folytonosság; megkülönböztetve a grammatikailag helyes nyelvhasználatától és a komplexitástól
- 3) tág értelemben: egy második nyelven történő beszéd képessége
- 4) nagyon tág értelemben: általános nyelvhasználati jártasság (Tavakoli, 2020: 7–8).

A fluensség meghatározásának négyféle megközelítése közül a nagyon tág értelemben történő meghatározás az első, amely a fluensséget mint általánosságban vett nyelvi jártasságot tekinti, amelybe nemcsak a beszédhelyesség, de mind a négy készség szerepet játszik. A második meghatározási szint szerint, amely már egy kicsit szűkebb értelmű, de még mindig tágan mondható, a beszédhelyesség a kompetens beszédhelyességgel egyenlő. A kifejezés jelentése olyan nyelvi beszélőre utal, aki magabiztosan kommunikál a második nyelven, hozzáértően és

koherensen adja át a saját maga által meghatározott üzenetet. Ebbe a meghatározásba Tavakoli és Hunter a kiejtést, valamint az érthetőséget is belefoglalja. A harmadik szinten a könnyedség és a folytonosság mellett a grammatikai komplexitás és a helyesség is megjelenik. A negyedik szinten pedig a mérhető tényezők jelennek meg, mint a sebesség vagy a szünetek tartása. Tavakoli megjegyzi, hogy ezzel a modellel megérthetjük az elmélet és a tanítási gyakorlat közötti különbséget is, hiszen míg a nyelvészek általában a legszűkebb, mérhető értelmezést használják kutatásaik során, addig a nyelvtanárok munkájuk közben egy szélesebb meghatározásban gondolkodnak (Tavakoli, 2020: 7).

#### **5.4 A beszédfolyékonyság a nyelvoktatásban**

A beszédfolyékonyságot [fluency] gyakran állították párhuzamba vagy szembe a grammatikailag helyes nyelvhasználat [accuracy] kérdésével a nyelvpedagógiai kutatásokban (vö. Sampson, 1977, Allington, 1983; Eskey, 1983). Az fluensség fogalmának egyik első meghatározásában (Byrne, 1976) a grammatikailag helyes nyelvhasználat még a beszédfolyékonyság kritériumai között szerepelt, a kommunikatív nyelvoktatás [Communicative Language Teaching] elterjedésével – amely az 1970-es évek második felére, végére tehető – azonban sokáig megfigyelhető egy tendencia, amely során a helyes és a fluens nyelvhasználatot egymáshoz képest próbálják meg definiálni. Brumfit a fluensség második nyelvre vonatkozó definíciójának megfogalmazása során ugyan hangsúlyozza, hogy a két fogalom elkülönítése nem azt jelenti, hogy a fluens nyelvhasználat nem lehet grammatikailag helyes is egyben, illetve, hogy a megkülönböztetés inkább metodológiai, mintsem nyelvészeti vagy pszichológiai (1984: 52). Ahogy korábban említésre került, Brumfit (1984) a folyékonyságot úgy határozta meg, mint a nyelv természetes használatának képességét, függetlenül attól, hogy a nyelv megértése és produkciója hasonló-e az anyanyelvi beszélőkéhez. Úgy véli, hogy a természetes nyelvhasználatot olyan tevékenységek végzésének kell tekinteni, amelyeket egy kompetens nyelvhasználó a mindennapi életben az anyanyelvén végezne, nem pedig hagyományos tantermi feladatnak. Ez azt jelenti, hogy a beszélőnek birtokolnia kell a nyelvet, a hallgatónak nincs szüksége közvetítőre az üzenet megértéséhez, és a beszélgetést mindkét résztvevőnek meg kell határoznia. Lehet benne improvizáció, parafrázis, önmaguk javítása csak úgy, mint egy hétköznapi beszélgetésben. A tevékenység célja ne a grammatikailag helyes nyelvhasználat legyen, a tanár pedig ne szóljon közbe, ne javítson, hiszen a „hibázás megkerülhetetlen része a második nyelv elsajátításának a folyamatában” idézi Brumfit Cordert (1975, idézi Brumfit, 1984: 56-57). Ez a meghatározás, amely a folyékony

nyelvhasználatot a kommunikatív kompetenciához hozzájáruló elemként tekinti, ma is több tanár és kutató nézetének központi eleme (Tavakoli, 2020: 5).

Ahogy Tavakoli (2020: 17–18) fogalmaz, a kutatások és a gyakorlat (hogyan fejlesszük, mérjük és értékeljük a fluensséget) gyakran nem állnak kapcsolatban egymással. A nyelvtanári gyakorlatban sokszor használják a beszédfolyékonyság fogalmát megfelelő elméleti megalapozás nélkül (Kirk, 2014:101). Habár ma már tudjuk, hogy a folyékonyság és a grammatikailag helyes nyelvhasználat nem különíthető el élesen egymástól (Thornbury, 2022: 220), még mindig sok nyelvkönyvben található meg ez a felosztás. Tanárjelöltként pedig hagyományosan úgy tanuljuk az egyetemi tanulmányaink során, hogy az úgynevezett kötött vagy irányított gyakorlatok [controlled practice] inkább a helyes nyelvhasználatot, míg a szabad gyakorlatok [free practice] a beszédfolyékonyságot fejlesztik. Természetesen ennél a kettősségnél sokkal komplexebb folyamatról van szó.

#### **5.4.1 A beszédfolyékonyság kapcsolata az első és második nyelvvel**

A beszédfolyékonyság meghatározásait bemutató részből (II.5.1–II.5.3 fejezet) láthattuk, hogy a fluensséget vizsgálták és vizsgálják az első és második nyelvvel kapcsolatban is. Legtöbbször a mérhető, időbeli tényezőket elemzik. Ilyenek például a korábban felsorolt beszédidő, a szünetek hossza vagy a beszéd és a szünetek aránya. Ha az adott kutatás a második nyelvre vonatkozik, azt vizsgálják, hogy mennyire tudja a nyelvtanuló átvenni és alkalmazni a kompetens nyelvhasználók automatizálás által megkönnyített beszédmódját. Emiatt a nézőpont miatt a második nyelven történő beszédfolyékonyság olyanná válik, mintha a nyelvtanuló első nyelvétől teljesen elkülönített jelenség lenne (Tavakoli, 2020: 22). Azonban meg kell jegyezni, hogy a második nyelven történő beszédfolyékonyságnak van kapcsolata az első nyelvhez. A nyelvtanuló beszédprodukciónak diszfluenciája gyakran nem a hiányos ismereteit mutatja, hanem a saját első nyelvének beszédmintázatát vagy valamiféle belső beszédstratégiát, amelyet a kommunikációs céljának megvalósulása érdekében alkalmaz (Tavakoli 2020: 33). Tavakoli Raupach 1980-as vizsgálatát idézi, aki német és francia első nyelvű diákok beszédprodukciónak elemezte és arra az eredményre jutott, hogy a második nyelven megjelenő diszfluencia az első nyelven is megtalálható, valamint hasonlóak a szünetek tartásának mintázatai is. Tehát a diákok első nyelvéhez köthető fluenssége nagy mértékben befolyásolja a második nyelven történő beszédprodukciónak. Tavakoli szerint a beszédfolyékonyság egyfajta személyiségjegye (Tavakoli, 2020:29). Raupach kutatásán kívül más vizsgálatokat is megemlít, amelyek ezt az elméletet megerősítik:

- Clark elmélete szerint (Clark – Fox – Tree, 2002) a szüneteket kitöltő elemek, a töltelékszavak, hümmögés vagy őzés konzisztensen ugyanúgy megjelennek az első és második nyelven is.
- Derwing et al. (2009) longitudinálisan is kimutatta, hogy a második nyelven történő beszédhelyettesítés az első nyelvre támaszkodik.
- De Jong et al. (2015) eredményei szerint pedig a beszédsebesség (a szótagok időtartama), a korábban Skehan (2003) által említett csendes szakaszok [breakdown fluency], amely a néma szünetek száma másodpercenkénti beszédidőben, a kitöltött szünetek száma másodpercenkénti beszédidőben, valamint a néma szünetek átlagos hossza, illetve a javítási fluencia [repair fluency] vagyis a javítások és ismétlések, újratekadések száma is hasonló volt mind a két nyelven (Tavakoli, 2020: 30).

A fenti kutatások eredményeiből arra a következtetésre juthatunk, hogy az első és a második nyelven történő beszédhelyettesítés átfedésben van egymással. A nyelvtanulók első nyelvi beszédmintázatai, valamint szokásai erősen befolyásolják a második nyelven történő beszédprodukciónkat.

A mérhető tényezőket vizsgáló pszicholingvisztikai kutatások mellett a leggyakrabban a második nyelv elsajátításával kapcsolatban vizsgálják a beszédhelyettesítés jelenségét. A kutatások során a nyelvi jártasságot egy négy összetevőből álló keretrendszer működésének különböző feladatok elvégzése, valamint ismétlése során történő vizsgálatával szokták elemezni. A keretrendszer egyik összetevője a beszédhelyettesítés, mely a szintaktikai és lexikai komplexitás, a grammatikailag helyes nyelvhasználat és az egyre nagyobb hangsúlyt kapó lexika mellett jelenik meg (vö. CALF keretrendszer – Complexity, Accuracy, Lexis, Fluency). Ugyanakkor elmondható az is, hogy mind a komplexitás, mind a grammatikailag helyes nyelvhasználat és a lexika is hozzájárul a beszédhelyettesítéshez (Tavakoli, 2020: 14–16). A legtöbb vizsgálat a komplexitást és a helyes nyelvhasználatot tartja meghatározónak, de a fent felsorolt kutatásokra hivatkozva Tavakoli (2020) azt állítja, hogy az első nyelv befolyása és a személyes beszédmintázatok miatt nem olyan nagy mértékben döntő tényezők. Fontosabb ennél, hogy a nyelvtanulók hogyan érik el a mentális lexikonjukban tárolt információkat valós időben történő kommunikáció során, valamint a lexikai egységek használata, illetve az, hogy a felsoroltakat hogyan lehet fejleszteni (Tavakoli 2020: 31). A beszédhelyettesítés minden bizonnyal kapcsolatban áll a formulaszerű egységek ismeretével és aktív, megfelelő helyű és gyakoriságú használatával (Nattinger – DeCarrico, 1992: 32). A valós idejű beszédfeldolgozással járó kognitív igénybevétel miatt a beszélgetéseknek ugyanis, ahogy

korábban már említettük „korlátozott és ismétlődő repertoárja” (Biber et al., 1999: 1049) van és erősen támaszkodik a formulaszerű, előre felépített lexikai egységekre (vö. Wray, 2002; Wray – Fitzpatrick, 2008; Wray, 2019).

#### 5.4.2 A beszéd folyékonyág értékelése a nyelvvizsgán

A nyelvtanári gyakorlatban a beszéd folyékonyág a nyelvvizsgáztatáskor, a nyelvvizsgákra való felkészítéskor is előkerül. A szóbeli produkciók értékelése kihívást jelentő feladat a nyelvvizsgáztatók számára. Bár egy meghatározott szempontrendszer alapján értékelnek, mégis időnként nehéz objektívnek maradni, mert a felsorolt kritériumok többféleképpen értelmezhetőek. Éppen emiatt van szükség többtagú vizsgabizottságokra. A Közös Európai Referenciakeretben (a továbbiakban KER) a szóbeli nyelvhasználat minőségi szempontjainál sincs teljesen objektív meghatározása a beszéd folyékonyágának. Nyelvi szintekre lebontva a következők olvashatók benne:

7. táblázat A folyékonyág meghatározása (KER, 2002: 36–37)

| Szint | Folyékonyág meghatározása   |
|-------|---|
| C2    | Természetes módon, hosszasan képes kifejezni magát, a hétköznapi fordulatokat természetes gördülékenységgel használja, és olyan gördülékenyen képes elkerülni bármilyen nyelvi nehézséget, hogy a beszélő partner alig veszi észre.             |
| C1    | Folyékonyan és természetes módon, szinte erőfeszítés nélkül tudja kifejezni magát. A természetes, gördülékeny beszédben csupán egy fogalmi szempontból nehéz téma akadályozhatja.   |
| B2    | Meglehetősen egyenletes beszédtempóval beszél. Bár időnként habozik, amikor szerkezeteken és kifejezéseken gondolkodik, emiatt csak kevés észrevehetően hosszabb szünet keletkezik.   |
| B1    | Érthetően tud folyamatosan beszélni, bár gyakran kell szünetet beiktatnia mondanivalójának nyelvtani vagy szókincsbeli megtervezése céljából, és mondanivalójának módosítása nagyon szembeűnő, elsősorban akkor, ha hosszabban beszél szabadon. |
| A2    | Nagyon rövid mondatokkal meg tudja értetni magát, bár gyakran megakad, újra kell kezdenie vagy át kell fogalmaznia mondanivalóját.  |
| A1    | Elboldogul nagyon rövid, különálló, többnyire előre betanult megnyilatkozásokkal, de sok szünetet tart, amikor a kifejezéseket keresgéli, a kevésbé ismerős szavakat próbálja kiejteni, vagy a mondanivalóját helyesbíteni.                     |

A fenti táblázatban (7. táblázat) láthatóak a nyelvi szintek elnevezései. A C2 az

anyanyelvi beszélőkhöz hasonló szintű nyelvtudást jelöli, a C1 a felsőfokú nyelvtudást, a B2 a haladó, a B1 a középhaladó, az A2 az alapozó szakaszt, és végül az A1 az abszolút kezdő vagy újrakezdő a KER rendszerben. A KER rendszerhez igazodva kerültek kidolgozásra a különböző nyelvvizsgák elvárásai Európában, valamint a vizsgán értékelő tanárok útmutatói is. Magyar mint idegen nyelvből jelenleg két nyelvvizsga-rendszerben lehetséges vizsgát tenni: a pécsi székhelyű ECL és a budapesti Origó nyelvvizsgaközpontok által kidogozottakban.

ECL nyelvvizsgát jelenleg A2–C1 nyelvi szinten lehet letenni magyar nyelvből. Ennél a vizsgatípusnál a beszédkészség értékelésére öt kritériumot határoztak meg: nyelvhelyesség, szóbeliség, szókincs, stílus és kommunikatív hatékonyság, amelyeket további öt-öt részre bontottak. Ahogy látjuk, itt is megjelenik a nyelvhelyesség és a fluensség elkülönítése. A beszédhelyesség az úgynevezett szóbeliség kritérium részét képezi és elsősorban a beszédtempóval, illetve a szünetek mennyiségével jellemezték. A szóbeliség másik része a kiejtés és a prozódia. A szóbeliséget minden nyelvi szinten a következőképpen határozzák meg:

**8. táblázat** Az ECL beszédkészség összetevőjének értékelési kritériumai (ECL, é.n.)

| <b>Pont</b> | <b>Szóbeliség - A beszéd folyékonyságának meghatározása</b>   |
|-------------|---|
| <b>5</b>    | Jó beszédtempó, kevés szünettel.  |
| <b>4</b>    | Ha a szóbeli produkció meghaladja a 3-as pontszámnál leírt teljesítményt, de nem feleltethető meg pontosan az 5-ös minősítéssel, akkor 4 pont az értékelés.                       |
| <b>3</b>    | Megfelelő beszédtempó; néhány helytelen szünet.   |
| <b>2</b>    | Ha a szóbeli produkció nem pontosan feleltethető meg a 3-as pontszámnál leírt teljesítménnyel, de felülmúlja az 1-es minősítést, akkor 2 pont az értékelés.                       |
| <b>1</b>    | Lassú és akadozó beszédtempó, önállóan beszédanyagot.   |
| <b>0</b>    | Ha a vizsgázó által nyújtott teljesítmény nem éri el az 1-es pontszámnál leírtakat, és a beszédprodukció vagy túl kevés, követhetetlen vagy irreleváns, akkor 0 pont az értékelés |

A beszédkészség értékelésének nehézsége nem a KER kidolgozatlanságából adódik, ha megnézzük egy másik nyelvvizsgarendszert az IELTS-et [International English Language Testing System], amely nem a KER alapján készült, láthatjuk, hogy ennek a vizsgának a leírásában sem részletes a beszédkészség egyik nyelvi szinthez kapcsolódó meghatározása sem. A beszédkészség értékelésénél pontot kapnak a vizsgázók az úgynevezett fluensség és koherencia [fluency and coherence] kritérium alapján, amely meghatározásában annyi szerepel csak, hogy a folyékonyság fő mutatói a beszédsebesség és a beszédfolytonosság. A KER szintjeinek megfelelően a következők olvashatók benne:

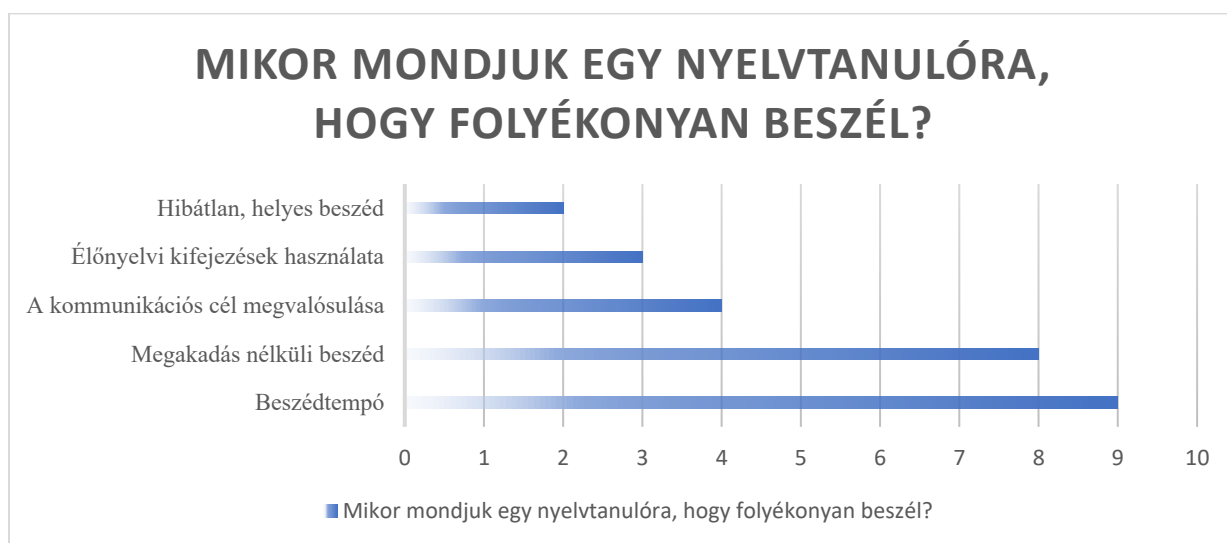
**9. táblázat** Az beszédkésztség értékeléseinek kritériumai (IELTS, 2018)

| Pont | Nyelvi szint neve                          | A KER szintek megközelítőleges megfeleltetése | Folyékony és koherens beszéd jellemzői   |
|------|--|---|--|
| 9    | Szakértői nyelvhasználó                    | C2  | Folyékonyan beszél, ritkán ismételt vagy javít, a hezitációja inkább a tartalommal kapcsolatos, mintsem a szavak keresésével vagy a grammatikával. Koherensen, összefüggően beszél. Beszédét megfelelően építi fel.    |
| 8    | Nagyon jó nyelvhasználó                    | C1  | Folyékonyan beszél, csak alkalmanként ismételt vagy javít. A hezitáció általában a tartalommal kapcsolatos és csak nagyon ritkán a nyelvvel. Koherensen és megfelelően építi fel a beszédét.                           |
| 7    | Jó nyelvhasználó                           |   | Különösebb erőfeszítés vagy koherenciavesztés nélkül, hosszasan beszél. Néha megjelenhet nála ismétlés, javítás vagy a nyelvtudás hiányosságaira utaló hezitáció. Rugalmasan használ sok kötőszót és diskurzusjelölőt. |
| 6    | Kompetens nyelvhasználó                    | B2  | Képes hosszán beszélni, bár néha nem koherensen az időnkénti ismétlés, javítás, hezitáció miatt. Sok kötőszót és diskurzus jelölőt használ, de nem mindig megfelelően.   |
| 5    | Szerény tudású nyelvhasználó               | B1  | A nyelvet folyamatosan használja általában, de ismételt, javítja magát, lassan beszél és halad tovább. Nyelvprodukciója egyszerű kommunikációnál folyékony, bonyolultabb beszélgetéseknél problémái adódnak.           |
| 4    | Korlátozott tudású nyelvhasználó           |   | Nem tud válaszolni észrevehető szünetek nélkül, lassan beszél, gyakran ismételt és javítja magát. A mondatokat össze tudja kötni, de csak egyszerű, ismételt kötőszókkal, a koherencia néha nem megfelelő.             |
| 3    | Rendkívül korlátozott tudású nyelvhasználó | A2  | Hosszú szünetekkel beszél, korlátozottan tudja összekötni a mondatait, csak egyszerű válaszokat ad és gyakran képtelen átadni az üzenetet.   |
| 2    | Időszakos nyelvhasználó                    | A1  | Sokszor, hosszán szünetet tart a legtöbb szó kimondása előtt. A kommunikáció csak kismértékben valósul meg.  |
| 1    | Nem nyelvhasználó                          |   | Nem lehetséges a vizsgázóval történő kommunikáció, nem minősíthető a nyelvismerete.  |
| 0    | Nem próbálta meg a vizsgát                 |   | Nem jelent meg.  |



Mivel a fenti feltételrendszer olyan kifejezéseket tartalmaz, mint a „természetes módon, hosszasan képes kifejezni magát” vagy „néha megjelenhet nála ismétlés, javítás” ennél a vizsgánál sem könnyű a vizsgáztató nyelvtanár feladata. Hiszen honnan tudjuk, hogy mit jelent pontosan a természetes módon történő kommunikáció? Vagy azt, hogy hány darab ismétlés és javítás után mondhatjuk, már gyakran teszi a vizsgázó? Érdekelt, hogy mit gondolnak erről a nyelvtanárok, ezért egy kérdőíves felmérés keretében megkérdeztem húsz pedagógus véleményét arról, szerintük mikor mondhatjuk egy nyelvtanulóra azt, hogy már folyékonyan beszél a célnyelvet. Arra kértem őket, hogy írjanak össze pár kritériumot, amely a megadott két perc alatt először az eszükbe jut. Húsz pedagógusból kilenc a beszédtempót, nyolc a megakadás nélküli beszédet, négy a kommunikációs cél megvalósulását, három az élőnyelvi kifejezések használatát említette meg a beszédfolyékonyság ismertetőjegyeként. Két pedagógus a hibátlan, helyes beszédet is felsorolta.

**5. ábra** Nyelvtanárok véleménye a beszédfolyékonyságról



Bár a felmérés nem reprezentatív, az eredmények mégis jól tükrözik azokat az irányelveket, nézőpontokat, amelyek a beszédfolyékonyság különböző definícióiban megjelentek az évek során.

## 6. A korpusznyelvészetről

A jelen értekezésben bemutatott korpuszalapú kutatáshoz szorosan hozzá kapcsolódik az alapjául szolgáló KorSzak Gyermeeknyelvi Korpusz építése is, amely egy speciális típusú korpusz. Ebből kifolyólag fontosnak tartom, hogy – mielőtt rátérnék a korpusz bemutatására – ismertessem a korpuszok típusait és azok tulajdonságait, valamint a kapcsolódó magyar és nemzetközi előzményeket.

### 6.1 A korpusz definíciója

Magyar nyelven a korpusznyelvészettel kapcsolatos legjobb összefoglaló Szirmai (2005) munkája. Maga a korpusznyelvészet kifejezés Aarts és Meijs 1984-ben megjelent könyvéhez köthető, bár csak sokkal később került bele a szakszótárakban (Szirmai, 2005: 17). Ugyanakkor elmondható, hogy az 1970-es évektől a korpusznyelvészet fejlődése lényeges változásokat hozott az alkalmazott nyelvészetben, és nagy hatást gyakorolt a nyelvek tanulmányozására (Hunston, 2022: 1). A számítógépek és az internet megjelenése lehetővé tette a korpusznyelvészeti kutatások szélesebb körű elterjedését, és ma már a nyelvoktatásban is számos módon hasznosítják.

Az évek során a korpusz szónak több meghatározása is született. Brazil (1995: 24) szerint a korpusz egy olyan gyűjtemény, amely az anyanyelvi beszélők által valóban használt nyelvet tartalmazza. A korpusz McCarthy definíciója szerint „írott- vagy beszélt nyelvi szövegek gyűjteménye, amelyet egy számítógépes adatbázisban tárolnak” (2004: 1). Tehát egy korpusz lehet írott nyelvi, amelyben könyvek, levelek, cikkek szövegei találhatóak. Illetve lehet beszélt nyelvi, amely többek között például beszélt nyelvi felvételek, televíziós vagy rádiós műsorok leiratait tartalmazza. Szirmai (2005) *Bevezetés a korpusznyelvészetbe* című könyvében összegyűjtötte az addig született meghatározásokat. A magyar nyelven készült definíciók közül kettőt említ:

1. A korpusz „meghatározott szempontok alapján kiválasztott szövegmennyiség, amelyen a nyelvész vizsgálatát végzi” (Kugler – Tolcsvai Nagy, 2000: 132).
2. „A korpusz ténylegesen előforduló írott, vagy lejegyzett beszélt nyelvi adatok gyűjteménye. A szövegeket valamilyen szempont szerint válogatják és rendezik. Nem feltétlenül egész szövegeket tartalmaz, és nem csak tárháza a szövegeknek, hanem tartalmazza azok bibliográfiai adatait, bejelöli a szerkezeti egységeket

(bekezdés, mondat). Emellett pedig feltünteti a szavak mellett szófaji kódjukat is” (MNSZ, 1998).

Szirmai az első meghatározást hiányosnak tartja, mert nem jelenik meg benne, hogyan tárolják, illetve milyen tartalommal rendelkezik egy korpusz. A másodikat már pontosabbnak gondolja, habár abból is hiányzik, hogy a szövegeket elektronikusan tárolják. Az angol szakirodalomban megjelenő definíciókat összefoglalva, Szirmai három fontos tényezőt emel ki, amelynek szerepelnie kell a korpusz meghatározásában.

A korpusz:

1. reprezentatív szövegek gyűjteménye,
2. elektronikus formában tárolt,
3. elemzése nyelvészeti szempontból hasznos lesz (vö. Stefanowitsch, 2020: 21–61).

Szerinte a „gondosan összeválogatott szövegfájlok tára korpuszt képez” (Szirmai, 2005: 18). Ilyen lehet például egy nyelvtanuló célnyelven írt szövegeinek gyűjteménye. Ugyanakkor fontos megjegyezni, hogy „a korpusz nem pusztán szövegek véletlen halmaza, hanem tudatosan megtervezett gyűjtemény, amelynek összeállításakor az ezen elvégezni kívánt nyelvi elemzést tartjuk szem előtt” (Szirmai, 2005: 23). Hunston (2022) Szirmaihoz hasonlóan azt gondolja, hogy egy digitalizált szövegyűjteménytől a korpuszok a szerkesztésükben térnek el. A korpuszt minden esetben úgy hozzák létre, hogy az a kutatás céljának megfelelő legyen, valamint egy számítógépes program alkalmazásával lehetővé váljon különböző mintázatok [patterns] megfigyelése, amely nem valósulhatna meg hagyományos módszerekkel (Hunston, 2022: 1).

### **6.1.1 A korpusz reprezentativitása és mérete**

A korpusz szó meghatározásának ismertetése után ebben a részben a korpuszok két tulajdonságának bemutatására kerül sor: a korpuszok reprezentativitása és mérete. A reprezentativitás az egyik fontos jellemzője volt a korpuszoknak. Több nyelvész azonban felhívta a figyelmet arra, nem biztos, hogy lehetséges a reprezentativitás tökéletes megvalósítása (Atkins et al., 1992 idézi Szirmai, 2005: 23) A 2000-es években úgy gondolták, hogy a reprezentativitás a korpusz méretén kívül a kiegyensúlyozottságával is összefüggésben áll, tehát hogy a céljainknak megfelelően felépített korpuszban, milyen arányban jelennek meg

a különböző alkorpuszok, a különböző szövegtípusok (Stefanowitsch, 2020: 28).

Az utóbbi időben azonban egy nézőpontbeli változás történt a korpuszok értelmezésében. Míg korábban a korpuszokra kész termékként, végleges produktumként tekintettek, ma már inkább az úgynevezett dinamikus vagy másnéven monitor korpuszokat alkalmazzák. Ezekből a korpuszokból soha nem lesz kész termék, hanem új anyagok rendszeres hozzáadásával folyamatosan bővülő szövegek gyűjteményeként működnek. A korpuszépítésben ma a hangsúly egy kiegyensúlyozott korpuszról inkább a megfelelő mennyiségű szöveg biztosítására és a rugalmasságra tolódott, hogy a nyelvészek a kutatásuk céljának megfelelő saját alkorpuszt tudjanak összeállítani belőlük (Hunston, 2022: 20).

A méreteket illetően a korpuszokat token és type szerint szokták meghatározni. A token azt mutatja meg, hogy a korpusz összesen, ismétlődésekkel együtt hány darab szóból épül fel. Ezzel szemben a type mutató azt jelzi, hogy hányféle szó található benne (Szirmai, 2005: 28). Egy korpusz lehet kicsi, például 50.000 tokent tartalmazó, illetve hatalmas, több millió tokent tartalmazó is, de jellemzően több tokent tartalmaznak, mint amennyit egy személy egyszerre el tudna olvasni és meg tudna jegyezni (Hunston, 2022: 1). Korábban a korpuszok méreteit behatárolta az azokat tároló számítógépek memóriájának mérete, azonban az akkori főleg grammatikai vizsgálatokhoz már egy egymillió szavas korpusz is megfelelő volt. Így sokáig az egymillió szavas korpuszokat tekintették az alaptípusnak. Változás ebben az 1980-as, majd az 1990-es években történt, amikor – a technikai fejlődés mellett – egyre jobban a lexikára, valamint a különböző mintázatok keresésére, a lexika és a grammatika kapcsolatának tanulmányozására helyeződött a hangsúly (ld. Sinclair, 1991), amelyekhez már nagyobb korpuszokat használtak (Hunston, 2022: 18–19). A nagy korpuszok, mint például a 100 millió szót tartalmazó British National Corpus (BNC), az egymilliárd szót tartalmazó Corpus of Contemporary American English (COCA), a 4,5 milliárd szót tartalmazó Collins Birmingham University International Language Database (COBUILD), az összesen 10 milliárd szót tartalmazó Corpus del Español, vagy magyar példát hozva, a Magyar Nemzeti Szövegtár (MNSZ2), amely több mint egymilliárd szavas, jól használhatóak a fent említett korpusznyelvészeti vizsgálatok megvalósítására, ugyanakkor ez nem jelenti azt, hogy a kisebb, saját gyűjtésű korpuszoknak nincs létjogosultságuk vagy kevésbé lennének alkalmasak nyelvészeti kutatásokra. A kisebb korpuszokat más célok miatt hozzák létre, legyenek akár egy nagyobb korpusz alkorpuszai vagy önálló kisebb korpuszok, amelyekben egy-egy jelenséget szeretnének megvizsgálni, amely csak bizonyos szövegtípusokban jelenik meg (vö. Stefanowitsch, 2020: 21–61). Ilyen lehet például a különböző regiszterek megfigyelése. Ezen a területen Biber (1988) ért el először áttörő eredményeket, amikor azonos nagyságú, más-más

regiszttert tartalmazó alkorpuszokat épített. Ez azért volt jelentős munka, mert előtte nehezen sikerült a nem azonos kritériumok alapján épített korábbi korpuszok újabbakkal való összehasonlítása (Hunston, 2022: 19–20).

Összefoglalva tehát elmondhatjuk, hogy a korpusz tulajdonságaival kapcsolatban a legfontosabb, hogy a céljainknak megfelelő mennyiségű szöveg álljon rendelkezésünkre egy-egy vizsgálathoz, illetve, hogy elég rugalmasan kezelhető legyen a korpuszunk.

### 6.1.2 A korpuszok típusai

A korpuszoknak különböző típusai léteznek (10. táblázat), amelyek között szerepelnek az általános és speciális korpuszok, szinkron és diakron, egynyelvű és többnyelvű, összehasonlítható, multimodális, pedagógiai, valamint tanulói korpuszok is. A következőkben bemutatásra kerül, hogy milyen jellegzetességekkel rendelkeznek ezek a korpuszok és milyen kutatási területeken alkalmazzák őket a nyelvészek.

10. táblázat A korpuszok típusai felhasználási módjuk szerint Hunston (2022: 21–30) és saját gyűjtésű példák alapján

| A korpusz típusa           | Tulajdonság   | Példa   |
|----------------------------|---|---|
| <b>Általános</b>           | Írott- és beszélnyelvi szövegek különböző fajtáinak gyűjteménye.  | Magyar Nemzeti Szövegtár  |
| <b>Speciális</b>           | Egy típusú, regiszterű szövegek gyűjteménye.  | Cambridge and Nottingham Corpus of Discourse in English (CANCODE), KorSzak Gyermekegyelvi Korpusz   |
| <b>Szinkron</b>            | Egy adott korszakból tartalmaz csak szövegeket.   | The English Web Corpus (enTenTen), KorSzak Gyermekegyelvi Korpusz   |
| <b>Diakron</b>             | Több korszakból tartalmaz szövegeket.   | Helsinki Corpus of English Texts (angol nyelvű szövegek a Kr.e. 9.századtól a Kr.u. 11. századig, a 12–15. századból, valamint a 16–18.századból) |
| <b>Egynyelvű</b>           | Csak egynyelvű szövegeket tartalmaz.  | Collins Corpus, KorSzak Gyermekegyelvi Korpusz  |
| <b>Paralel, többnyelvű</b> | Több nyelven tartalmazza az adott szövegeket, hogy össze lehessen őket hasonlítani. Elsősorban a fordítások | European Comparable and Parallel Corpora of Parliamentary Speeches Archive (ECPC Corpus)  |

|                          |   |   |
|--------------------------|---|---|
|                          | vizsgálatának céljából állítják össze őket.   |   |
| <b>Összehasonlítható</b> | Azonos kritériumok alapján összeállított szövegek, több nyelven. Elsősorban a szavak és kifejezések ekvivalenciájának megtalálása céljából vizsgálják őket. | The Child Language Data Exchange System (CHILDES)   |
| <b>Multimodális</b>      | Nemcsak írott- és beszélt nyelvi, hanem vizuális információk is megtalálhatóak benne.   | Hungarian Multimodal Corpus (szimulált állásinterjúk és beszélgetések személyesebb témákról), KorSzak Gyermeznyelvi Korpusz |
| <b>Pedagógiai</b>        | Pedagógiai céllal létrehozott korpuszok.  | KorSzak Gyermeznyelvi Korpusz   |
| <b>Tanulói</b>           | A nyelvtanulók beszédprodukcióit és írásaikat tartalmazó korpusz.   | Cambridge Learner Corpus (CLC), a KorSzak Tanulói korpusz   |

Biber (1988) más-más regisztert tartalmazó alkorpuszai hozzájárultak ahhoz, hogy ma már számos típusú korpusz álljon a rendelkezésünkre. Kategorizálásuk során kétféle megközelítést alkalmazhatunk: a mintavétel módja szerint, illetve a felhasználás módja szerint csoportosíthatjuk őket.

A mintavétel módja szerint egy korpusz lehet statikus, vagyis olyan korpusz, amely az összeállítása után már nem változik. Ezeket a korpuszokat jól lehet összehasonlító vizsgálatokra használni. Ilyen például a fenti táblázatban (10. táblázat) látható, angol nyelvű szövegeket tartalmazó *Helsinki Corpus of English Texts*. A második típus a dinamikus korpusz, amely folyamatosan fejlődik, bővül. A harmadik típus, a monitor korpusz Szirmai (2005: 32) szerint a statikus és a dinamikus ötvözeteként, időnként bővül. Ezzel szemben a fogalmat Hunston (2022: 26) a dinamikus korpusz szinonimájaként használja. A dinamikus vagy monitor korpusz típusára jó példa a fenti táblázatban lévő *Cambridge Learner Corpus*, amely a Cambridge nyelvvizsgán vizsgázók szövegeivel minden évben meghatározott időben, a nyelvvizsgák kiértékelése után bővül.

A korpuszok felhasználási módja szerint beszélhetünk általános korpuszokról és speciális korpuszokról, amelyeknek szintén több típusa létezik. Az általános és a speciális korpusz közötti különbség nem éles, inkább céljaikban térnek el egymástól. Egy általános korpusz feladata, hogy bemutasson egy nyelvet vagy egy nyelvváltozatot, míg egy speciális korpusz arra szolgál, hogy egy adott témával kapcsolatos kérdésekre választ találjunk belőle.

Egy speciális korpusz egy bizonyos típusú szöveget, vagy egy bizonyos kritériumrendszer alapján összegyűjtött szövegeket tartalmaz, így az adott témához, regiszterhez vagy szituációhoz kapcsolódó nyelv tanulmányozható a vizsgálatával. Az általános korpuszokat a korpusznyelvészeti kutatásokban gyakran használják referenciakorpuszként, hogy egy speciális korpuszsal összehasonlítva többek között megfigyeljék egy adott nyelvi elem jellemző gyakoriságában felelhető különbségeket (Hunston, 2022: 21–23). Ezenkívül még számos módon kategorizálhatjuk a korpuszokat, de az értekezés keretei miatt csak azoknak a típusoknak a tulajdonságait részletezem a továbbiakban, amelyek a kutatás alapjául szolgáló korpuszt is leírják vagy valahogyan kapcsolatban állnak vele.

A fenti táblázatban (10. táblázat) a kutatás alapjául szolgáló KorSzak Gyermeknyelvi korpuszt több kategóriában is említtem. A KorSzak Gyermeknyelvi Korpusz felhasználási módja szerint speciális, a mintavétel módja szerint pedig szinkrón. A korpusz egynyelvű, ugyanakkor multimodális, mert videófelvételeket is tartalmaz. Ezenkívül gyermeknyelvi és pedagógiai korpusz is egyben.

## **6.2 Gyermeknyelvi korpuszok**

A fenti táblázatban (10. táblázat) látható, hogy a jelen kutatás alapjául szolgáló KorSzak Gyermeknyelvi Korpuszt a pedagógiai korpuszok közé sorolom. Bár a nevében szerepel a gyermeknyelvi kifejezés – mivel gyermekek szövegeit tartalmazza – a korpusz felhasználási céljai miatt inkább a pedagógiai korpuszok közé sorolható. A következőkben bemutatom, hogy az eddigi gyermeknyelvi korpuszok miben térnek el a KorSzak Gyermeknyelvi Korpusztól és miért volt szükséges a kutatás céljainak elérése érdekében egy új korpusz építése.

A gyermeknyelvi korpuszok általában az anyanyelvi nyelvfejlődést vizsgálják. Az ilyen típusú korpuszokban egy bizonyos időben vagy időablakban vizsgálják meg, hogy milyen a megfigyelt gyermekek szókincsének fejlődése, hogyan változik a grammatikai és pragmatikai tudása. A nyelvfejlődés hatékony megfigyelése érdekében a korpusz építéskor jól és előre megtervezett struktúrára van szükség, ezért időben el kell döntenünk, hogy a vizsgálatok két fajtája közül – keresztmetszeti [cross-sectional] és longitudinális – melyiket választjuk. Egy keresztmetszeti korpuszban minden gyermek beszédét egyénileg rögzítik természetes vagy természeteshez hasonló körülmények között. Az ilyen típusú korpuszokban több résztvevő adatainak átlagolásával vonnak le következtetéseket. Amíg ebben a típusban minden gyermek beszédét a fejlődésének csak egy bizonyos pontján rögzítik, ezzel szemben a longitudinális korpuszokban egyetlen vagy kevés számú gyermekkel rendszeres időközönként készülnek

felvételek, és ennek a gyermeknek, gyermekeknek a korpuszban található mintái alapján vonnak le következtetéseket a nyelvfejlődésükkel kapcsolatosan. A kutatás folyamata általában több hónapon vagy akár éveken keresztül is eltarthat (Stoll – Schikowsky, 2020: 505).

**11. táblázat** A gyermeknyelvi korpuszvizsgálatok típusai és azok jellemzői

| <b>Korpusz típusa</b>                             | <b>Keresztmetszeti</b>  | <b>Longitudinális</b>  |
|---|---|--|
| <b>Résztevők száma</b>                            | Több résztvevő.   | Egy, esetleg néhány gyermek.                                   |
| <b>Felvételek rendszeressége</b>                  | Egy résztvevő gyermekkel sokszor csak egyetlen felvétel készül. | Több, rendszeres időközönkénti felvétel készül a résztvevőkkel |
| <b>A kutatás időtartama</b>                       | Sokszor rövidebb, mint a longitudinálisé.                       | Általában hosszabb időkerettel dolgoznak.                      |
| <b>Az következtetéshez hozzájáruló eredmények</b> | Több gyermek mintáinak átlaga.                                  | Általában egy, esetleg néhány gyermek mintáinak megfigyelése.  |

Ha a fenti táblázatban (11. táblázat) is bemutatott kétféle kutatási típus alapján szeretnénk a KorSzak Gyermeknyelvi Korpuszt meghatározni, akkor az egy keresztmetszeti vizsgálatnak felel meg. Bár vannak olyan adatközlő gyermekek, akikkel többször is készültek felvételek a korpuszpépítés eddigi három éve alatt, nem egy-egy gyermek nyelvi fejlődésének megfigyelése volt a korpusz készítésének célja, hanem egy korosztály (11–15 évesek) nyelvhasználatának vizsgálata. A korpuszban jelenleg (2022 októberében) 78 felvétel található, amelyen meghatározott témák alapján beszélget a 29 gyermek adatközlő (9 fiú és 20 lány).

A kutatások kiterjedten tanulmányozzák az anyanyelv elsajátítása mellett több nyelv egyidejű elsajátítását is. Nagy számban találhatóak meg ebben a kategóriában a többnyelvű gyermekekkel készült hanganyagok és leirataik (vö. Deuchar, 2022: 58–81). A KorSzak Gyermeknyelvi Korpusz ebben sem egyezik a többi gyermeknyelvi korpuszsal. Ezenkívül a felvételeken szereplő gyermekek életkorában is eltérések figyelhetők meg a kategória más korpuszaihoz képest. A hagyományos gyermeknyelvi korpuszok adatközlői ugyan több életkorból kerülnek ki, túlnyomó részben a kisgyermekek (0–5 évesek) jelennek meg bennük. Ennek ellenére úgy gondolom, hogy érdemes megemlíteni néhány kiemelkedően jelentős gyermeknyelvi korpuszt, mert bár nem azonos céllal és módon épültek, mégis a KorSzak Gyermeknyelvi Korpusz előzményeinek tekinthetőek.



### 6.2.1 CHILDES

A Gyermeknyelvi adatcsere rendszer<sup>10</sup> [Child Language Data Exchange System] vagyis a CHILDES adatbázis egy olyan gyermeknyelvi adatbázis, amely nagy része angol nyelvű, de huszonkét más nyelvű korpuszt is tartalmaz. Az angol, német, dán, svéd, norvég, francia, olasz, görög gyermekektől származó hanganyagok átíratái mellett több magyar nyelvű vagy olyan kétnyelvű felvétel is található benne, amelynek az egyik nyelve a magyar (Neuberger, 2014: 55). A CHILDES weboldalán<sup>11</sup> ingyenesen férhetünk hozzá a korpuszhoz és a kutatásunkhoz szükséges korpusznyelvészeti eszközökhöz is (Szirmai, 2005: 87), valamint a megfelelő annotáció, vagyis kódolás után, mi is feltölthetjük a saját felvételeinket (Neuberger, 2014: 56). A CHILDES több okból is kiemelkedő jelentőségű projekt, de talán a legfontosabb fejlesztése az, hogy olyan eszközöket és protokollt terveztek meg hozzá, amelyek megkönnyítik a gyermekek felvételeinek átírását, növelik az átírás megbízhatóságát, megkönnyítik az adatelemzést egyes folyamatok automatizálásával, valamint az adatok megosztását is egyszerűsítették. Ezzel jelentős változásokat hozva a gyermeknyelv kutatásának módszereiben (MacWhinney, 2001: 1). A projekt több korai próbálkozás után, 1984-ben indult el Brian MacWhinney és Catherine Snow vezetésével. Az első szakaszban az adatgyűjtésre fókuszáltak a kutatók, de hamar kiderült, hogy szükség van az adatok gyűjtésének, az átíratok elkészítésének egységesítésére, ezért a második szakaszban, 1986 és 1991 között kidolgozták a fent említett eszközöket (MacWhinney, 2001: 2). A szövegek átíratait az úgynevezett CHAT szabvány szerint formázták és a szintaktikai pontosságukat ellenőrizve töltötték fel az adatbázisba. Az adatbázis teljes mérete ebben az időszakban megközelítőleg 140 millió karakter volt, amely azóta tovább növekedett. A korpuszt kezdetben hat fő könyvtárba rendezték: angol nyelvű felvételek, nem angol nyelvű felvételek, narratívák, nyelvi zavar, kétnyelvűség elsajátítása és könyvek. Ma ezeknél már több található a felületen, sajnos nem túl felhasználóbarát módon, vagyis nem könnyen átlátható struktúrában. 2001-től a projekt az NSF Infrastructure támogatásának köszönhetően újabb eszközöket fejlesztett ki, illetve növekedett az adatbázis mérete is, ezért ettől az évtől kezdve már egy új néven, TalkBank rendszerként hivatkoznak rá (MacWhinney, 2022: 11). A korpusz jelenlegi adatairól, méreteiről összefoglalóan a TalkBank felületén nem derül ki sok információ, egy része viszont már az általam is használt és a későbbiekben (III.2.1 fejezet) részletesen is bemutatásra kerülő Sketch Engine elnevezésű korpuszeszközöket tartalmazó felületére is felkerült, ahol a következő

---

<sup>10</sup> A korpusz elnevezésének magyar fordítását Neuberger (2014) könyvéből kölcsönöztem.

<sup>11</sup> <https://childes.talkbank.org/>

adatokat olvashatjuk:

**12. táblázat** A Sketch Engine felületen megtalálható CHILDES korpuszok

|     | <b>Korpusz neve</b>       | <b>Szavak száma</b> |
|-----|---------------------------|---------------------|
| 1.  | CHILDES Thai Corpus       | 243.939             |
| 2.  | CHILDES Afrikaans Corpus  | 26.020              |
| 3.  | CHILDES Catalan Corpus    | 209.525             |
| 4.  | CHILDES Croatian Corpus   | 300.832             |
| 5.  | CHILDES Danish Corpus     | 285.231             |
| 6.  | CHILDES English Corpus    | 22.693.506          |
| 7.  | CHILDES Estonian Corpus   | 313.457             |
| 8.  | CHILDES Farsi Corpus      | 120.527             |
| 9.  | CHILDES French Corpus     | 2.583.460           |
| 10. | CHILDES Gaelic Corpus     | 16.848              |
| 11. | CHILDES German Corpus     | 5.941.266           |
| 12. | CHILDES Hebrew Corpus     | 807.657             |
| 13. | CHILDES Hungarian Corpus  | 247.881             |
| 14. | CHILDES Italian Corpus    | 459.881             |
| 15. | CHILDES Japanese Corpus   | 1.578.068           |
| 16. | CHILDES Korean Corpus     | 36.056              |
| 17. | CHILDES Norwegian Corpus  | 56.827              |
| 18. | CHILDES Polish Corpus     | 1.041.300           |
| 19. | CHILDES Portuguese Corpus | 216.407             |
| 20. | CHILDES Russian Corpus    | 48.791              |
| 21. | CHILDES Spanish Corpus    | 802.743             |
| 22. | CHILDES Swedish Corpus    | 520.478             |
| 23. | CHILDES Tamil Corpus      | 15.490              |
| 24. | CHILDES Turkish Corpus    | 178.100             |

Az eredeti TalkBank weboldalon a korpuszok nagy részénél csak a leirat található meg, egy részénél, köztük néhány magyar felvételnél a hanganyagot is meghallgathatjuk, amíg a legújabbaknál már videókat is lehet látni. A videók lejátszásánál a leiratban az éppen beszélő személy szövegét sárgával emelik ki, ami nagy segítséget jelenthet a kutatóknak. A hanganyagok és a videók megjelenésével a kutatási témák is kiszélesedtek, hiszen így már több olyan aspektust is megfigyelhetünk a gyermekek beszédfejlődésében, amelyet csak a leiratok alapján nem lehetett volna. Ilyen többek között a nonverbális eszközök használata.

6. ábra Egy spanyol kisfiú és szüleinek felvétele a CHILDES adatbázisban

The screenshot shows the CHILDES database interface. On the left, there is a file browser showing a folder structure: 'chilides / Spanish / Ornat /'. Below it is a video player showing a young boy. The main area displays a transcript and participant information.

**Transcript:**

| CHAT    | path                           | filename | languages | media | date | pid                | design type | activity type | group type |
|---------|--------------------------------|----------|-----------|-------|------|--------------------|-------------|---------------|------------|
| 010700m | chilides/Spanish/Ornat/010700m | 010700m  | spa       | video | -    | 11312/c-00032724-1 | long        | toyplay       | TD         |

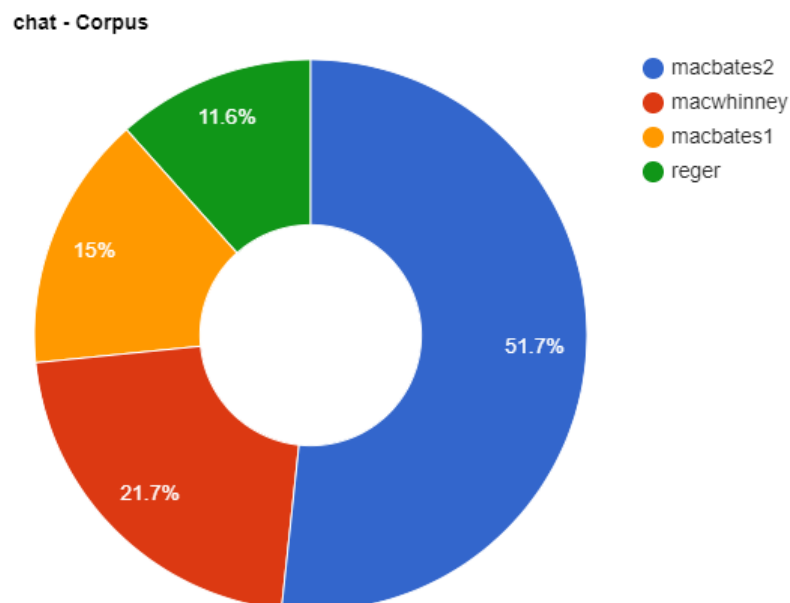
**Participants:**

| participant | role         | name  | language | age     | sex |
|-------------|--------------|-------|----------|---------|-----|
| CHI         | Target_Child | Maria | spa      | 1;07.00 | -   |
| MAD         | Mother       | -     | spa      | -       | -   |
| PAD         | Father       | -     | spa      | -       | -   |

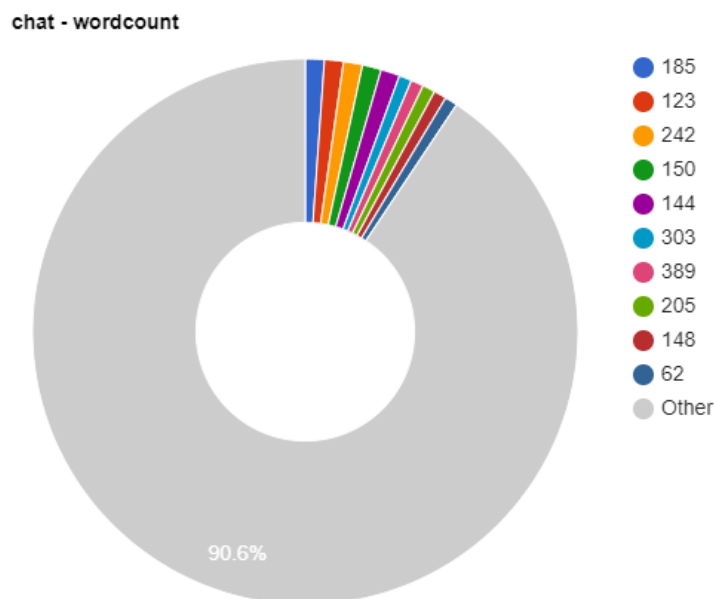
Below the tables, there is a 'View dependent tiers:' checkbox and a list of metadata tags such as @UTR#, @PID, @Begin, @Languages, @Participants, @ID, @Media, @Comment, @Tape Location, @Situation, and @Types.

Sajnos a CHILDES korpusz nem minden modern korpusznyelvészeti eszközzel használható. Ha például megnézzük a Sketch Engine felületén megtalálható részét, láthatjuk, hogy az azon található programok a CHAT szabványnak csak egy részét ismerik fel. A CHAT adatait a rendszer metaadatként rögzítette ugyan (7. ábra), de nem minden esetben lehet ebből az információból megbízható vagy értelmezhető adatokat kinyerni (8. ábra).

7. ábra CHILDES magyar nyelvű korpusz a Sketch Engine felületén a korpusz készítői szerint



8. ábra CHILDES magyar nyelvű korpusz a Sketch Engine felületén a szószámok szerint



Ezenkívül a funkciók közül nem érhető el a kollokációk [Word Sketch] keresése, valamint két elem kollokációinak összehasonlítása [Word Sketch Difference] sem. Ezeket a korpuszeszközöket is bővebben be fogom mutatni a Sketch Engine működésének leírása során (III.2.1 fejezet), ugyanakkor ennél a résznél is fontosnak gondoltam megemlíteni. Arra a kérdésre, hogy miért nem működnek ezek a funkciók a CHILDES korpuszon, kétféle választ kaptam. A Sketch Engine terméktámogatással foglalkozó ügyintézője szerint a korpuszból hiányzik néhány alapvető annotáció, amely megakadályozza bizonyos eszközök használatát (például nem lehet meghatározni egyes szavak lemmáját vagy szófaját). Ez valószínűleg a korpusz korából és méretéből adódik, amely alkalmatlanná teszi a korpuszt a fejlettebb eszközökkel való feldolgozásra. A másik választ MacWhinney professzor úrtól kaptam, aki szerint a Sketch Engine még nem elég fejlett az ilyen jellegű korpuszok és a magyar nyelv elemzéséhez, ezért ajánlatos a TalkBank felületén található korpuszeszközöket használni helyette. A TalkBank felületén valóban rengeteg korpusznyelvészeti kutatásokhoz használható szoftver található. Többek között találhatóunk leiratokat készítő, kódoló és elemző (CLAN), a CHAT-ben átírt fonológiai és fonetikai adatok elemzésére szolgáló (Phon), egyéb kódoló (NVivo7), beszédelemző (Praat), valamint különböző JAVA scripteken alapuló programokat, amelyek segítik a korpusz XML formátumban történő kezelését.

## 6.2.2 Magyar gyermeknyelvi korpuszok

A magyar nyelvet illetően a korábbi korpusznyelvészeti kutatások, amelyek a gyermekek nyelvhasználatával foglalkoztak szintén elsősorban az első nyelv elsajátítási folyamatának megértése és tanulmányozása céljából készültek. Ezeknek a korpuszoknak egy része megtalálható a már említett TalkBank rendszerben (13. táblázat). A hangfelvételek között találhatóak egynyelvű (vö. MacWhinney, 1970; Réger, 1990; Bodor, 1994) és többnyelvű (vö. Navracsics, 1994–1997; Bíró, 2013) gyermekekkel készítették is.

A legnagyobb magyar gyermeknyelvi korpuszok azonban – mint a 62 óvodás bevonásával készült MONYÉK, vagyis a Magyar Óvodai Nyelvi Korpusz (vö. Mátyus – Orosz, 2014), a 3–18 évesek felvételeivel folyamatosan épülő GABI (vö. Bóna, 2017) vagy a 16–18 évesekről készült felvételeket tartalmazó Beszélnyelvi Adatbázis (BEA) TiniBea része (vö. Gyarmathy – Neuberger, 2015) – nem képezik részét a CHILDES rendszernek. Ugyanakkor hasonlóan az ott található felvételekkel, az ezeken alapuló kutatások nagy része szintén az első nyelv elsajátításának fázisait vizsgálják, illetve fonetikai kutatásokat végeznek rajtuk.

A jelen kutatás céljaihoz a fent említett GABI és TiniBea korpuszok állnak a legközelebb. A GABI korpusz (Bóna, 2017), ahogy említésre került 3–18 éves gyermekek felvételeit tartalmazza, amelyek a BEA korpusz alapján, de kétféle protokollal készültek (3–9 évesek és 9–18 évesek számára) (vö. Bóna, 2017: 42–43). Mindkét protokoll többféle feladtból áll, ilyenek többek között a 15 különféle modalitású, hosszúságú, fonetikailag változatos mondat utánmondása, spontán narratíva, szavak és kifejezések definiálása, a történetmesélés képsorozat alapján, vagy a vita két gyermek között megadott témában. A felvételek között monológok és dialógusok is találhatóak, de arányaiban inkább az előbbire koncentrálnak. A korpuszépítők célja, hogy minden korosztállyal 50 - 50 felvétel készüljön, 2017-es könyvében Bóna Judit 405 felvételtől számolt be. A másik, a TiniBea korpusz az idősebb korosztállyal, 16–18 éves gimnazistákkal készült felvételeket tartalmaz, szintén a BEA korpusz alapján összeállított saját protokollal rendelkezik. A feladatok között megtalálható a mondatismétlés, a narratíva, a véleménykifejtés, a történetmesélés és az úgynevezett zsebszöveg feladat (Gyarmathy – Neuberger, 2015: 214–215), amelyben a gyermekeknek érdekes szituációkra kell reagálniuk párokban. A szerzők 2015-ös adatai alapján a korpusz 18 felvételt tartalmaz. Tudományos szempontból mindkét korpusz hiánypótló és nagy jelentőségű. Mindkettő kiválóan használható különféle beszédfejlődés-vizsgálatok alapjaként, ugyanakkor egyik sem teljesen alkalmazható a jelen kutatás céljainak megvalósításához.

13. táblázat A CHILDES adatbázisban megtalálható magyar vonatkozású gyermeknyelvi korpuszok összehasonlítva a KorSzak Gyermekegyelvi Korpuszal

| Név                                       | Kutatás típusa              | Résztevők száma             | Cél  | Kor                        | Nyelv                    | Gyűjtő  | Gyűjtés helye                          | Gyűjtés ideje | Média típusa                   | Elérhetőség  | Email                         |
|---|-----------------------------|-----------------------------|--|----------------------------|--------------------------|---|--|---------------|--------------------------------|--|-------------------------------|
| <b>KorSzak Gy. Korpusz</b>                | keresztmetszeti             | 29                          | MID tananyag készítése gyermekek számára   | 11-15                      | magyar                   | KorSzak Gyermekegyelvi munkacsoport   | Mo., (jelenleg: Baranya vm., Pest vm.) | 2020-jelenleg | videó, audió                   | elérhető a KorSzak Gyermekegyelvi munkacsoport számára | korszakgyermeknyelv@gmail.com |
| <b><u>Biro Trilingual Corpus</u></b>      | longitudinális, természetes | 1                           | Egy háromnyelvű nyelvi gyermek fejlődésének dokumentálása.                         | 1;7.2<br>8 -<br>3;7.2<br>5 | magyar, spanyol, katalán | Andrea Biró, Translation and Language Sciences, University Pompeu Fabra, Barcelona  | Spo., Mo.                              | 2012-2014     | videó                          | elérhető, CHILDES                                      | andreabiro@gmail.com          |
| <b><u>Navracsecs Bilingual Corpus</u></b> | longitudinális, természetes | 2 + felnőttek               | Egy testvérpár nyelvi fejlődésének dokumentálása.                                  | 3-5;<br>2-4                | magyar, perzsa, angol    | Judit Navracsecs, Department of Applied Linguistics, Veszprém University  | Mo.                                    | 1994-1997     | nem elérhető                   | elérhető, CHILDES                                      | navracscj@almos.vein.hu       |
| <b><u>MacBates1 Corpus</u></b>            | keresztmetszeti, képleírás  | 120 / 40 magyar /30 gyermek | Egynyelvű gyermekek nyelvi fejlődése 3-6 év alatt, mondatszerkezetek megfigyelése. | 3;4;5                      | magyar (angol, olasz)    | Brian MacWhinney, Department of Psychology, Carnegie Mellon University; Elizabeth Bates, Department of Psychology, University of California San Diego | Mo., USA, Olaszó.                      | 1978          | színes képek (stimulus), audió | elérhető, CHILDES, Sketch Engine                       | macw@cmu.edu                  |
| <b><u>MacBates2 Corpus</u></b>            | filmléírás                  | 173 + 103 felnőtt           | Egynyelvű gyermekek nyelvi fejlődése.  | 3;6;10                     | magyar (angol)           | Brian MacWhinney, Department of Psychology, Carnegie Mellon University; Elizabeth Bates, Department of Psychology, University of California San Diego | Mo., USA                               | 1978          | film (stimulus), audió         | elérhető, CHILDES, Sketch Engine                       | macw@cmu.edu                  |

| Név                             | Kutatás típusa  | Résztevők száma | Cél   | Kor       | Nyelv  | Gyűjtő   | Gyűjtés helye                          | Gyűjtés ideje | Média típusa | Elérhetőség   | Email                         |
|---------------------------------|-----------------|-----------------|---|-----------|--------|--|--|---------------|--------------|---|-------------------------------|
| <b>KorSzak Gy. Korpusz</b>      | keresztmetszeti | 29              | MID tananyag készítése gyermekek számára                  | 11-15     | magyar | KorSzak Gyermeknyelvi munkacsoport                                     | Mo., (jelenleg: Baranya vm., Pest vm.) | 2020-jelenleg | videó, audió | elérhető a KorSzak Gyermeknyelvi munkacsoport számára | korszakgyermeknyelv@gmail.com |
| <b><u>Bodor Corpus</u></b>      | természetes     | 1               | Saját egynyelvű gyermek nyelvi fejlődésének megfigyelése. | 2-2;10    | magyar | Péter Bodor, Department of Social Psychology, ELTE University          | Mo.                                    | 1994-1995     | audió        | elérhető, CHILDES                                     | peter.bodor@gmail.com         |
| <b><u>MacWhinney Corpus</u></b> | természetes     | 5 + 1 felnőtt   | Egynyelvű gyermekek nyelvi fejlődésének megfigyelése.     | 1,3-2;2-3 | magyar | Brian MacWhinney, Department of Psychology, Carnegie Mellon University | Mo.                                    | 1970-1971     | audió        | elérhető, CHILDES                                     | macw@cmu.edu                  |
| <b><u>Réger Corpus</u></b>      | esettanulmány   | 1               | Egynyelvű gyermekek nyelvi fejlődésének megfigyelése.     | 1;11-2;11 | magyar | Zita Réger, Linguistics Hungarian Academy of Sciences                  | Mo.                                    | 1991-1992     | audió        | elérhető, CHILDES                                     | -                             |

### 6. 3 A pedagógiai célú korpuszok

Ahogy fent említettem a jelen kutatás alapjául nem a már meglévő gyermeknyelvi korpuszokat használtam, amelynek az oka az, hogy tervezésük célja eltér és nem egyeztethető össze a kutatás céljaival. Ahogy Hunston (2022) megállapítja, alapvető fontosságú annak megértése, hogy a korpusz mindig olyan objektum, amelyet célirányosan építettek fel. A benne található szövegek úgy vannak megválasztva, hogy a korpuszban konkrét információkat lehessen vizsgálni. A szövegek korpuszba kerüléséhez készült kritériumok meghatározásakor figyelembe kell venni a korpusz célját, amely a kutatás céljával azonos (Hunston, 2022: 18).

A KorSzak Gyermeknyelvi Korpusz egy pedagógiai céllal épülő korpusz, amelynek szövegeit, valamint a ráépülő korpuszalapú kutatás eredményeit gyermekeknek szóló tananyag készítéséhez szeretnénk felhasználni. A jelen doktori értekezésben a pedagógiai korpusz nyelvoktatásra készült autentikus szövegek halmazát jelenti. A korpuszban szereplő szövegek összeállítása tudatosan történt a tananyagkészítést mint elsődleges célt szem előtt tartva. A korpuszépítés során kizárólag a témák meghatározására került sor, amely a beszélgetések során a gyermekek által használt nyelvi eszközöket nem befolyásolta, így a természetes nyelvhasználat jól megfigyelhető benne. A témák meghatározására a gyermekek bevonásával történt, amelyről bővebben a III. 1. fejezetben lesz szó.

Habár különböző értelmezésekben, pedagógiai korpuszokkal többen is foglalkoztak a korpusznyelvészet és a nyelvoktatás területén. Az egyik legismertebb Willis (1998, 2003: 163), aki könyvében azt javasolja, hogy az olyan korpuszra, amelyet a tanulók által ismert szövegekből (tankönyvek, olvasókönyvek) állítunk össze, pedagógiai korpuszként tekintünk. Szerinte a tanári feladatok egyik legfontosabb része lenne a tanulókat megismertetni a pedagógiai korpuszok használatával. Hiszen ahogy a nyelvészek és lexikográfusok is korpuszt vizsgálva írják le a nyelvet vagy készítenek szótárakat, hozzájuk hasonlóan a nyelvtanulók is fejleszthetnék a saját nyelvtudásukat, szókincsüket, grammatikai ismereteiket egy-egy ilyen korpusz használatával (Willis – Willis, 2007: 187–189). A kutatási célra összeállított korpuszokat is használhatnák a nyelvtanulók erre a célra, de sokkal hasznosabb, ha már ismert szövegekkel találkoznak újra és újra (vö. II.2. és II.3. fejezet). A tankönyvi szövegek korpusz formájában történő bemutatása, a szoftverek önálló használata, arra ösztönözné a tanulókat, hogy olyan nyelvhasználati mintázatokat is megfigyeljenek, amelyek a tanórán talán észrevétlenek maradnának (Hunston, 2022: 26). Az ilyen típusú korpuszok használata hasznos, mert a nyelvtanulók a tanult szavakat kontextusban maguk fedezhetik fel, és használatuk közben játékosan gyakorolhatnak (Szirmai, 2005: 35). Ugyanakkor meg kell jegyeznünk, hogy



a tankönyvi szövegekből készült pedagógiai korpuszok használata a nyelvtanórán akkor válik igazán hasznossá, ha az adott tankönyv korpuszinformált, vagyis a természetes nyelvhasználatot tükrözi (ld. II.6.5.1 fejezet). Az ilyen pedagógiai korpuszok építésével kapcsolatban Szita (2021: 73) négy alapelv követését ítéli elengedhetetlennek. Az első, hogy a szövegek természetes nyelvhasználatot reprezentálják (alacsonyabb nyelvi szinteken néhol rövidítve és adaptálva). Visszautalva az előzőekben bemutatott lexikai előfeszítés elméletére, Hoey is azt tanácsolja a nyelvtanároknak, hogy lehetőség szerint autentikus adatokat használjanak a nyelvtanítás során (II.3.4 fejezet). Erre azért van szükség, mert a korpuszok megfelelő használata elősegíti a lexikai előfeszítés jelenségének fejlődését. A második alapelv, hogy a nyelvtanulók számára releváns témákat és szövegtípusokat foglaljanak magukban. A harmadik, hogy a szövegek nyelvezete illeszkedjen megfelelően a tanuló nyelvi szintjéhez. A negyedik pedig, hogy a méretét tekintve kellő számú megnyilvánulást tartalmazzon egy-egy témához tartozó jelenség megfigyeléséhez (Szita, 2021: 73).

Willis értelmezése mellett a pedagógiai korpuszok fogalmának más meghatározásai is születtek. Az egyik ilyen, amely szerint nem a tankönyvi szövegek gyűjteményei ezek a korpuszok, hanem sokkal inkább „témák szerint rendezett korpuszok, amelyek inkább pedagógiai, mint nyelvi reprezentativitásra törekszenek, és megkérdőjelezzik a hagyományos korpuszkeresési folyamatokat” (Crosthwaite, 2019: 69). Ezenkívül pedagógiai korpusznak nevezik azokat a videó- és hanggyűjteményeket is, amelyeket a nyelvtanulók közvetlenül használhatnak a nyelvtanulásuk során (pl. Widmann – Kohn – Ziai, 2011).

### **6.3.2 Gyermek számára készült pedagógiai korpuszok**

Crosthwaite (2019) könyvében említi, hogy a gyermekek adataalapú módszer [Data-driven learning – DDL] általi nyelvtanítását még nem sokan alkalmazzák és még csak néhányan vizsgálták. Ennek főleg az az oka, hogy az adataalapú nyelvtanuláshoz szükséges számítógépes programoknak a fiatalabb korosztállyal történő megismertetését túl nagy feladatnak találják a nyelvtanárok.

Sinclair tervezte egy nyelvtanórán használható pedagógiai korpusz készítését általános iskolás korú Skóciában tanuló gyermekek számára, amelyet 2007-es halála miatt már nem tudott megvalósítani. A *PhraseBox* elnevezésű projektről a következőket írta a *West Word* című helyi újságban:

„...olyan mintha egy óriási, rendkívül jól rendezett, egyszerre elérhető emlékgyűjteményt adnánk minden egyes gyermeknek, amely annak a gyűjteménye, ahogyan sok éven keresztül emberek ezrei fejezték ki magukat. A gyermekek fokozatosan elsajátítják azt, amire szükségük van ebből a forrásból, amely által magabiztosabban tudják majd kifejezni magukat, ugyanakkor ez a forrás mindig elérhető lesz számukra, amikor csak szükségük van rá” (Sinclair, 2006).

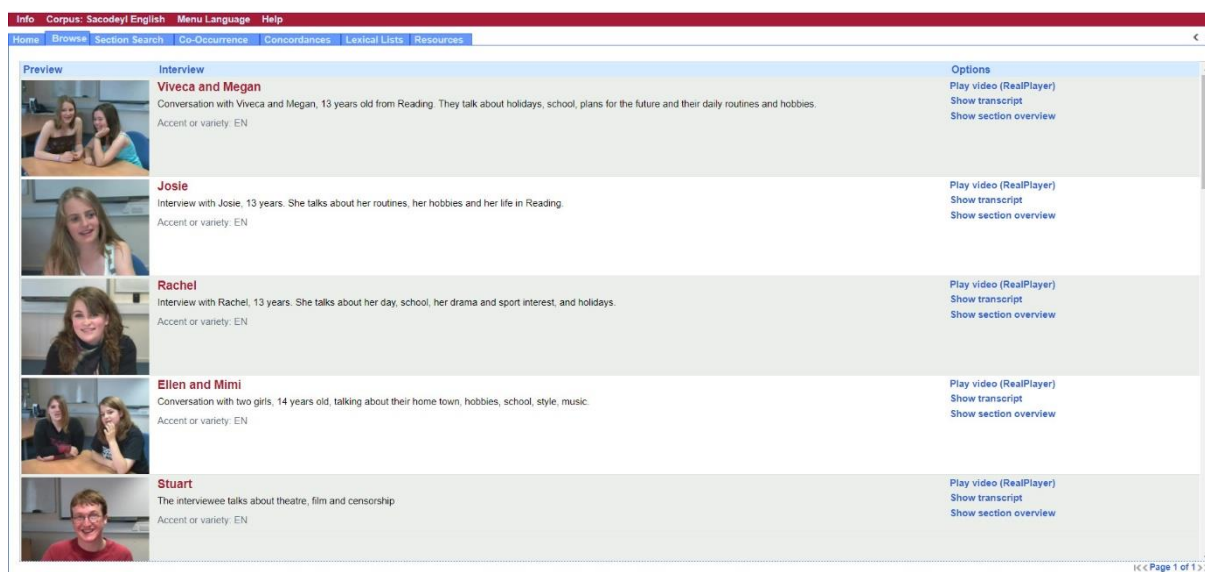
Később mások is próbálkoztak ilyen korpuszok összeállításával, elsősorban tinédzserkorú gyermekek számára. Az egyik ilyen próbálkozás az ELISA, vagyis a második nyelven használható angol nyelvű interjúk korpusza [English Language Interview Corpus as a Second-Language Application] (Braun, 2007). Ez a korpusz 25 videóinterjúból állt, amelyet angol anyanyelvi beszélőkkel készítettek. Braun (2007: 322) huszonhat 14–15 éves angol nyelvet tanuló diákkal, egy hónapig próbálta ki a korpusz használatát és megállapította, hogy a hozzá készített korpuszinformált feladatok hatékonyak bizonyultak.

Egy másik, a SACODEYL, azaz az európai ifjúsági nyelv rendszersegített összeállítása és nyílt terjesztése [System-Aided Compilation and Open Distribution of European Youth Language] elnevezésű gyermekek számára készült pedagógiai korpuszt az Európai Unió Minerva projektjének keretén belül hoztak létre 2005 és 2008 között. A korpusz azzal a céllal jött létre, hogy a középiskolai nyelvoktatásban használható korpuszok hiányát pótolják. Ahogy a projekt weboldalán olvasható „a SACODEYL pedagógiai közvetítőnek tekinti magát a fiatal európaiak nyelvtanulási folyamatában, amely az IKT-forrásokat felhasználva biztosítja a lehetőséget a konstruktivista megközelítésen alapuló adatvezérelt nyelvelsajátításhoz” (SACODEYL, 2008).

A SACODEYL egy hét részből álló, témák szerint összeállított korpusz, amelyben az angol, francia, német, olasz, litván, román és spanyol nyelvek jelennek meg (Braun, 2007). A korpusz kialakítása Braun (2005) javaslatait követte, aki szerint ez a nyelvtanulásban betöltött közvetítő szerep úgy valósítható meg, ha az adott pedagógiai korpusz releváns tartalommal bír, korlátozott a korpusz mérete, így nem veszünk el benne tanulás közben, valamint a könnyebb kezelhetőség érdekében multimédiás formátummal és pedagógiai annotációval rendelkezik. A SACODEYL angol része például ötvenhárom ezer szót és huszonegy videóinterjút tartalmaz. A benne található videófelvételeken tinédzserek beszélnek az érdeklődési körükről, tapasztalataikról, barátaikról és családjukról (Thompson, 2010: 4). Ezek a videók többféleképpen is hasznosíthatóak a nyelvórákon, jól használható a hallott szövegértés fejlesztésére, mert a videókhoz készültek feladatok és leiratok, amellyel a tanulók

ellenőrizhették a megértésüket. A videók egyben vagy a nyelvórákhoz és a pedagógiai célokhoz igazodva több kisebb részben is lejátszhatóak (Braun, 2005). Ezenkívül a weboldal tartalmaz egy könnyen használható keresőt is, amely szintén megkönnyíti a tanulók és tanáraik munkáját. A témák szerinti annotáció mellett, a korpusz grammatikai jelenségek, szófajok szerint is annotálva van és akár konkordancia kereséseket is lehet rajta végezni (Thompson, 2010). A SACODEYL korpuszt nem azért készítették el, hogy azzal a tinédzserek beszédét reprezentálják, hanem azért, hogy az interjúalanyok általi nyelvhasználatot más az adott nyelvet tanuló gyermekek számára modellként, forrásként biztosítsák (Pérez-Paredes – Alcaraz, 2009).

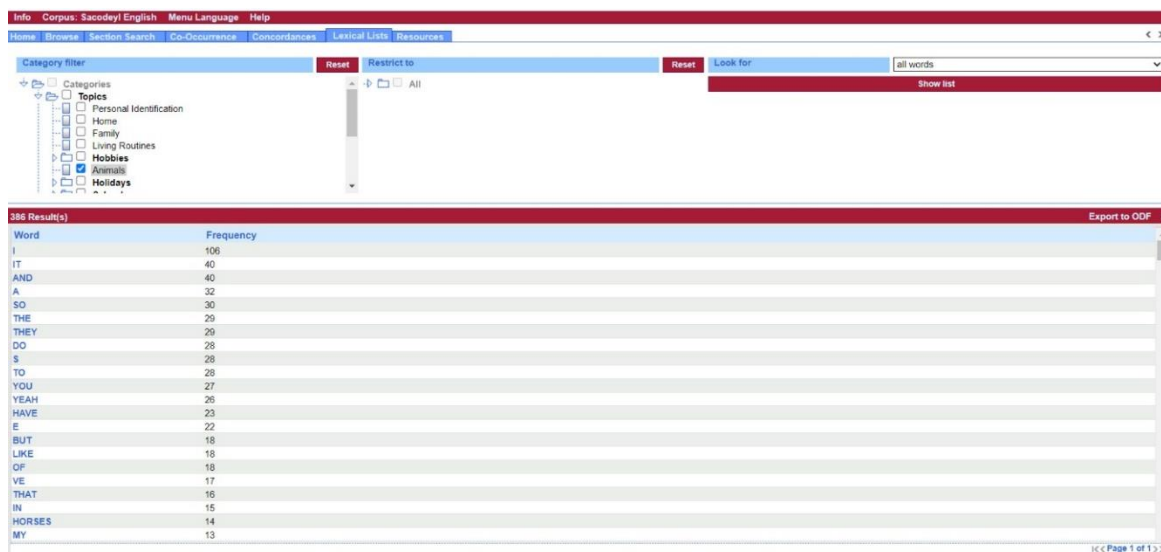
9. ábra A SACODEYL angol nyelvű alkorpuszának kezdőképernyője



Ehhez a korpuszhoz hasonló a szintén Európai Unió projekt során létrejött BACKBONE korpusz, amelyet 2009 és 2011 között építettek és a tartalomalapú nyelvoktatás [Content and Language Integrated Learning – CLIL] támogatása volt a célja. Ebben az úgynevezett élethosszig tartó tanulás [Lifelong Learning] projektben felnőttekkel készítettek interjúkat hét nyelven (angol, francia, német, lengyel, spanyol, török). Hasonlóan a SACODEYL korpuszhoz, ehhez is készültek feladatlapok, illetve tanulást segítő leiratok. Sajnos a feladatlapok ma már egyik korpuszhoz sem érhetőek el. A BACKBONE korpuszban két fájl formátumban voltak elérhetőek a videóinterjúk, amelyek közül ma már csak a Windows Média fájlok játszhatóak le, a Real Video formátumúak már nem. Jó hír azonban, hogy bár az eredeti weboldalaikon már nem, a SACODEYL egy részének leiratai (angol, francia, és spanyol gyermekekkel készült interjúk) és a teljes BACKBONE korpusz keresőeszközökkel együtt

megtalálható a Tübingeni Egyetem weboldalán<sup>12</sup>.

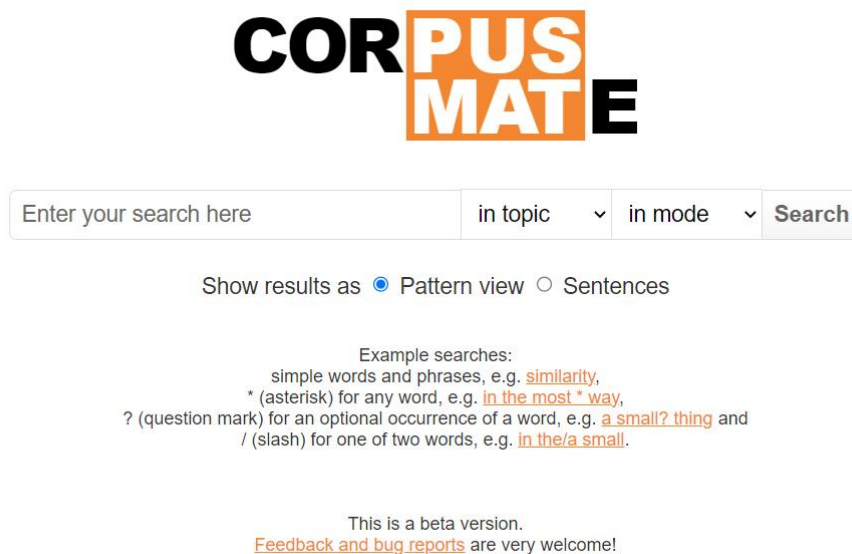
### 10. ábra A SACODEYL keresője: a leggyakoribb szavak az állatok témakörében



| Word   | Frequency |
|--------|-----------|
| I      | 106       |
| IT     | 40        |
| AND    | 40        |
| A      | 32        |
| SO     | 30        |
| THE    | 29        |
| THEY   | 29        |
| DO     | 28        |
| S      | 28        |
| TO     | 28        |
| YOU    | 27        |
| YEAH   | 26        |
| HAVE   | 23        |
| E      | 22        |
| BUT    | 18        |
| LIKE   | 18        |
| OF     | 18        |
| VE     | 17        |
| THAT   | 16        |
| IN     | 15        |
| HORSES | 14        |
| MY     | 13        |

Végül pedig egy éppen készülő projektet is érdemes megemlítenünk a gyermekeknek szóló pedagógiai korpuszok kapcsán. A *CorpusMate*<sup>13</sup> nevű felületet Peter Crosthwaite (University of Queensland) és Vít Baisa fejleszti (a SkELL korpusz platform készítője is, amelyről a 6.5 fejezetben lesz szó bővebben).

### 11. ábra A CorpusMate platform kezdőlapja



**CORPUSMATE**

Enter your search here    in topic    in mode    Search

Show results as  Pattern view  Sentences

Example searches:  
simple words and phrases, e.g. [similarity](#),  
\* (asterisk) for any word, e.g. [in the most \\* way](#),  
? (question mark) for an optional occurrence of a word, e.g. [a small? thing](#) and  
/ (slash) for one of two words, e.g. [in the/a small](#).

This is a beta version.  
[Feedback and bug reports](#) are very welcome!

<sup>12</sup> <http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp>

<sup>13</sup> <https://corpusmate.baisa.cz/>

A *CorpusMate* egyszerűsített nyelvi adatelemzést biztosít középiskolás angolt mint második nyelvet tanuló gyermekek számára. A projekt célja a jelenleg elérhető legjobb korpuszeszközök legjobb tulajdonságainak integrálása egy könnyen kezelhető digitális környezetbe. A felületen elérhető korpusz szövegei kapcsolódnak az középiskolai és egyetemi tantárgyakhoz, és hat különböző forrásból származnak:

1. BAWE (egyetemi hallgatók írásainak gyűjteménye),
2. TED Talks (TED előadások leiratai),
3. Egyszerű angol Wikipédia (előre meghatározott, a tantárgyakhoz kapcsolódó kulcsszavak alapján kiválasztott szövegek)
4. BBC Teach (a BBC weboldalára nyelvtanításhoz készült videók leiratai)
5. Elsevier OA CC-BY (az Elsevier folyóirataiból származó, a tantárgyakhoz kapcsolódó 40 000 tudományos cikk),
6. BNC 2014 Spoken (a British National Corpus beszélt nyelvi része, 11,5 millió szó)

A szövegeket normalizálták, tokenizálták és szófajok szerint címkézettek. Egy speciális eszközzel ellenőrizték a szövegeket, eltávolították a trágár szóhasználatot, valamint az összes sértő mondatot. Az elemzéshez a korpuszba bevitt szövegeket előzetesen meghatározott témákhoz, vagyis a tantárgyakhoz rendelték, amelyekhez kulcsszavakat használtak, és amelyek alapján a keresés során szűkíthetjük a találatokat (a 11. ábrán a témakörökben való keresés funkció [in topic]). A témák szerinti keresés mellett, lehetőség van külön az írott és külön a beszélt nyelvi szövegekben történő keresésre is (a 11. ábrán a módokban történő keresési funkció [in mode]). A folyamatban lévő projekt célja, hogy egy hatékonyabb és célzottabb nyelvtanítási eszközt biztosítson a fiatal nyelvtanulók számára (CorpusMate, 2023).

A fentieken kívül gyermekek számára készült és még elérhető pedagógiai korpuszról nincs tudomásom. Kevesen foglalkoznak ezzel a korosztállyal, amelynek okai talán azok a nehézségek, amelyekkel szembe kell néznünk, amikor kiskorúakkal dolgozunk. Ilyen a személyes adatok kezelése, a szülőktől való engedélykérések és a felvételek megszervezése, valamint a korpuszok pedagógiai felhasználásával kapcsolatban már említett technikai ismeretek hiánya.

### **6.3.3 Korpusznyelvészeti kutatások gyermekek számára készült pedagógiai korpuszon**

Habár a SACODEYL – a KorSzak Gyermeeknyelvi Korpuszal ellentétben – alapvetően nem

abból a célból jött létre, hogy korpusznyelvészeti kutatásokat végezzenek rajta, több vizsgálatra is használták, amelyek közül kettőt, Buntinx és Van Goethem, (2018), valamint Vanderweerd és Keijzer (2008) kutatását emelem ki és mutatom be ebben a részben.

Buntinx és Van Goethem (2018) kutatásában a melléknévfokozást vizsgálta meg kontrasztív nyelvészeti keretben francia anyanyelvű angolul és hollandul tanuló diákok nyelvhasználatában. A vizsgálatuk egyik referenciakorpuszaként szintén a SACODEYL korpuszt használták, mert a második nyelvként angolul beszélő franciák nyelvhasználatát az abban található angol anyanyelvű gyermekek nyelvhasználatához hasonlították. A francia anyanyelvű, hollandul tanuló diákok esetében arra az eredményre jutottak, hogy a tanulók kevésbé ismerik a holland ifjúsági nyelvben használt informálisabb és kifejezőbb fokozott melléknéveket. Az angolt második nyelvként beszélőknél viszont a kapott eredmények hasonlóak voltak az angol anyanyelvű beszélőknél tapasztaltakhoz, amely a célnyelv nagyobb befolyására utalhat. Összehasonlították a CLIL és nem CLIL módszerrel tanulók teljesítményét is, ahol nem mutatkoznak jelentős különbségek a fokozás formai és szemantikai típusai tekintetében. Ugyanakkor kimutatták, hogy a CLIL módszerrel tanulók a határozószók változatosabb választékát használták.

A másik vizsgálat a jelen értekezésben bemutatott kutatás szempontjából még az előzőnél is érdekesebb. Vanderweerd és Keijzer (2008) a SACODEYL francia részét – amelynek a mérete körülbelül 45 ezer szó és 53 ezer tokent tartalmaz - használta arra, hogy megvizsgálják a gyermekek interjúiban lévő formulaszerű nyelvhasználatot, majd a talált lexikai egységeket összevessék a kezdő francia mint idegen nyelv könyvekben találhatóakkal, hogy megtudják, mennyire reprezentálják a könyvekben található dialógusok a természetes nyelvhasználatot. Ezt azért tudták megtenni, mert a SACODEYL francia nyelvű alkorpuszában található 22 interjú és 2 csoportos beszélgetés tematikájában megfelelnek a KER szerinti A1–B1 nyelvi szinteknek. A lexikai egységek kinyerésére a CLAN szoftvert (MacWhinney, 2012) alkalmazták. A korpuszban összesen 103.292 négyszavas lexikai egységet találtak, amelyek a teljes beszélt nyelvi korpusz 77,72%-át tették ki. A leggyakoribb 103 lexikai egységet négy funkció szerint kategorizálták (Biber, 2006 alapján): referenciális, diskurzus-szervező, álláspont kifejezése, és kommunikatív, ahogy ez a jelen kutatásban is történt (ld. III.4. fejezet) Ezt vetették össze a vizsgált tankönyvekben található szövegekkel. Eredményeik alapján kimutatták, hogy a gyermekek nyelvhasználatában szignifikánsan több lexikai egység volt megtalálható, mint a tankönyvekben, így arra a következtetésre jutottak, hogy azok nem reprezentálják jól az autentikus nyelvhasználatot.

## 6.4 A korpusznyelvészet módszer vagy elmélet?

A korpusznyelvészeti kutatásokkal kapcsolatban meg kell említenem, hogy néhány szakirodalom (vö. Tognini – Bonelli, 2001; Teubert, 2005) különbséget tesz az úgynevezett korpuszalapú [corpus-based] és a korpuszvezérelt [corpus-driven] megközelítések között. Ezt a kettősséget részletesen tárgyalja könyvében McEnery és Hardie (2012: 5–6; 147–152). Tognini – Bonelli (2001: 65–66) úgy határozta meg, hogy a korpuszalapú megközelítést használó kutatások egy teóriát tesztelnek az adott fókuszkorpuszon [top-down], amíg a korpuszvezérelt megközelítést választó kutatások készítői a korpuszt veszik alapul és annak vizsgálatából vonnak le olyan következtetéseket, illetve alkotnak hipotéziseket, amelyekből végül általánosítás által megszületik az új elmélet [bottom-up] (2001: 84-85). Teubert (2005: 2) a korpusznyelvészetet egyenesen a nyelv tanulmányozásának egyik elméleteként említi. Ahogy Stubbs (1997: 301) is írja, „a korpusz nem csak eszköz, hanem egy jelentős nyelvelméleti fogalom is”. A két fogalom többször keveredik a szakirodalomban, a korpuszalapú kutatásokban a korpusz szerepe nagyon változó, vannak olyan kutatások, amelyekben jobban, amíg másokban kevésbé támaszkodnak a korpuszokból nyert információkra (McEnery – Hardie, 2012: 151). A korpuszvezérelt megközelítés elveti a korpusznyelvészet kialakulása és elterjedése előtt megalkotott elméleteket, mert követői szerint ez elvonhatja a figyelmet a korpuszban található adatokból kikövetkeztethető valódi elméletekről (Sinclair, 2004: 191). Ezzel szemben a korpusznyelvészetet módszerként értelmezők a korpuszokat olyan empirikus adatok forrásának tekintik, amelyek (az elmélet kidolgozásának módjától függetlenül) bármely nyelvelmélet alátámasztására vagy megdöntésére szolgálhatnak (McEnery – Hardie 2012: 148). A korpuszvezérelt megközelítést képviselők, Sinclair (2004) „trust the text”, vagyis „bízz a szövegben” mottóját sajátjuknak tekintik. Sinclair és követői három pontra hívják fel a figyelmet. Az első, hogy a korpuszépítés közben hasznos teljes szövegekkel dolgozni, ahogy Sinclair írja: „a teljes dokumentumokból felépített korpusz a nyelvészeti vizsgálatok szélesebb körét teszi lehetővé, mint a rövid minták gyűjteményei” (1991: 19). Sinclairnek ez azért volt fontos, hogy a teljes diskurzust meg lehessen vizsgálni egy-egy dokumentumban, ugyanakkor meg kell jegyezni, hogy a vizsgálatok során a korpuszeszközök nem a szöveg részeinek egymásutániségát nézik, hanem összeolvasztva, a teljes korpuszban vizsgálják meg őket (McEnery – Hardie, 2012: 153). Itt meg kell jegyezni, hogy beszélt nyelvi korpuszok építése során szinte lehetetlen csak teljes szövegekkel dolgozni. Ha erre törekednénk, azzal kockáztatnánk a természetes nyelvhasználat megfigyelésének lehetőségét, hiszen így a szövegek szerkesztettek lennének.

A második pont, amire felhívják a figyelmet az a korpusz nyelvészeti annotációjának elkerülése, amely ma már nem releváns. Ahogy McEnery és Hardie is megemlíti a mai korpuszeszközökkel már nem kell attól félni, hogy a címkék [tags] problémát jelentenek (2012: 154). A jelen kutatás során használt Sketch Engine ráadásul automatikusan annotálja a korpuszokat, amelyre a III. 2.1 fejezetben térünk ki részletesebben.

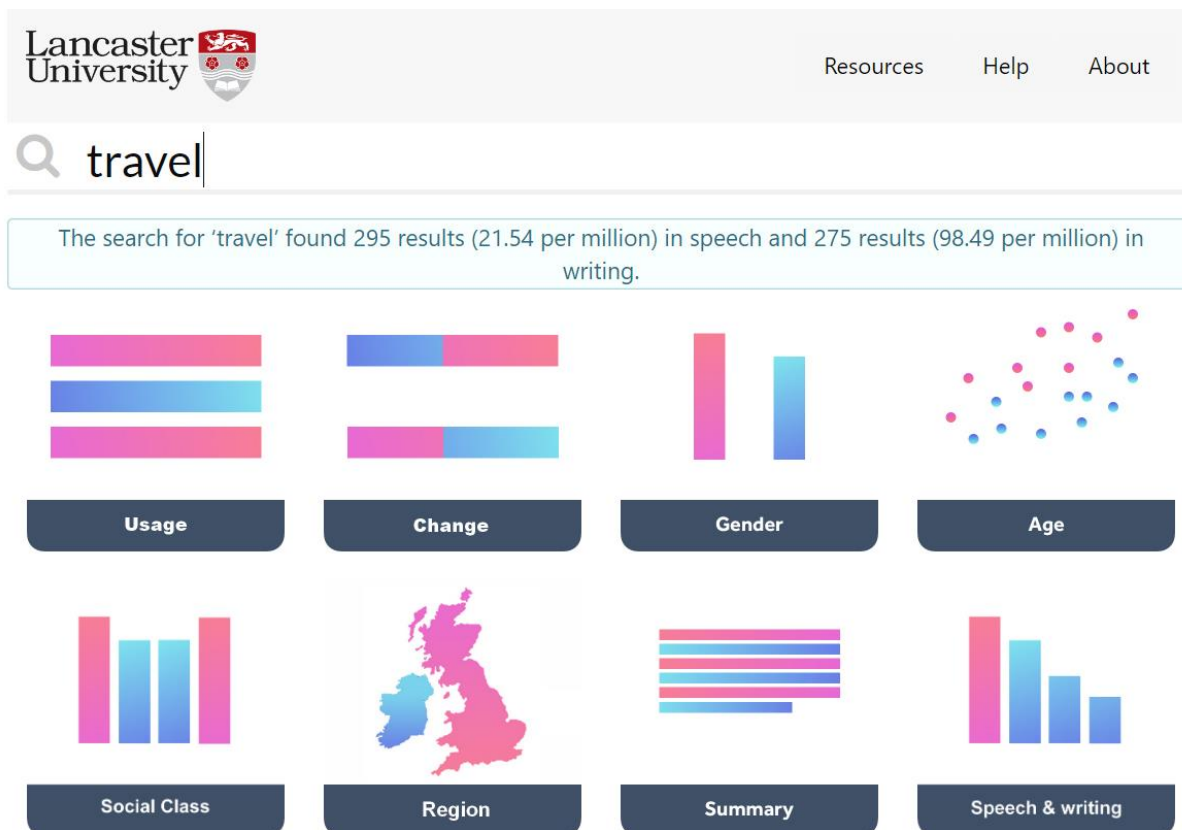
A harmadik pont, amelyet megemlítenek, a korpusznyelvészet megjelenése előtti elméletek használatának elvetése. Véleményem szerint nem szabad teljesen elvetni a korábbi elméleteket és ahogy McEnery és Hardie, én sem értek egyet azzal, hogy teljesen szétválasszuk egymástól a korpuszalapú és korpuszvezérelt megközelítést. Egy korpusznyelvészeti kutatás során mindkét megközelítés fontos lehet. Egy korpuszalapú megközelítést használó kutatásban ugyan egy már meglévő elméletet próbál bizonyítani a kutató, de ezt az elméletet meg tudja változtatni, igazítani tud rajta az elemzés során kapott eredmények felhasználásával (McEnery – Hardie, 2012: 150). Ezért szerintem is érdemes a korpuszalapú megközelítés fogalmának meghatározását kitágítani és ezt használni minden korpusznyelvészeti kutatásra, ugyanakkor természetesen tisztában kell lennünk vele, hogy ezek a kutatások többféleképpen közelíthetik meg a korpuszt.

## **6.5 A korpusznyelvészet alkalmazása a nyelvoktatásban**

A nyelvtanárok a korpuszokat kétféleképpen alkalmazhatják az órán: egyrészt indirekt módon, másrészt pedig direkt módon. Az indirekt mód a korpuszinformált tananyagok, szótárak használata, amelyekről még bővebben lesz szó a későbbiekben (II.6.5.1–II.6.5.2 fejezet), amíg a direkt módon történő alkalmazás, amikor magát a korpuszt használják a diákok, vagy abból származó részleteket, például egy konkordanciasort visz be a tanár az adatalapú módszert alkalmazó nyelvórára. Erre a célra használhatunk egy általános korpuszt is mint a BNC vagy magyar példát említve az MNSZ2, illetve használhatjuk a szándékosan erre a célra létrehozott pedagógiai korpuszokat is. Annak ellenére, hogy jól látható a pedagógiai korpuszok használatának jelentősége, a nyelvoktatásban sok helyen „nem terjedt el, mert a forrásként használható korpuszok többsége sem nem pedagógiailag orientált, se nem felhasználóbarát” idézi Frankenberg – Garciát Crosthwaite (2019: 69). Véleményem szerintem ehhez még az is hozzájárul, hogy nem minden nyelvtanár kapott olyan képzést, amelyben a korpuszok használatának lehetőségeiről szó volt, például arról, hogyan készítsünk megfelelően hatékony feladatokat az órára.



12. ábra A BNClab keresőfelületének kezdőlapja



A korpuszok nyelvtanuláshoz történő optimalizálására jó kezdeményezés a Lancaster University által létrehozott BNClab<sup>14</sup>, amely egy egyszerűsített keresőplatform (12. ábra). A BNClab-hez tartozó két korpusz a BNC és a BNC2014 összesen 10 millió szavát tartalmazza. A felületet használva a nyelvtanulók felfedezhetik a jelenleg beszélt brit angol nyelvi mintákat, és megvizsgálhatják, hogy az olyan társadalmi tényezők, mint a nem, az életkor, a társadalmi osztály és a régió hogyan befolyásolják a nyelvhasználatot (Corpus for Schools, é.n.). A statisztikai adatokon és jól áttekinthető vizualizációkon kívül a BNClab segítségével konkordanciasorokat is készíthetünk (13. ábra), amelyet szűrhetünk név, nemi hovatartozás, életkor, szociális helyzet, régió és korpusz szerint is.

<sup>14</sup> <http://corpora.lancs.ac.uk/bnclab/search>

### 13. ábra Konkordanciasor a BNCLab felületén

Display info about speakers  
 Name  Gender  Age  Social Class  Region  Corpus

| Index | Left   | Node   | Right  |
|-------|--|--------|--|
| 1     | once S: it was actually a few pounds cheaper S: yeah S: to | travel | first class than second S: mm do n't know how they         |
| 2     | experience no S: of the erm S: no I do n't S: rail         | travel | S: no S: and the places that you see which presumably will |
| 3     | it S: was for the travel industry had been in the          | travel | industry since before he met my sister and they met        |
| 4     | of like and it was just to catch onto S: free              | travel | S: if you were called and it was just like S: or           |
| 5     | not like a travel company ? S: erm as like a               | travel | company S: mm S: it will be that you just book a           |
| 6     | 's S: not the the world 's best road to S: no              | travel | on S: no it is n't no S: and we got that                   |
| 7     | a S: little bit later because the train 's got to          | travel | from London to Ashford S: oh yeah yeah S: so that just     |
| 8     | and S: then they can like and all they do is               | travel | around the country and then if they need to they           |

Egy másik hasonló projekt, a SkELL<sup>15</sup> [Sketch Engine for Language Learning], amely jelenleg hat nyelven (angol, cseh, észt német, olasz, orosz) érhető el. Ez a Sketch Engine korpuszeszközök tartalmazó felület egyszerűsített, ingyenes eszköze, ahol a tanulók és nyelvtanárok könnyen ellenőrizhetik, hogy egy adott kifejezést vagy szót a célnyelv valódi beszélői használnak-e, illetve, ha igen, hogyan használnak (SkELL, é.n.). A SkELL alapjául szolgáló korpusz nem teljes dokumentumokat tartalmaz, hanem úgynevezett jó szótári példákat [Good Dictionary Examples – GDEX], vagyis olyan kiválasztott mondatokat, amelyeket az algoritmus bizonyos szempontok alapján megfelelőnek értékelt (a nyelvhasználati példákat hosszuk, lexikájuk bonyolultsága, ellentmondásos témakörei kizárhatják a jó szótári példák közül). A könnyen kezelhető felületen jelenleg három korpuszeszközt használhatunk: megnézhetjük egy keresett szó vagy kifejezése előbb említett jó szótári példamondatait (14. ábra), hogy milyen szavakkal együtt és milyen grammatikai környezetben jelenik meg (15. ábra), valamint mik a keresett szóhoz hasonló szavak (16. ábra), amelynek találati eredményeit szemléletes szófelhő formájában is megjeleníti az eszköz.

<sup>15</sup> <https://skell.sketchengine.eu/#home?lang=en>

## 14. ábra Jó szótári példamondatok [GDEX] a SkELL felületén

travel 162.57 hits per million

1. Our current retail **travel** packages are listed below.
2. His thoughts were therefore turned upon **travelling**.
3. **Travel** is restricted through inland border checkpoints.
4. Region trusts are lastly **travelled** existing various illness significant areas.
5. British army trains **travelled** nine million track miles every month.
6. A humble rat peasant **traveling** between fields.
7. This facilitates higher **travel** speeds while protecting against atmospheric contamination.
8. That memory suggests future air **travel** possibilities.
9. Most major **travel** insurers offer such policies.
10. Official air **travel** is normally accomplished using commercial air transportation.
11. New technologies allow extremely fast underwater **travel**.
12. The occupant thus faced backwards during **travel**.

## 15. ábra A szavak kollokációinak és grammatikai mintázatainak keresése a SkELL felületén

travel verb ✓ Show context

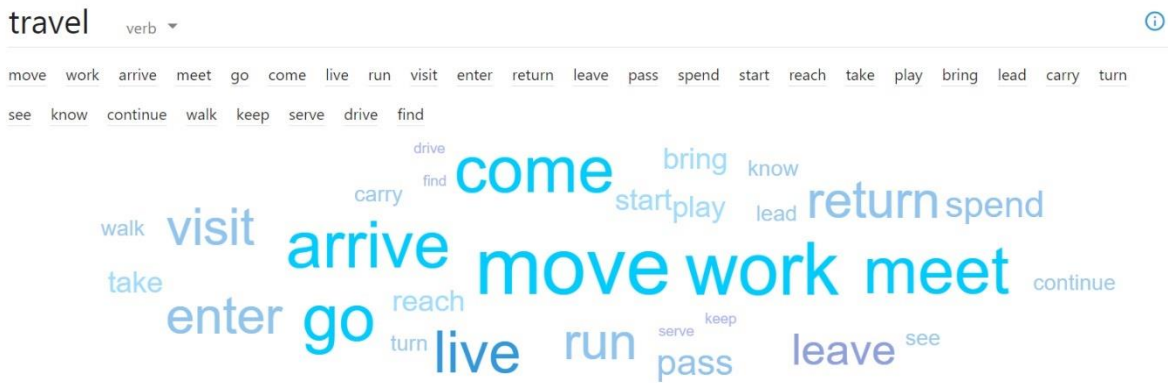
| subject of travel                         | object of travel                          | phrasal                             | phrasal with object                   |
|---|---|-------------------------------------|---------------------------------------|
| 1. <b>passenger</b> passengers travelling | 1. <b>distance</b> travel long distances  | 1. <b>through</b> to travel through | 1. <b>around</b> to travel around the |
| 2. <b>distance</b> the distance traveled  | 2. <b>mile</b> miles traveled             | 2. <b>around</b> travel around      | 2. <b>down</b> travel down            |
| 3. <b>train</b> train travelling          | 3. <b>companion</b> traveling companion   | 3. <b>across</b> to travel across   | 3. <b>along</b> travel along the      |
| 4. <b>wave</b> waves travel               | 4. <b>salesman</b> a traveling salesman   | 4. <b>along</b> travel along        | 4. <b>over</b> traveling over         |
| 5. <b>vehicle</b> vehicles traveling      | 5. <b>exhibition</b> traveling exhibition | 5. <b>in</b> travel in and out of   | 5. <b>up</b> travel up                |
| 6. <b>bus</b> bus traveling               | 6. <b>world</b> travel the world          | 6. <b>on</b> travel on              | 6. <b>through</b>                     |
| 7. <b>car</b> car travelling              | 7. <b>circus</b> a traveling circus       | 7. <b>down</b> travel down          | 7. <b>unto</b>                        |
| 8. <b>mile</b> vehicle miles traveled     | 8. <b>mi</b> traveled 70 mi               | 8. <b>over</b> travel over          | 8. <b>across</b>                      |
| 9. <b>ship</b> ships traveling            | 9. <b>exhibit</b> traveling exhibit       | 9. <b>up</b> travel up to           | 9. <b>off</b> travel off the          |
| 10. <b>light</b> light travels            | 10. <b>public</b> the traveling public    | 10. <b>away</b> to travel away from | 10. <b>in</b>                         |
| 11. <b>time</b> time traveling            | 11. <b>km</b> km travelled                | 11. <b>out</b> travel out of        | 11. <b>away</b>                       |
| 12. <b>road</b> road less traveled        | 12. <b>globe</b> travel the globe         | 12. <b>upon</b>                     |                                       |
| 13. <b>boat</b> boat travelling           | 13. <b>expense</b> travelling expenses    |                                     |                                       |
| 14. <b>ball</b> the ball travels          | 14. <b>route</b> traveled route           |                                     |                                       |

## 16. ábra A hasonló szavak keresése a SkELL felületén

travel noun

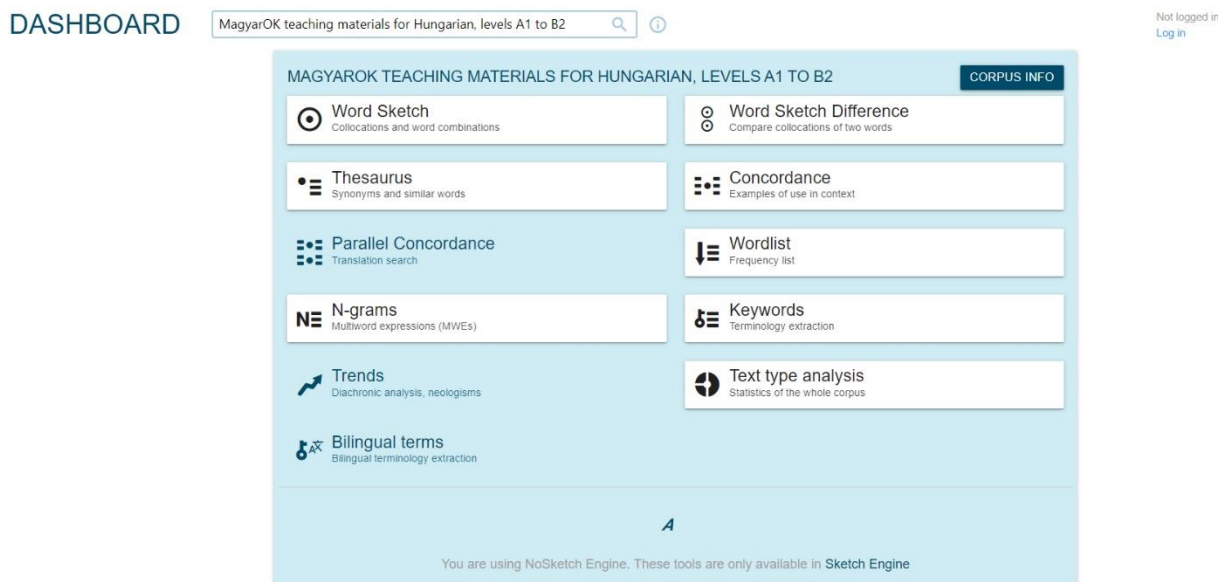
service transportation business development activity training education trip access work health trade research care cost information  
 transport use communication life time job technology sale food project resource experience program production

production program resource  
 project sale research development trip trade education  
 care  
 transport communication business service transportation  
 technology activity life training health cost work use time  
 experience information food job access



Sajnos a SkELL magyar nyelven még nem érhető el, de magának a Sketch Engine felületnek – ahogy láthatjuk majd a későbbi bemutatásában (III.2.1 fejezet) – sem olyan bonyolult a használata, hogy a nyelvtanulók egy kis segítséggel ne tudnák megtanulni az alkalmazását. Egy másik lehetőség a magyar nyelv korpusz segítségével történő tanítására a Sketch Engine-en megtalálható a Szita Szilvia – Pelcz Katalin: *MagyarOK* című tankönyvsorozat szövegeiből készült nyílt korpusz (vö. Szita, 2020), amelyet a magyar mint idegen nyelv órákon, de akár a tanulók nyelvórán kívül és a tanárok feladatlapok készítéséhez is használhatnak (Szita – Pelcz, é.n.).

17. ábra A MagyarOK nyílt korpusznak a kezdőoldala a Sketch Engine-en



A fenti ábrán (17. ábra) látható korpuszeszközöknek a használatára a Sketch Engine-t bemutató részben (ld. III.2.1 fejezet) térek majd ki részletesen, de itt mindenképpen szeretném

megjegyezni, hogy amint látható számos lehetőség közül választhatunk a korpusz tanulmányozása közben.

1. Megismerhetjük a szavak és kifejezések kollokációit, a grammatikai környezetét [Word Sketch],
2. nyelvi elemeket hasonlíthatunk össze egymással [Word Sketch Difference],
3. konkordanciasorokat készíthetünk [Concordance], amelyekből megfigyelhetjük a használati mintázatokat,
4. szólistákat [Wordlist],
5. kulcsszavakat [Keywords] és többelemből álló kifejezéseket [N-grams] kereshetünk.

A másik két funkció, a szinonimakeresés [Thesaurus] és a korpusz statisztikája [Text type analysis] nem működik megfelelően, de ez egyáltalán nem jelent problémát, és a fenti felsorolásra tekintve kijelenthetjük, hogy a *MagyarOK* nyílt korpuszának használata során sokkal több korpuszeszköz közül választhatunk, mint a SkELL felületén. A *MagyarOK* ezenfelül azért is hatékonyabb, mert nyelvi szintenként (A1–B2), valamint tankönyvi fejezetekként is kereshetünk benne (Szita, 2020, 2021).

A MagyarOK nyílt korpusza tehát a KER szerinti nyelvi szintek alapján négy részre tagolódik (A1, A2, B1, B2), és háromféle szövegtípust fog tartalmazni. A jelenleg már elérhető részei a tankönyvek szövegeit, valamint anyanyelvi beszélők félautentikus szövegeit tartalmazzák. Ezenkívül hamarosan elérhető lesz egy olyan folyamatosan bővülő alkorpusza is, amely személyes szövegekből, például közösségi oldalakon folytatott beszélgetésekből áll majd (Szita, 2021: 74).

Kezdő nyelvtanulóknál, valamint gyermekek tanításánál talán eleinte megfelelőbb a direkt módszer másik fajtájának, azaz a korpusz egy részének és ahhoz készült feladatlapoknak órán történő használata. A magyar mint idegen nyelv területén – a későbbiekben említésre kerülő korpuszinformált tananyagokat nem számítva – jelenleg nincs arról tudomásom, hogy ilyen jellegű segítség elérhető lenne a tanárok számára. Az angol nyelvterületen érdemes megismerni Ken Lackman<sup>16</sup> munkásságát, aki a tanártovábbképzések mellett, rengeteg ötletet – többek között a korpuszok nyelvórán történő (játékos) használatáról – gyűjtött össze letölthető foglalkoztató könyveiben. Egy másik jelentős kezdeményezés a Lancaster University Corpus for Schools<sup>17</sup> [’korpusz iskolák számára’] projektje, amely során a már említett BNClab segítségével készült és annak használatával kapcsolatos feladatlapokat, valamint tanárok számára készült jegyzeteket osztanak meg a projekt weboldalán. A két felsorolt forráson kívül

---

<sup>16</sup> <http://kenlackman.com/>

<sup>17</sup> <https://wp.lancs.ac.uk/corpusforschools/>

hasonló ötleteket találhatunk több olyan módszertani könyvben is, amelyek a lexikai egységek játékos tanításával foglalkoznak. Ilyen például Leo Selivan *Lexical Grammar: Activities for Teaching Chunks and Exploring Patterns* című angol nyelvű, valamint Eveline Schwartz *55 Spiele mit Chunks* című német nyelvű könyve is.

### 6.5.1 Korpuszinformált tananyagok

A korpuszokat a nyelvtanításban általában a nyelvtanulóknak készült szótárak elkészítéséhez alkalmazták, de ma már egyre több tananyag készül a felhasználásukkal. Ebben a részben a tananyagkészítéssel kapcsolatban a korpuszok felhasználásának jelentőségét és lehetőségeit, valamint egy korpuszinformált tankönyvcsaládot mutatok be röviden.

Ahogy már korábban említésre került a korpusznyelvészeti vizsgálatokról tudjuk, hogy a számítógépek elterjedésével, illetve azok memóriájának növekedésével váltak lehetővé. Ugyanakkor fontos megemlíteni, hogy már a számítógépek megjelenése előttről is származnak korpuszinformált tananyagok (Vermeki, 2022a). A XX. század elején készültek olyan angol nyelvkönyvek, amelyeknek a szerzői leveleket, írott szövegeket gyűjtöttek össze és ezeket használták a tananyagaik forrásaként (Palmer, 1999). A már számítógép segítségével készült tananyagok közül a London – Lund Corpus-t használó, 1985-ben megjelent *Comprehensive Grammar of the English Language* (Quirk et al., 1985) című könyvet tekinthetjük az első grammatikának, amely a korpuszból nyert adatokat szisztematikusan használta. A későbbi a komplexebb számítógépes folyamatok általi elemzés elterjedése Biber (1988) kutatásának köszönhető, aki az angol nyelv variációit vizsgálta beszélt és írott nyelvi műfajokban (Meunier – Reppen, 2015: 501). A nyelvkönyvek közül a *Collins Cobuild English Course* (Willis – Willis, 1988) a COBUILD korpuszra épülő tankönyv az egyik első korpuszinformált tananyag (McCarthy – McCarten, 2022). Meg kell jegyezni azonban, hogy a teljes egészében a korpusznyelvészeti kutatások eredményeire támaszkodó szerzők ebben a könyvben a lexikát is kizárólag gyakorisági vizsgálatok alapján választották ki, amelynek következményeként az egy szemantikai mezőhöz tartozó lexikai elemek a tankönyv különböző fejezeteiben jelentek meg. Így például a színek elnevezéseit ebből a könyvből nem egyszerre tanulhatták meg a nyelvtanulók (Szirmai, 2005: 139). Egy másik érdekes példa Scott Thornbury *Natural Grammar – The keywords of English and how they work* (2004) című könyve, amelyben a szerző gyakorisági vizsgálatok segítségével kiválasztotta a leggyakoribb száz angol szót, amelyekhez grammatikai magyarázatokat, kollokációkat és kifejezéseket köt, majd ezeket különböző feladatok elvégzésén keresztül gyakoroltatja.

## a/an

[indefinite article] used with singular nouns:

- (1) to introduce a new person or thing into a text:  
*An Englishman, a Scotsman and an Australian were on a plane ...*  
(2) to show that a person or thing is a member of a group:  
*Fatima's a Muslim.*  
(3) to talk generally about a class of things: *An octopus has eight legs.*

NOTE: Use *a* in front of a word that starts with a consonant sound, and use *an* in front of a word that starts with a vowel sound:

*a plane; a Muslim; a house; a university student*  
*an octopus; an hour; an umbrella; an MA / ,em'eɪ /*

### Grammar patterns

- 1 NP | + is/was | + a/an | + classifying noun

*My flatmate is a Swede.*  
*Jane Austen was an English writer.*  
*The kangaroo is a marsupial.*

- ▲ for classifying a person or thing, e.g. saying what their nationality or job is

- 2 there | + is/was etc. | + a/an | + NP | (+ adverbial)

*There is a supermarket next to the station.*  
*There was an immediate reaction.*  
*There was a strange smell in the kitchen.*

- ▲ for introducing a new topic, e.g. in a description  
(→ there)

- 3 NP | + has/'s got | + a/an | (+ adjective ) | + NP

*Milan has a Gothic cathedral.*  
*Jan's got a nice smile.*  
*Has Vigo got an airport?*

- ▲ for describing places, people, things

- 4 NP | + is/was | + such | + a/an | + NP

*Kevin is such a hard worker.*  
*It's such a shame.*

- ▲ to emphasize how you feel about something or someone

### Collocations

There are many 'verb + noun' combinations beginning *have a ...* or *take a ...*. For example:  
*have a drink, have a bite (to eat), have a laugh, have/take a look, have a go, have a talk, have a say, have a try, have a ride, have/take a walk, have/take a rest.*

*Can you take a look at my homework?*  
*We stopped at a roadside café and had a bite to eat.*  
*I had a go on Robin's new motorbike.*  
*Everyone should have a say in the peace process.*

### Set phrases

- a few / a little / a lot (of) etc. → little

*We had a few friends round for dinner.*  
*There's a little salad left, if you'd like some.*  
*How much time have we got? ~ A lot.*

- ▲ for talking about quantities

- a kilo / a metre / a dozen etc.

*Petrol costs 3 euros a litre.*  
*How much are the eggs? ~ £1.50 a dozen.*

- ▲ for talking about prices

- a coffee / a fruit juice / a Coke etc.

*Would you like a coffee?*  
*He bought me a tomato juice and ordered a Scotch for himself.*

- ▲ to refer to a cup, glass, bottle, or can, of a certain drink

- a hundred / a thousand / a million etc.

*A hundred and twenty people attended the meeting.*  
*There were more than a million viewers.*

- ▲ a less emphatic form of one

- a year / a week / an hour etc.

*She checks her e-mail three or four times a day.*  
*I go to the gym once a week.*

*They get paid less than ten dollars an hour.*

- ▲ to talk about how often something happens, or the rate of something (Here *a/an* means 'every')

Az ilyen típusú tananyagokat a korpuszsal való kapcsolatuk alapján korpuszalapúaknak [corpus-based teaching materials] vagy korpuszinformáltaknak [corpus-informed teaching materials] nevezik. Ezt a két fogalmat sokszor egymás szinonimájaként is szokták használni a szakirodalomban. A többféle elnevezés azért terjedhetett el, mert a tananyagok szerzői különböző módokon használták a korpuszokat és más mértékben támaszkodnak a vizsgálatuk által nyert adatokra. Meunier és Reppen (2015) meghatározása szerint a pedagógiai célra készült anyagokat egységesen korpuszinformáltak kellene neveznünk, amíg – ahogy korábban már említettem (ld. II.6.4 fejezet) – a korpusznyelvészeti kutatások lehetnek korpuszalapúak [corpus-based] vagy korpuszvezéreltek [corpus-driven], illetve ezek kombinációi. A tananyagok korpuszinformált megnevezését azért tartják megfelelőbbnek, mert a tananyagkészítés során több elem ötvözése történik. A több korpuszon végzett kutatások

eredményeinek beépítése mellett, a tartalom kiválasztására (szókincs, nyelvi szerkezetek és mintázatok), a megfelelő szövegek kiválasztására, valamint a korpuszból nyert információk alapján különböző döntések meghozatalára is sor kerül (Meunier – Reppen, 2015: 499). Az általuk említett indokok miatt a korpuszinformált elnevezést használom én is a továbbiakban.

### **6.5.1.1 A korpuszok felhasználási lehetőségei a tananyagok írásánál**

Ahogy az előző részben bemutattam, a korpuszhasználat jelentőségének felismerése következtében ma már egyre több korpuszinformált tananyag jelenik meg a nyelvkönyvek között. Bármilyen nyelvről is legyen szó, azt kijelenthetjük, hogy a korpuszok forrásként való alkalmazása több okból is fontos a tananyagírás során (McCarthy – McCarten, 2022). Az egyik legfontosabb, hogy a kompetens nyelvhasználóvá váláshoz elengedhetetlen, hogy a tanulók megismerjék azokat a kifejezésmódokat, amelyeket az anyanyelvi beszélők alkalmaznak hétköznapi szituációkban (Szita – Pelcz, 2017: 263). Az anyanyelvi beszélők természetes nyelvhasználata pedig beszélt és írott nyelvi korpuszokra támaszkodva könnyen figyelhető meg (Conrad, 2000: 548).

1990 és 2010 között számos kutatás született (ld. Biber – Conrad – Reppen, 1998; Harwood, 2005), amelyekben több nyelvkönyvet vetettek össze a kurrens korpusznyelvészeti kutatások eredményeivel, és amelyek eredményei közül most bemutatok néhány jelentősebbet. Mindent összevetve a vizsgálatokból arra következtettek, hogy az elemzett nyelvkönyvek tartalma és a korpusznyelvészeti kutatások eredményei között jelentős eltérés áll fenn. Azt vették észre, hogy a tananyagokban használt mintaszövegek, párbeszéddek nagyban eltérnek a korpuszokban található valós nyelvhasználatától. Di Vito (1991) a francia nyelvkönyvekben található tagadó szerkezetek, tárgyias névmások, vonatkozó mellékmondatok használatát, valamint a kérdésalkotási stratégiákat hasonlította össze különböző korpuszokkal. Bebizonyította, hogy nem minden szerkezet alkalmas egyformán a különféle beszélt és írott nyelvi kontextusokban történő használatra. Kutatásából megtudhatjuk, hogy az akkor széles körben használt nyelvkönyvek a gyakorisági mutatók és a nyelvi szerkezetek funkcióinak különbségeit általában figyelmen kívül hagyták. Példáikban gyakran található olyan grammatikai szerkezet, amelynek létezése és gyakori használata egyáltalán nem támasztható alá anyanyelvi beszélők természetes nyelvhasználata által. Cheng és Waren (2007) azt mutatták ki, hogy a tankönyvek olyan nyelvi formákat tartalmaznak, amelyeket ritkán vagy egyáltalán nem használnak valós diskurzusban. A tananyagok szerzőit túlságosan befolyásolják a tudományos műfajú írások. A korpuszokban található leggyakoribb nyelvi formák a tankönyvekben szereplőknél



egyszerűbbek és kevésbé explicitek. A hétköznapi beszélgetések stratégiáiban is eltérések találhatóak (ilyen például a reakciók hosszának aránya, vagy a diskurzusjelölők használata). A régebbi tankönyvek nem tükrözték pontosan az autentikus interakciókat, mert a tananyagok szerzői hagyományosan hajlamosak voltak a dialógusokat médiumként használni bizonyos grammatikai szerkezetek, a szókincs vagy egy nyelvi funkció bemutatására. Az autentikusság helyett inkább a strukturális, funkcionális pedagógiai céljaikra helyezték a hangsúlyt (Gilmore, 2004). Hunston (2002), valamint Biber és Reppen (2002) szerint a tananyagok készítői túlzott mértékben támaszkodtak saját megérzéseikre, intuícióikra ahelyett, hogy a korpuszok empirikus adatait használták volna fel forrásként (vö. Stefanowitsch, 2020: 15–17).

Legtöbbször a grammatika területét emelték ki ezekben a vizsgálatokban. A szerzők ugyanis szinte automatikusan inkább a tradicionális, grammatikaközpontú megközelítést választották a tankönyvek készítése során. Habár nem tudatosan választották a hagyományos megközelítést, sok esetben észre sem vették, hogy a tananyag készítése során bármiféle döntést is meghoztak volna (Biber – Reppen, 2002: 200). Az utolsó kutatás, amelyet lényeges megemlítenünk, Koprowski (2005) kutatása, aki tanulmányában három angol nyelvkönyvet vizsgált meg a bennük található lexikai elemek hasznosságával kapcsolatban. Kutatásának eredménye, hogy a vizsgált tankönyvekben található kifejezések egyharmada csak korlátozottan használható hétköznapi szituációkban egyrészt a gyakorisági mutatóik, másrészt pedig a regiszter miatt, amelyben megjelennek. Azt javasolja, hogy csak olyan kifejezéseket használjanak a szerzők a nyelvkönyvekben, amelyeknek a tudását maximálisan ki tudják használni a tanulók, és hagyják ki azokat a lexikai elemeket, amelyek felszínesek, illetve ritkán használtak (Koprowski, 2005: 331).

Ezekből fakadóan felmerül a kérdés, hogyan használhatók fel a korpuszok a tananyagok készítésének folyamatában. Amikor a szerzők korpuszkutatásukat értelmezik, McCarthy szerint „három átfogó célt tartanak szem előtt:

1. a hiteles, motiváló nyelv azonosítását,
2. hogy ezeket a megállapításokat egy gondosan kidolgozott tantervbe fonják,
3. valamint céljuk, hogy ismerős szerkezetű és könnyen használható tankönyveket hozzanak létre” (McCarthy, 2004: 15).

A korpuszok használata több területen is támogathatja a tananyagkészítők munkáját (vö. McCarthy – McCarten, 2022). Használatukkal el tudják kerülni a megérzésekre való túlzott támaszkodást és ezáltal a rosszul reprezentált nyelvhasználatot is (McEnery – Xiao – Tono, 2006), segíthetnek egy szintezett lexikogrammatikai tanmenet elkészítésében, valamint megfelelő szövegek és valós szituációk kiválasztásában, amelyeknek az eljátszása során a

nyelvtanulók gyakorolhatják a tanult nyelvi elemeket (McCarten, 2010: 415; McCarthy – McCarten, 2022). A nyelvi formák gyakoriságának vizsgálatával a korpusznyelvészet segítséget nyújthat olyan döntések meghozatalában is, amelyek által a tananyag a valódi nyelvhasználatot fogja tükrözni (Conrad, 2000). Ezeken kívül a korpuszok alkalmazása más területeken is nagyon hasznos lehet. Ilyen a hasznos szókincs kiválasztása, a szókincs kapcsolata a különböző szövegtípusokkal (formális – informális stílus, írott vagy beszélt nyelv), grammatikai mintázatok (Kaltenböck – Mehlmauer – Larcher, 2005), és a diskurzusjelölők használatának megfigyelése (McCarten, 2010). A korpuszhasználat megkönnyíti a szerzők munkáját olyan területeken is mint az vizsgált jelenség szövegben való megjelenési arányának meghatározása, vagy a tanulandó nyelvi formák sorrendjének kiválasztása (Meunier – Reppen, 2015: 501). Biber és Reppen ezeken felül azt javasolja még, hogy a tananyagok szerzői támaszkodjanak a gyakorisági vizsgálatok adataira, hogy növeljék a tanulók rendelkezésére álló jelentéssel teli nyelvi inputot (Biber – Reppen, 2002: 207). A szerzőknek érdemes összevetni a tanítandó szókincs a témával, a kontextussal, amelyben használják őket a kompetens beszélők, illetve a grammatikai mintázatokkal, amelyek gyakran megjelennek a környezetükben, így a korpuszokban található szövegek modellként is szolgálhatnak a tanulók számára (Kaltenböck – Mehlmauer-Larcher, 2005: 72).

A legfontosabb pontokat összefoglalva tehát a tananyagok szerzői a munkájuk során a következő területeken alkalmazhatják a korpuszokat:

1. a hasznos lexika kiválasztásakor,
2. a tanított nyelvi formák arányának és sorrendjének kialakításakor, valamint egy színtezett lexikogrammatikai tanmenet elkészítésekor,
3. a témához kapcsolódó lexika, válaszvariációk és szövegtípusok kapcsolatának vizsgálatakor,
4. valós és a tanulók számára releváns szituációk kiválasztásakor,
5. a saját intuícióik helyett a korpuszadatokra támaszkodva a természetes nyelvhasználatot tükröző szövegek készítésekor.

Kijelenthetjük tehát, hogy a korpuszokat forrásként használó tananyagok modellként a valós nyelvhasználatot tükrözik, ellentétben a sokszor egy-egy szerző intuícióra és hagyományos megközelítésekre épülő más, korábbi tananyagokkal. Az így készült tananyag szövegeiben az élőnyelv, valódi, autentikus nyelvhasználat jelenik meg különböző kontextusokban (Vermeki, 2022a). Ez még akkor is így van, ha a nyelvi szintekhez történő igazítás közben némi módosítást végzünk rajtuk. Ezek a tananyagok a felhasználás

szempontjából nem térnek el a korábban megjelent nem korpuszinformált tananyagoktól, használatuk korpuszinformált jellegük miatt nem kíván külön felkészülést a tanároktól. Ugyanakkor autentikus nyelvhasználatuk révén sokkal hatékonyabban alkalmazhatóak még forrásnyelvi környezetben is. A tanulók számára pedig relevánsabbak a beszédszerveződés megismerésében. Az ilyen tananyagok a tanulók motiválásában is nagy szerepet játszhatnak, hiszen a diákok biztosak lehetnek benne, hogy a szövegekben megjelenő nyelv a természetes nyelvhasználaton alapul, a szituációk pedig olyan mindennapi szituációk, amelyek akár velük is megtörténhetnek, hiszen az autentikus vagy szemiautentikus, tehát részeiben módosított, nyelvi szinthez igazított szövegek, amelyek megjelennek a könyvekben, valós szituációkon alapulnak. Ennek következtében nemcsak a tanárok, hanem a tanulók is észreveszik, hogy a szövegek hiteles forráson alapulnak (McCarthy, 2004: 6). A korpuszokból azt is megtudhatjuk, hogyan oldunk meg különböző feladatokat, olyan funkciókat is megfigyelhetünk, mint a kérdés, a kérdésfeltevés, a beszélgetések elkezdésének és befejezésének módjai, a témaváltás, vagy a beszélgetőpartner félbeszakításának stratégiái (McCarthy, 2004: 6). A korpuszokon történő gyakorisági vizsgálatok alapján, olyan magas kommunikációs értékkel rendelkező, hasznos szavak és nyelvi szerkezetek kerülhetnek bele a tananyagba, amelyek elősegítik a nyelvtanulók természetes nyelvhasználatra hasonlító, folyékony beszédének fejlődését.

A természetes nyelvhasználatához hasonló kommunikáció elérésének céljából fontos például az is, hogy a dialógusok ne csak a tradicionális osztálytermi kérdés-válasz szekvenciát kövessék, hiszen a valós beszélgetésekhez hozzátartoznak még a párbeszédekben történő reakciók is (például *Ó!*, *Tényleg?*, *Szerintem is.*) (McCarthy, 2009: 122–125). Ezeket a reakciókat szintén könnyen megismerhetjük a korpuszokból, amelyek felhasználásával a szerzők még hétköznapiabbá tehetik a korábbi tankönyvekben sokszor túlzottan mesterkéltnek hangzó dialógusokat (McCarthy, 2004: 8–9). Jó példa erre a következő tankönyvi szövegrészlet, amelyből azt is láthatjuk, hogy már kezdő nyelvi szinten is természetesebb dialógusokat kaphatnak így a nyelvtanulók:

- Tímea:** És miért tanulsz csehül?  
**Anna:** Mert a barátom cseh.  
**Tímea:** *Tényleg?* És milyen nyelven beszéltek otthon?  
**Anna:** Magyarul és egy kicsit csehül. Patrik, a barátom nagyon jól tud magyarul.  
**Tímea:** *Érdekes.* Az én barátom is cseh. Mi otthon csehül beszélgetünk.

(Szita – Pelcz, 2013: 26)

## **6.5.2 Korpuszinformált tananyagok a magyar mint idegen nyelv oktatásának területén**

Habár ezen a területen még nem bővelkedünk korpuszinformált tankönyvekben, ahogy más nyelveknél, a magyar mint idegen nyelv tanítása során is elengedhetetlen az élőnyelvi példák alapuló szövegek használata. A következőkben a Szita Szilvia és Pelcz Katalin által írt *MagyarOK* című, négyrészes korpuszinformált tankönyvcsaládot (2013–2019) szeretném röviden bemutatni, azokra a szempontokra koncentrálva, amelyek mintaként szolgálnak a korpuszok nyelvoktatásban való felhasználásához és a korpuszinformált tananyagok elkészítéséhez.

### **6.5.2.1 A hasznos lexika kiválasztása a tananyag számára**

A *MagyarOK* szerzői számos forrást használtak fel munkájuk elkészítéséhez, például a Magyar Nemzeti Szövegtárt, a Hunglish néven ismert angol-magyar korpuszt, közösségi médiából (Twitter és Facebook) származó szövegeket, saját gyűjtésű anyagokat és kompetens nyelvhasználókkal rögzített interjúkat (Szita, 2014: 115). A lexika kiválasztása során mindig az előre meghatározott témára és a kommunikációs szituációra alapoztak (Szita – Pelcz, 2020). Ebből kiindulva a fent említett források lehetővé tették a számukra, hogy a lexikát gyakoriság és relevancia alapján válasszák ki a korpuszokon végzett statisztikai mérések segítségével. A lexika kiválasztásának e módszerével lehetővé vált annak megfigyelése, hogy az egyes lexikai elemek milyen tipikus kontextusokban fordulnak elő, amely a tankönyvek hatékonyságának növelésében fontos tényező volt (Szita, 2014: 115), és amelyre a 6.5.2.3 fejezetben térünk ki bővebben.

### **6.5.2.2 A tanított nyelvi formák arányának és sorrendjének kialakítása, egy szintezett lexikogrammatikai tanmenet elkészítése**

A szerzők a Magyar Nemzeti Szövegtárt, a Hunglish és az általuk készített korpuszokat többféleképpen is alkalmazták. A hasznos és releváns lexika mellett azt is a korpuszok segítségével határozták meg, hogy mi kerüljön bele a tananyagba és a kiválasztott elemek pontosan hova kerüljenek. A nyelvi szintekhez kötött grammatikai struktúrákat – mivel a szerzők véleménye szerint a használatukat és funkciójukat nagy mértékben a lexika határozza meg – a szintén gyakorisági alapon kiválasztott témákhoz és kommunikációs helyzetekhez

kapcsolódó lexikához illesztették (Szita, 2014: 115). Ezenkívül a nyelvi formák elhelyezésének kérdésében a hasznosságuk, gyakoriságuk és funkciójuk is segítette a szerzőket. Az így elkészült lexikogrammatikai tanmenet előrehaladása kisebb lépésekből áll a kezdeti szakaszban, majd progresszív módon történik, amikor a diákok már rendelkeznek némi nyelvi ismerettel. Erre jó példa az A1-es könyvben (2013) a határozatlan és a határozott igeragozás tanítása. A határozatlan igeragozás megismertetése fejezetenkénti bontásban történik, vagyis minden egyes fejezet egy adott igitípust tárgyal (az elején két részre bontott paradigmával). A határozott igeragozás bemutatása előtt (6. fejezet) a teljes rendszert összefoglalják. A határozott igeragozást is részletesen tárgyalják, de a teljes paradigma egy leckében jelenik meg, mivel a diákok ekkorra már ismerik az alapvető fogalmakat, és az eddig tanult birtokos személyjelek és ragozott személyes névmások is segítenek nekik (Szita – Pelcz, 2020).

### 6.5.2.3 A témához kapcsolódó lexika, válaszvariációk és szövegtípusok kapcsolatának beépítése a tananyagba

A *MagyarOK* tankönyvcsalád a szerzők által kidolgozott, modellalapú nyelvoktatás módszerét alkalmazza. A modellalapú nyelvoktatás egy összetett módszer, amelynek jellemzője, hogy tematikus ismétlésekre és variációkra épít, lexikogrammatikai, spirális felépítésű, valamint nagy nyelvi inputot ad. A tankönyvek minden fejezetében modelldialógusok, mondatmodellek találhatóak, amelyeknek a „célja, hogy mintát adjon a nyelvtanulónak, hogyan reagálhat az adott közléshelyzetben” (Szita – Pelcz, 2017: 264). Az egy-egy közléshelyzethez összegyűjtött gyakori kifejezések ezt követően többször is újabb variációkban jelennek meg, bővülnek az adott könyvben és a magasabb nyelvi szintekre íródott könyvekben, valamint más nyelvi regiszter tanításakor is. Ezt a szerzők spirális bővítésnek nevezik, amely azért hatékony, mert a nyelvtanuló már meglévő tudására építve kapcsolja össze az újabb és újabb elemeket (Szita – Pelcz, 2017: 265).

14. táblázat Példa az ismétlődő előfordulásokra a MagyarOK tananyagaiban

|                        |   |                       |        |
|------------------------|---|-----------------------|--------|
| <b>Lexikai egység:</b> | <i>ha van egy kis időm / szabadidőm</i> |                       |        |
| <b>Gyakorisága:</b>    | 12                                      | 59,68 / egymillió szó | 0,006% |

| Sz. | Könyv / Szint | Fejezet / Forrás | Példa   |
|-----|---------------|------------------|---|
| 1.  | MagyarOK A1   | 4.               | <b>Ha van egy kis időm</b> , futok a közelben.          |
| 2.  | MagyarOK A1   | 6.               | <b>Ha van egy kis időm</b> , telefonálok a barátnőmnek. |
| 3.  | MagyarOK A1   | 6.               | <b>Ha van egy kis időm</b> , sportolok.                 |

| Sz. | Könyv / Szint | Fejezet / Forrás | Példa   |
|-----|---------------|------------------|---|
| 4.  | MagyarOK A1   | 7.               | <b>Ha van egy kis szabadidőm</b> , szeretek vásárolni, barátnőmmel találkozni, és imádok Operába menni. |
| 5.  | MagyarOK A1   | videó            | <b>Ha van egy kis időm</b> , mindig játszom rajta.  |
| 6.  | MagyarOK A1   | videó            | <b>Ha van egy kis szabadidőm</b> , zongorázom.  |
| 7.  | MagyarOK A1   | videó            | <b>Ha van egy kis szabadidőm</b> , énekelek.  |
| 8.  | MagyarOK A1   | videó            | <b>Ha van egy kis időm</b> , természetesen számítógépezek.  |
| 9.  | MagyarOK A1   | videó            | <b>Ha van egy kis időm</b> , természetesen gördeszkázom.  |
| 10. | MagyarOK A2   | 1.               | <b>Ha van egy kis szabadidőm</b> , olvasok vagy festek.   |
| 11. | MagyarOK A2   | 3.               | <b>Ha van egy kis időm</b> , általában tévém.   |
| 12. | MagyarOK A2   | 3.               | <b>Ha van egy kis időm</b> , inkább olvasok.  |

A tananyagban lévő feladatszekvenciákat úgy tervezték meg, hogy a nyelvtanulóknak először lehetőségük legyen megfigyelni a különböző szituációkhoz, interakciókhoz köthető nyelvet, majd a többféle, támogatásként szolgáló mintamondat segítségével saját magukra szabhatják a szövegeket. Ezek a modellmondatok különböző variációkban vissza-visszatérnek a tankönyv fejezeteiben, így segítve a könnyebb memorizálást. Ennek több előnye is van, például a tanulóknak nem kell saját maguktól új megnyilatkozásokat alkotniuk. Ez nem pusztán utánpótlás, hanem – hasonlóan ahhoz, amit a használatalapú elméleteknél a nyelvelsajátítással kapcsolatban Tomasello (2003, 2013) kijelentett az imitációról (ld. II. 2.1 fejezet) – az anyanyelvi beszélők által használt nyelv értelmes és funkcionálisan megfelelő megismétlésével egyenlő. A teljesen új mondatok készítése sok időt venne igénybe és nagy kognitív terheléssel járna. Valamint – ahogy már korábban említésre került (ld. II.4 fejezet) – a kompetens nyelvhasználók sem így beszélnek, főleg előre felépített, formulaszerű szerkezeteket használnak beszéd közben (vö. Wray, 2002; Wray – Fitzpatrick, 2008; Wray, 2019). A tanulók a modellmondatokra támaszkodhatnak a saját mondataik megalkotásakor és közben biztosak lehetnek abban, hogy az adott szituációban ezek a mondatok valóban elhangozhatnak a kompetens nyelvhasználók szájából (Szita – Pelcz, 2017: 264). Egy másik előnye, hogy nagy gyakoriságú mondatvariációkat kapnak készen a tanulók, amelyek a természetes nyelvhasználatot tükrözik, ezáltal az ő beszédük is közelíthet az anyanyelvi beszélőkéhez (Szita, 2014: 116). A fentieket jól illusztrálja a következő részlet, ahol a születésnap szervezésének témakörében készült feladatban, a meghívással kapcsolatban (meghívás és reakció) és a telefonbeszélgetésekre is jellemző lexikai egységeket is felfedezhetünk.

**Zsuzsanna:** Szia, Csilla! Zsuzsa vagyok. Nem zavarlak?

- Csilla:** Szia, Zsuzsa! Egyáltalán nem zavarasz. Mi újság?
- Zsuzsanna:** Azért hívlak, mert vasárnap lesz Zalán születésnapja...
- Csilla:** Nem felejtettem el. Már ajándékot is vettem.
- Zsuzsanna:** Hogy te milyen figyelmes vagy! Szóval: vasárnap rendezünk egy kis ünnepséget, és szeretnék meghívni benneteket. Ráértetek?
- Csilla:** Persze. Délelőtt vagy délután?
- Zsuzsanna:** Délután három körül.
- Csilla:** Ott leszünk. Ha valami közbejön, még kereslek.
- Zsuzsanna:** Rendben várlak benneteket. Szia!
- Csilla:** Szia!

(Szita – Pelcz, 2013: 174)

#### 6.5.2.4 Valós és a tanulók számára releváns szituációk beépítése a tananyagba

Ahogy korábban említésre került (6.5.2.1–6.5.2.3 fejezet) a szerzők mindig egy adott témát, kommunikációs szituációt vettek alapul. A kiválasztott témák és kommunikációs szituációk egy anyanyelvi beszélő mindennapjaiban előforduló hétköznapi helyzetek, illetve emiatt a nyelvtanulók számára is relevánsak. Ilyen például az előző részben látható telefonbeszélgetés, amelyből két helyzetet (meghívás, telefonbeszélgetés) és háromféle csoportba sorolható nyelvhasználati modelleket is megismerhetnek a nyelvtanulók. Egyrészt megtudhatják azt, hogy az anyanyelvi beszélők hogyan hívnak meg valakit egy eseményre, másrészt azt, hogyan lehet erre reagálni, harmadrészt pedig, milyen a telefonbeszélgetésekre jellemző keretes szerkezet, és melyek a bennük megjelenő leggyakoribb lexikai egységek (15. táblázat).

**15. táblázat** A születésnapi meghíváshoz, valamint a telefonbeszélgetésekhez tartozó lexikai egységek az előző párbeszédben (Szita – Pelcz, 2013: 174. alapján) (Vermeki, 2022a: 262)

| Születésnapi meghívás   | Reakciók a meghívásra  | Telefonbeszélgetés  |
|---|--|---|
| <ul style="list-style-type: none"> <li>• vasárnap lesz Zalán születésnapja</li> <li>• vasárnap rendezünk egy kis ünnepséget</li> <li>• szeretnék meghívni benneteket</li> <li>• délután három körül</li> <li>• várlak benneteket</li> </ul> | <ul style="list-style-type: none"> <li>• nem felejtettem el</li> <li>• már ajándékot is vettem</li> <li>• ott leszünk</li> <li>• ha valami közbejön, még kereslek</li> </ul> | <ul style="list-style-type: none"> <li>• Szia! ... vagyok. Nem zavarlak?</li> <li>• Szia, ...! Egyáltalán nem zavarasz.</li> <li>• Mi újság?</li> <li>• azért hívlak, mert...</li> <li>• ...még kereslek</li> </ul> |

A születésnap témaköre egy későbbi feladatban (19. ábra) is előkerül, amikor a tanulóknak a legutóbbi saját születésnapjukról kell beszélniük. Egy ilyen történet elmesélése szintén releváns lehet a nyelvtanulók számára. Ennek a feladatnak a megoldásához hasznos mondatmodelleket kapnak segítségül (Szita – Pelcz, é.n.: 131). A *MagyarOK* mondatmodelljei (a 19. ábrán kék színnel jelöltek) – ahogy ezt az előző fejezetben már említettük – arra szolgálnak, hogy segítsék a tanulókat egy adott témával kapcsolatos beszélgetés közben. Ki kell emelnünk, hogy ezek a mondatmodellek nem egy mondattípus variációi, hanem a témához kapcsolódó választípusoké. Tehát arra példák, hogy egy anyanyelvi beszélő hogyan reagál, reagálhat a témával kapcsolatban feltett kérdésekre (Szita – Pelcz, 2020).

**19. ábra** Példa a mondatmodellek bemutatására, variációk egy témára (Szita – Pelcz, 2013: 177)

20. Ön következik. Mit csinált egyik emlékezetes születésnapján? Meséljen és írjon róla!

Mikor volt a születésnapja?

Kivel ünnepelt?

Kitől milyen ajándékot kaptott?

Hogyan töltötte a napot?

Július 5-én volt a (harmincadik) születésnapom.  
Nagy bulit rendeztem.  
Minden barátomat meghívtam.  
Moziba mentem a feleségemmel.  
A családommal ünnepeltem.  
Sok ajándékot kaptam.  
Az öcsémtől egy wellness-hétfőjét kaptam ajándékba.

Vettem magamnak egy ruhát.  
Egy luxusétteremben vacsoráztam.  
Énekeltem és táncoltam.  
Tortát ettem és pezsgőt ittam.  
Nem csináltam semmit.  
Én adtam ajándékot másoknak.  
Nagyon boldog voltam.  
Nagyon szép nap volt.

**6.5.2.5 Saját intuíciók helyett a korpuszadatokra támaszkodva készített, a természetes nyelvhasználatot tükröző szövegek beépítése a tananyagba**

A természetes nyelvhasználatot tükröző szövegekkel kapcsolatban említettük, hogy a szerzők – a fenti korpuszok (Magyar Nemzeti Szövegtár, Hunglish) mellett – saját gyűjtésű anyagokat és kompetens nyelvhasználókkal rögzített interjúkat (Szita, 2014: 115) is használtak a *MagyarOK* tankönyvcsalád készítésekor. Így készültek a tankönyvek szövegei, amelyek az alacsonyabb nyelvi szintre didaktizáltak, például rövidítettek, ugyanakkor nem veszítették el a természetes nyelvhasználatra jellemző tulajdonságokat (Szita, 2021). Ennél a résznél szükséges megemlítenem, hogy a *MagyarOK* tankönyvcsaládhoz készül egy szintén



korpuszalapú kiegészítő, olvasókönyv-sorozat, amelynek jelenleg az A1-es része érhető el, és amely akár más tankönyvekkel vagy önállóan is használható. Az *Olvasókönyv a MagyarOK A1+ és kompakt* fejezeteihez kapcsolódva további tíz-tíz modellszöveget tartalmaz. Ezek a szövegek a tankönyvi fejezetek összefoglaló feladataira készült változatok, amelyeket tíz nyelvtanár írt arra törekedve, hogy a tankönyvben szereplő lexika minél magasabb arányú használatával hozzák létre saját szövegeiket, ugyanakkor ne egyszerűsítsék le nyelvhasználatukat, hanem maradjon természetes a kifejezőmódjuk (Szita, 2021).

A modellszövegek olvasása által a nyelvtanulók megfigyelhetik, hogy a kompetens nyelvhasználók milyen tipikus nyelvi fordulatokat használnak egy-egy témakörrel kapcsolatosan. A szövegekhez szókinccs- és szövegértést fejlesztő feladatok is tartoznak. A fejezetek három részből állnak, a már említett modellszövegek, a szavak kontextusban rész, valamint a hasznos mondatok elnevezésű rész. A fejezetek második részének utolsó feladata a *Lépjünk tovább!* című feladat a *MagyarOK* nyílt korpuszának használatára ösztönzi a tanulókat (Szita, 2022: 3–4). Az *Olvasókönyv* a szerkesztésében is segíti a nyelvi fordulatok alapos megfigyelését, a modellmondatok a konkordanciasorokhoz hasonlóan rendezettek (ld. KWIC, III.2.1.4.1 fejezet), az igék pirossal és félkövéren szedettek. Ezáltal a mintázatok könnyebben észrevehetőek, gyorsabban memorizálhatóak, illetve a hasonló szerkesztés által a korpuszok tanuláshoz való használatába is bevezeti a nyelvtanulókat.

## 20. ábra Példa a Szavak kontextusban című részből (Szita, 2022: 16)

### 1. A *vagyok* használata • The use of *vagyok* (I am)

#### a) Olvassa el a mondatokat! • Read the sentences.

Gergő **vagyok**.

Veronika **vagyok**.

Orvos **vagyok**.

Fordító és tanár **vagyok**.

Szobrász **vagyok**, a férjem pedig tanár.

Angolul jól beszélek, angoltanár is **vagyok**.

Tanár **vagyok** és újságíró.

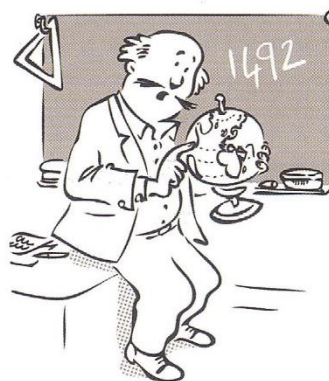
Tanár **vagyok** a Belgrádi Egyetemen.

Fordító **vagyok** Londonban.

Programkoordinátor **vagyok** a Pécsi Egyetemen.

Magyar **vagyok**, de Szerbiában dolgozom.

Magyar **vagyok**, de Franciaországban élek.



A *MagyarOK* tankönyvesalád nyelvtanulással kapcsolatos szemlélete, korpuszinformált jellege és modellalapú módszertana által jelentős változást hozott a magyar mint idegen nyelvi tananyagok területén, ezért még bőségesen lehetne elemezni. A fenti rövid összefoglalóval a szándékom a jelen doktori értekezésben bemutatott kutatáshoz kapcsolódó leglényegesebb információk kiemelése volt, vagyis, hogy milyen korpuszokat és hogyan használtak a szerzők a tankönyvek elkészítéséhez, milyen módszertant fejlesztettek ki a hatékony nyelvtanulás célját szem előtt tartva, illetve, hogyan tükrözi a tananyag a természetes nyelvhasználatot.

### III. A kutatás

#### 1. A KorSzak Gyermeeknyelvi Korpusz

A KorSzak Gyermeeknyelvi Korpusz egy olyan dinamikusan fejlődő, multimodális pedagógiai korpusz, amelynek célja, hogy autentikus szövegeket biztosítson a gyermekek számára készülő tananyagokhoz, valamint az annak elkészülését megelőző korpuszalapú kutatásokhoz. Ahogy már korábban említésre került, a jelen doktori értekezésben a pedagógiai korpusz fogalmát a nyelvoktatáshoz készült autentikus szövegek halmazaként értelmezzük. A korpuszban szereplő szövegek összeállítása tudatosan, az elsődleges cél, vagyis a tananyagkészítés szem előtt tartásával történt. A korpusz általános adatait és jellemzőit, építésének céljait, tervezését és folyamatát ebben a fejezetben ismertetem részletesen.

##### 1.1 Néhány szó a Korpusznyelvészeti és Szakmódszertani Munkacsoportról

A jelen doktori értekezésben bemutatott kutatást a KorSzak Gyermeeknyelvi Korpuszon végeztem. A korpusz a Korpusznyelvészeti és Szakmódszertani Munkacsoport (rövidítve KorSzak) által épített korpuszok részét képezi. A KorSzak Munkacsoport 2020 februárjában jött létre azzal a céllal, hogy „hidat építsen a korpusznyelvészeti kutatások és a magyar mint idegen nyelv szakmódszertana között” (Baumann et al., 2021: 32). Az alapító tagok arra a felismerésre jutottak, hogy sokszor az empirikus kutatások eredményei nem tükröződnek a magyar mint idegen nyelv szakmódszertanában, pedig az elengedhetetlen lenne a szakmódszertan fejlődésének és a hatékony nyelvoktatás érdekében. Ahogy korábban már említésre került (ld. II.6.5.1 fejezet) a korpuszalapú vizsgálatok eredményeinek felhasználása számos előnnyel jár a nyelvoktatás és a tananyagkészítés területén, hiszen a korpuszokon végzett kutatások eredményeiből megtudhatjuk többek között, hogy mely nyelvi elemet milyen gyakran használjuk, milyen tipikus kontextusban és grammatikával jelenik meg (vö. Hoey, 2005). A korpuszvizsgálatok során tehát információt kaphatunk a valódi, természetes nyelvhasználatról, a lexikai, grammatikai, valamint pragmatikai kérdésekre vonatkozóan is. Az így nyert információkat felhasználva, többé nem kell a saját egyéni intuíciónkra támaszkodni, hanem megbízható, autentikus adatok segítségével tehetjük módszertanunkat hatékonyabbá

(Hoey, 2005; Hunston, 2009; O’Keeffe et al., 2007). A KorSzak munkacsoport is ezeken a területeken szeretne hozzájárulni a nyelvtanárok, nyelvtanulók, valamint a tananyagkészítők munkájához. Ezen célok elérése érdekében a Munkacsoport tagjai több almunkacsoport létrehozásával, több korpusz építését kezdték meg (Baumann et al., 2021: 32–33). A jelenleg épülő korpuszok a következők:

1. Élőnyelvi korpusz Pelcz Katalin vezetésével.
2. Tanulói korpusz Baumann Tímea vezetésével.
3. Gyermeknyelvi korpusz Vermeki Boglárka vezetésével.

A KorSzak munkacsoport 2020 februárjában, a Pécsi Tudományegyetem Nemzetközi Oktatási Központjában megtartott alakuló ülése után, szinte azonnal elkezdte a fent említett korpuszok létrehozását. Azóta rengeteg adatmennyiség gyűlt össze, a korpuszok folyamatosan bővülnek. A Munkacsoport taglétszáma pedig az alakulás óta több mint duplájára emelkedett. Rendes, támogató, valamint tiszteletbeli tagjai között egyre több magyar (Hegedűs Rita, Nádor Orsolya) és külföldi kutatót (Michael Hoey†, Alex Boulton, Maggie Charles, Anne O’Keeffe, Szirmai Mónika, Susan Hunston) köszönthet (Vermeki, 2023c, megjelenés alatt).

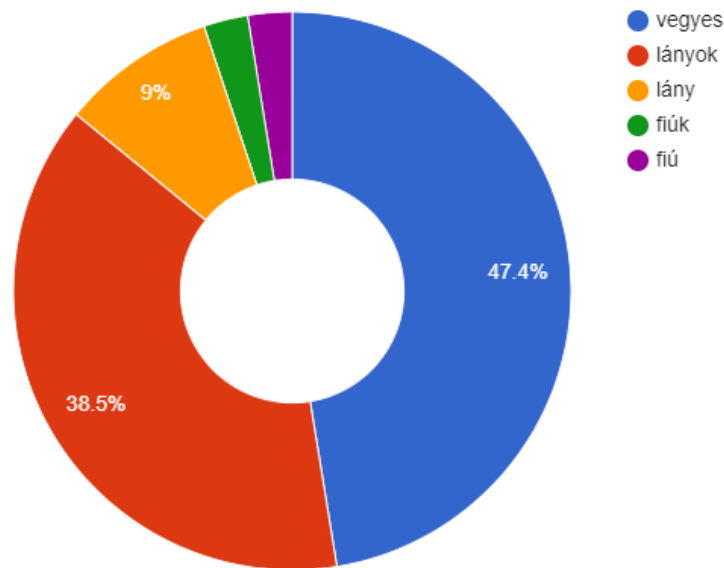
## **1.2 A KorSzak Gyermeknyelvi Korpusz építésének célja és folyamata**

A KorSzak Gyermeknyelvi almunkacsoport célja olyan pedagógiai célú korpusz felépítése, amely különböző, a gyermekek számára is releváns és érdekes témákkal foglalkozik. Elsődleges célkitűzése, hogy biztosítsa azt a korpuszt, amelynek vizsgálata során nyert adatok felhasználásával a magyart idegen- vagy származásnyelvként tanuló gyermekek számára tananyag készülhet (Baumann et al., 2021: 33).

Mivel a korpuszépítés egyik fontos célkitűzése a korpuszinformált tananyagok készítése, a KorSzak Gyermeknyelvi Korpusz a mintavétel módja szerint dinamikus, a felhasználási módja szerint pedig egy speciális, pedagógiai célú korpusz, a kategóriájára jellemző minden tulajdonsággal rendelkezik (ld. II.6.3 fejezet). A korpusz dinamikus, mert folyamatosan növekszik. Jelenleg, 2022 októberében hetvennyolc felvételt tartalmaz, huszonkilenc 11–15 éves gyermek (kilenc fiú és húsz lány) beszélgetéseit (Vermeki, 2023c, megjelenés alatt).

A diagrammon (21. ábra) látható a rögzített felvételeken szereplő gyermekek nemek szerinti megoszlása. A vegyes kifejezés arra utal, hogy az adott felvételeken lányok és fiúk is hallhatóak, a lányok és fiúk jelentése pedig, hogy a rögzített felvételeken több gyermek is hallható, amíg az egyes számban szereplő lány és fiú kategória a monológokban szereplőkre utal.

**21. ábra** A lányok és fiúk aránya a KorSzak Gyermeknyelvi Korpuszban (2022. október)



### 1.2.1 A témakörök

A korpusz másik fő tulajdonsága, hogy pedagógiai céllal létrehozott, vagyis a szövegek előre meghatározott témakörökről szólnak, amelyek a tananyag témakörei is lesznek. Ezek jelenleg a következők:

**22. ábra** A KorSzak Gyermeknyelvi Korpuszban szereplő szövegek témakörei

|  |  |
|--|--|
| <ul style="list-style-type: none"><li>• állatok (kutyák),</li><li>• család,</li><li>• ételek,</li><li>• filmek és sorozatok (Harry Potter),</li><li>• hobbi,</li><li>• influenzszerek,</li><li>• kirándulás,</li></ul> | <ul style="list-style-type: none"><li>• saját történet,</li><li>• sport,</li><li>• személyleírás,</li><li>• találmányok,</li><li>• utazás,</li><li>• zene.</li></ul> |
|--|--|

Az eredetileg előre meghatározott témakörök mellé került egy egyéb kategória is, amely az adatközlők által különböző általuk választott témákban készült beszélgetések és monológok gyűjteménye (Vermekei, 2023c, megjelenés alatt).

**16. táblázat** A KorSzak Gyermeznyelvi Korpuszban lévő témakörök tokenszám szerint

| Sorsz. | Témakör             | Altémakör*                                  | Felvételek száma | Tokenek száma |
|--------|---------------------|---|------------------|---------------|
| 1.     | Állatok             | Kutyák                                      | 10               | 24361         |
| 2.     | Filmek és sorozatok | Filmek, Sorozatok, Harry Potter, Hétvége    | 9                | 13942         |
| 3.     | Hobbi               | Kreatív, Sportolás, Zenélés, Videójáték     | 12               | 10683         |
| 4.     | Influenzerek        |   | 1                | 5065          |
| 5.     | Saját történet      |   | 8                | 4606          |
| 6.     | Egyéb               | Ajándék, Barátság, Szeretet, Az osztályom** | 14               | 4120          |
| 7.     | Sport               |   | 4                | 2875          |
| 8.     | Utazás              |   | 4                | 2154          |
| 9.     | Kirándulás          |   | 3                | 1915          |
| 10.    | Személyleírás       | Tanár, Családtag, Barát bemutatása          | 4                | 566           |
| 11.    | Család              |   | 5                | 510           |
| 12.    | Ételek              |   | 2                | 286           |
| 13.    | Találmányok         |   | 1                | 279           |
| 14.    | Zene                |   | 1                | 184           |

\* Nem minden meghatározott témakörhöz tartozik altémakör.

\*\* Az egyéb kategóriában lévő alkategóriák nem mind kerültek felsorolásra, mert ezek szinte mind egy-egy felvételt jelentenek.

A KorSzak Gyermeznyelvi Korpusz építése a fenti témakörök meghatározásával kezdődött, amely előtt készítettem egy felmérést 11–14 éves gyermekek körében. Erre azért volt szükség, hogy valóban olyan témakörök kerüljenek kiválasztásra, amely a gyermekek számára érdekesek és relevánsak. Ezt három okból tartottam fontosnak. Elsősorban a készülő

tananyagra gondolva azért, mert a tananyagot később használó gyermekek számára is várhatólag érdekesek és relevánsak lehetnek az ily módon kiválasztott témakörök. Ezenkívül azért, hogy a gyermekek természetes nyelvhasználatának rögzítését elősegítse, amely sokkal jobban megfigyelhető, ha a dialógusok közben olyan témákról van szó, amelyekről szeretnek beszélgetni. Végül pedig azért, hogy az adatközlő gyermekek jól érezzék magukat a felvételek alatt. Bár a felmérésben résztvevő gyermekek közül csak néhányan vettek részt a rögzítés folyamatában is, megfigyelhető volt, hogy a témakörök többsége tetszett a felvételeken szereplő többi gyermek számára is. A felmérésben lévő kérdések között szerepelt tehát, hogy milyen témákról szeretnek beszélgetni, melyek a kedvenc szabadidős tevékenységeik, valamint, hogy kikkel töltik szívesen a szabadidejüket. A felmérés során, amelyben 138 általános iskolás vett részt, a gyermekeknek párban vagy kiscsoportban ki kellett egészíteniük egy gondolatterképet, és nyitott kérdésekre kellett válaszolniuk. A témákat életkor és gyakoriság szerint válogattam ki, majd a leggyakrabban előfordulók közül a munkacsoport tagjaival elsőként tizenkét témát határoztunk meg (Vermekei, 2023c, megjelenés alatt).

**23. ábra** A felmérés eredményeinek összesítése közben készült Excel tábla részlet

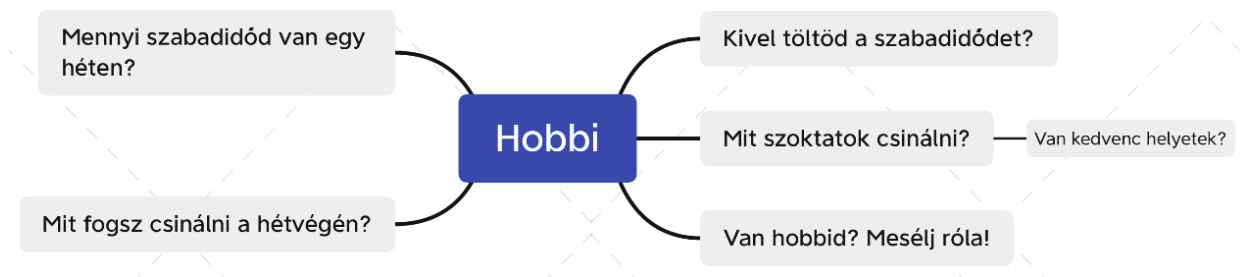
|    | A                | B                | C                | D   |
|----|------------------|------------------|------------------|---|
| 1  | <b>11 éves</b>   | <b>12 éves</b>   | <b>13 éves</b>   | <b>14 éves</b>                              |
| 2  | állatok          | állatok          | állatok          | „Sztrímerek” nézése                         |
| 3  | barátokkal lenni | anime            | Ariana Grande    | barkácsolás, szerelés                       |
| 4  | BFF              | Ariana Grande    | család           | buli  |
| 5  | biciklizés       | barátokkal lenni | Family Guy       | érdekes dolgok a neten                      |
| 6  | Brawl stars      | BFF              | festés, rajzolás | Facebook                                    |
| 7  | család           | biciklizés       | Fifa             | Family Guy                                  |
| 8  | cserkészlet      | Brawl Stars      | gitár            | GTA   |
| 9  | csocsó           | bunkerépítés     | görkori          | Insta                                       |
| 10 | divat            | dínók, régészet  | Insta            | kutya                                       |
| 11 | édességek        | éneklés          | kutya            | League of legends                           |
| 12 | ékszerek         | Facebook         | ló, lovaglás     | Messenger                                   |
| 13 | éneklés          | Family Guy       | Pizza            | Minecraft                                   |
| 14 | filmek           | FB Messenger     | PS4              | online játékok                              |
| 15 | foci             | festés, rajzolás | Rick és Morty    | sorozatok, filmek (Dexter, Supernatural...) |

A kiválasztott tizenkét téma később, az első felvételi folyamat lezárulta után további témákkal bővült, illetve elrendezésében (téma, altéma) átalakult, hogy jobban illeszkedjen a korpusz építésének céljához.

### 1.2.2 A felvételek módja

A beszélgetések rögzítése előtt minden témakörhöz egy gondolattérkép (61. ábra) készült, rövid kérdésekkel a beszélgetés hosszan elnyúló megakadásának elkerülése érdekében. Amikor a gyermekek elakadnak a beszélgetés közben, a kérdések megtekintésével új ötleteket kaphatnak a folytatáshoz (Baumann et al., 2021: 33–35). Ugyanakkor minden beszélgetés felvétele előtt felhívtuk az adatközlők figyelmét arra, hogy ne ragaszkodjanak ezekhez a gondolattérképekhez, csak támaszként használják őket.

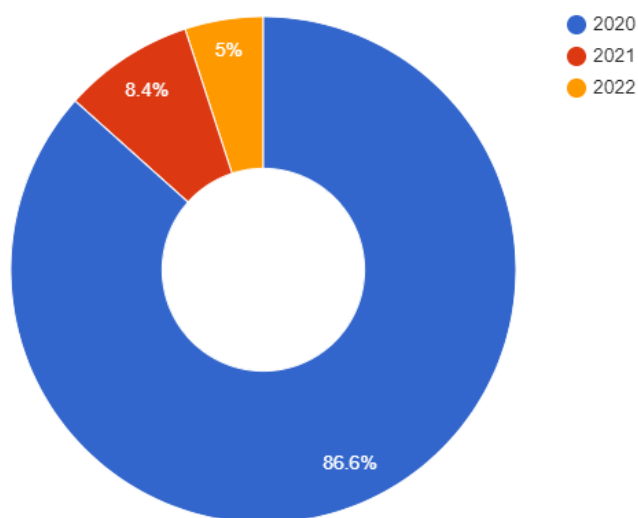
24. ábra A hobbi témakörhöz készült gondolattérkép



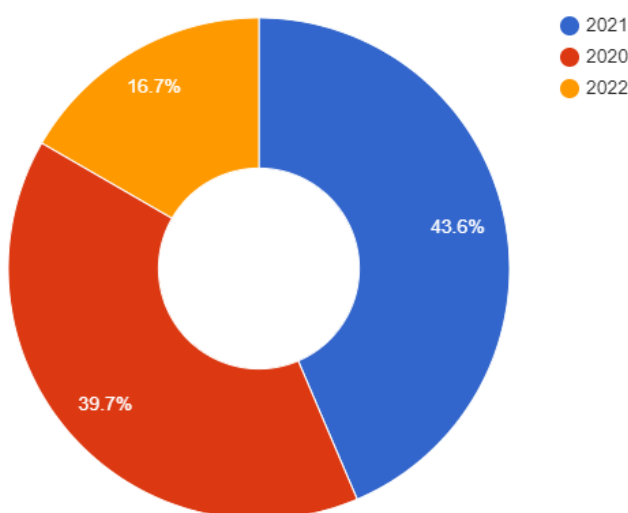
A beszélgetések rögzítésének folyamata 2020 tavaszán kezdődött és azóta is folyamatosan zajlik. Először a Zoom videokonferencia program segítségével, később személyes találkozások alkalmával. A leghosszabb, legtöbb tokent tartalmazó felvételek a 2020-as évben készültek (25. ábra), a legtöbb felvétel pedig 2021-ben (26. ábra). Ennek az eloszlásnak egyik oka lehet, hogy a Zoom programon keresztül történő felvételek rögzítése a koronavírus járvány idején valósult meg, amikor mindenki sokat volt otthon, így a felvételek is hosszabbak lehettek. Ezzel szemben a személyesen történő felvételek délutáni szakkör és egyéb foglalkozások keretein belül történtek, így bár mennyiségileg több, ugyanakkor időtartamában rövidebb beszélgetések születtek.



**25. ábra** A felvételek tokenek száma szerinti évenkénti eloszlása



**26. ábra** A felvételek számának évenkénti eloszlása



### 1.2.3 A természetes nyelvhasználat megfigyelhetőségének kérdése

Mind a kutatás, mind a tananyag érdekében lényeges a gyermekek természetes nyelvhasználatának megfigyelése. Ugyanakkor kutatásetikai megfontolásból – amelyre a III. 5. fejezetben fogok részletesebben kitérni – a felvétel tényéről nemcsak a szülőket, hanem a gyermekeket is tájékoztatni kellett. Felmerülhet a kérdés, hogyan lehet így a természetes nyelvhasználatot rögzíteni.

A természetes nyelvhasználat rögzítése érdekében több intézkedést is megtettünk:

1. Megfelelő témakörök kiválasztása: ahogy az előző részben említettem már a témakörök kiválasztása is a természetes nyelvhasználat rögzítésének elősegítése érdekében történt a fent (III.1.2.1 fejezet) említett módon.
2. A gyermek adatközlők kiválasztása: a felvételeken a KorSzak Munkacsoport tagjainak családtagjai, családi barátok gyermekei, közeli tanítványok és olyan ismerősök szerepelnek, akik közelebbi kapcsolatban állnak a felvétel rögzítőjével.
3. A felvételek módja: A felvételek közben a gyermekeket magukra hagytuk. A felvétel rögzítője egy idő után kikapcsolta a kameráját, egy kicsit távolabb vonult a helyiségben.

### 1.3 A KorSzak Gyermekegyelvi Korpusz általános adatai és jellemzői

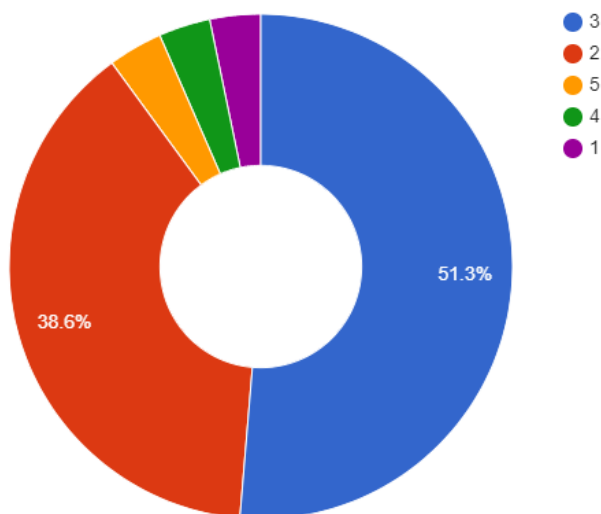
A KorSzak Gyermekegyelvi Korpusz jelenleg, 2022 októberében több mint 71 000 tokent tartalmaz. A videó- és hangfelvételek során a gyermekek párokban vagy kiscsoportokban szabadon beszélgetnek egy-egy témáról. A feladat időtartama nem volt meghatározva, ezért vannak 1-2 perces és ennél jóval hosszabb, akár 45 perces felvételek is, a már említett 2020-ban Zoom videokonferencia program segítségével készült felvételek. A korpusz hatvannyolc dialógusból és tíz monológból áll.

**17. táblázat** A KorSzak Gyermekegyelvi Korpuszban lévő szövegek típusai

| Típus    | Tokenek száma |
|----------|---------------|
| Dialógus | 69101         |
| Monológ  | 2445          |

A monológok a 2021-es évben először a járvány miatt keletkeztek, majd később a személyleírás témakör bevezetése miatt további felvételekkel gyarapodott a korpusz. A dialógusok során egyszerre 2 - 5 adatközlő beszélget egymással az adott témáról. A felvételeken nagy arányában két vagy három gyermek beszélget egymással (27. ábra). Mindösszesen 3,3 százalékukban szerepel négy, és 3,5 százalékukban öt beszélő.

27. ábra Beszélők számának felvételenkénti eloszlása a KorSzak Gyermeknyelvi Korpuszban

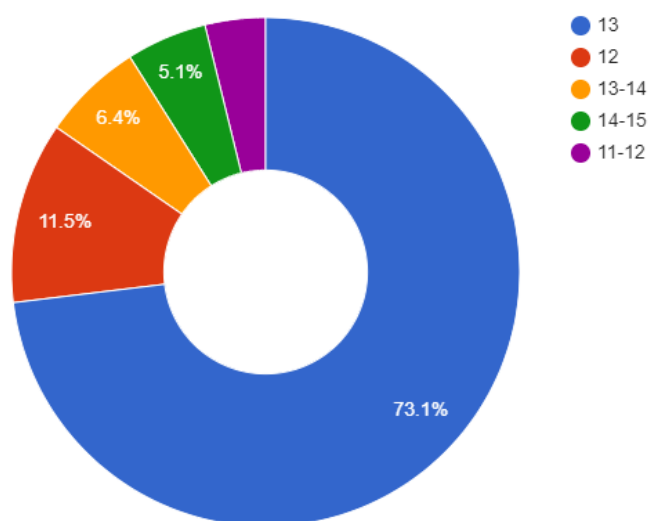


A huszonkilenc adatközlő gyermek magyarországi vidéki iskolákban tanuló osztálytársak és barátok, vagy csak barátok. A felvételek 64,1%-a Pest megyében, 35,9%-a Baranya megyében készült. Az első rögzítési fázisban, 2020-ban főleg Baranya megyei iskolások vettek részt. Ahogy már említettem, ekkor készültek a hosszabb felvételek, ezért, ha a tokenek számának lefedettségét nézzük, a korpusz 84,1%-a Baranya megyéből származik (Vermekei, 2023c, megjelenés alatt).

A legtöbb gyermek egynyelvű magyar anyanyelvű. Két gyermeknek (5. és 16. adatközlő) az édesapja nem magyar származású (holland és olasz), de otthon magyar nyelven kommunikálnak. Egyikük jelenleg egy nyelviskolában tanulja az olasz nyelvet, hogy a nagyszülőkkel tudjon kommunikálni. Az adatközlő gyermekek az általános iskola felső tagozatán angol (83%) vagy német (17%) nyelvet tanulnak idegen nyelvként.

A felvételek nagy része a 13 éves korosztállyal készült, de a korpuszban szereplő gyermekek életkori megoszlása 11–15 éves korig terjed. Az egyes felvételeken az osztálytársak életkorában lehet életérés, valamint vannak olyan adatközlők is, akik többször is adtak mintát a három év során, így az adatbázisban több életkorral is szerepelnek.

**28. ábra** A KorSzak GyermeKNyelvi Korpusz felvételein szereplő gyermekek életkori megoszlása



## 1.4 A KorSzak GyermeKNyelvi Korpusz összeállításának módja

### 1.4.1 A beszélt nyelvi szövegek átírása

A beszélt nyelvi korpuszok mennyisége arányaiban sokkal kevesebb az írott nyelvekhez képest, amelynek egyik indoka a rögzített beszéd átírásának kihívásai. A leírt szöveget számos módon, különböző hangsúllyal, hangszínnel és hangmagassággal lehet felolvasni, ezért nehéz írásban visszaadni azt, ami az adott pillanatban szóban elhangzott (Szirmai, 2005: 37). A jelen kutatás céljának megvalósulásához nincs a fenti tulajdonságok vizsgálatára szükség, ezért a felvételek átírásához elegendő volt egy beszédfelismerő program használata. Ehhez az Alrite mesterséges intelligenciára épülő beszédfelismerő programot alkalmaztam, amely a magyar nyelvre optimalizált, így 95%-os pontossággal képes dolgozni (Alrite, 2020). A szövegeket az automatikus átírás után lektoráltuk. A leiratokon nem kerültek jelölésre a beszélt nyelvre vonatkozó tulajdonságok (intonáció, hangsúly). Azok a megnyilatkozások, amelyek átfedték egymást, egymás utáni sorokban kerültek leírásra. A szövegek lektorálás után a későbbiekben (ld. III.2.1 fejezet) bemutatott Sketch Engine korpusznyelvészeti eszközöket tartalmazó online elérhető felületre kerültek fel.

## 1.4.2 Annotáció

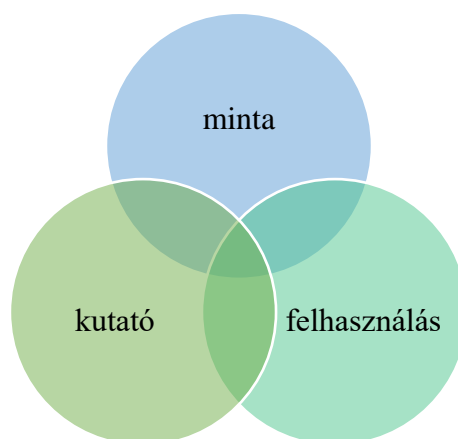
Minden olyan információt, amely megkönnyíti és meggyorsítja a korpuszban történő keresést, korpuszannotációnak nevezünk (Szirmai, 2005: 38). A korpuszok általában szófajok szerint annotáltak, vagyis szófaji címkézést [tagging] hajtanak végig rajtuk, illetve lemmákként jelöltek, de a felhasználási módjuk szerint speciális korpuszoknak minősülő korpuszok esetében előfordul másfajta annotáció is. A nyelvtanulói korpuszokban például gyakran megjelölik a tévesztéseket és a hibákat vagy a nyelvi szintet. Ahogy a Sketch Engine-t bemutató részben (ld. III.2.1 fejezet) is részletesen lesz szó róla, ma már a modern korpusznyelvészeti eszközök felületei automatikusan végzik el a szófajok szerinti címkézést, a tokenizációt és a lemmák megjelölését, így a felhasználóknak már csak a saját kutatásukat szem előtt tartva úgynevezett metaadatok megadását szükséges megtenniük, hogy a munkájuk még könnyebbé váljon. Ez a megoldás azért is jó, mert ahogy Szirmai (2005: 40) is írja a korpuszokon vizsgálatokat végezni kívánó kutatók mind más és más célt tűztek ki maguk elé, ezért nem mindig értenek egyet az eredeti annotációval. A szófajok és a lemmák szerinti annotáció viszont olyan általános jellegű tulajdonságokat jelöl, amely mindenki számára megfelelő. Ha a kutatóknak más információra is szükségük lenne a metaadatok megadásával vagy saját alkörpuszok összeállításával leszűkíthetik a kapott eredményeket a saját elképzeléseik szerint (vö. Stefanowitsch, 2020: 38–45). A KorSzak Gyermeknyelvi Korpusz metaadatai közé az alábbiakat vettem fel: életkor, lakóhely (megye), nem, felvétel készítésének éve, témakör, altémakör, beszélők száma, a szöveg típusa, tanult idegen nyelv. A metaadatokra még bővebben kitérünk a III.2.1 fejezetben.

## 2. A kutatás módszertana

A jelen doktori értekezésben egy korpuszalapú kutatás és annak eredményei kerülnek bemutatásra. A korpuszalapú kutatások tradicionálisan kvantitatív kutatómódszertant alkalmaztak, hiszen a korábbi ilyen jellegű vizsgálatok elsődleges célja az volt, hogy a kevésbé megbízható emberi intuícióval szemben (vö. Stefanowitsch, 2020: 15–17) megfogható, számokkal igazolható adatokat szolgáltatson a nyelvészek számára (Martinez, 2020: 211). Több tanulmány is íródott abból a célból, hogy meggyőzze olvasóközönségét ennek fontosságáról. Ilyen volt például a már említett (ld. II.6.5.1 fejezet) Koprowski (2005) kutatása, aki a tananyagokat vetette össze korpuszadatokkal, és bizonyította be, hogy azok nem tükrözik megfelelően a valódi nyelvhasználatot. Egy másik jelentős kutatás, Anderson (2007) vizsgálata szintén kvantitatív módszertanú, aki pedig az anyanyelvi beszélők intuícióit és a korpusz adatokat hasonlította össze. Ő azt bizonyította be, hogy az anyanyelvi beszélők segítség nélkül nem ítélik meg megfelelően a lexikai elemek gyakoriságát (Martinez, 2020: 211).

A természetes nyelvhasználat fontosságának hangsúlyozása érdekében korábban jellemzően tehát kvantitatív kutatások születtek. Ugyanakkor ahhoz, hogy megfelelően tudjuk értelmezni a kapott eredményeket, illetve, hogy a kapott eredmények hatékonyan legyenek felhasználva, elkerülhetetlen a kvalitatív kutatómódszertan alkalmazása.

**29. ábra** A korpuszalapú kutatások három elemének interakciója Martinez (2020: 214) alapján



Ahogy Martinez (2020: 215) fogalmaz, az ember, a kutató saját tanári és nyelvhasználati tapasztalatai adják az értékes kvalitatív részét a kutatásnak. Állítását egy érdekes példával Michael West *GSL* [General Service List] alapszókincs listájának történetével támasztja alá. A

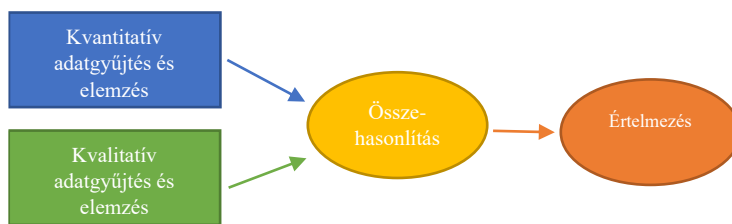
jelenleg 2000 szót tartalmazó listát eredetileg Edward Thorndike állította össze az 1910-es évek környékén azzal a céllal, hogy a pedagógusok tisztában legyenek azzal, hogy a gyermekek milyen szókészlettel találkoznak a tanulmányaik során. A lista összeállításához Thorndike egy olyan korpuszt használt, amelyet az általános iskolában használt szövegekből állított össze. Az eredeti lista azonban több okból is problémás volt: 20.000 szót tartalmazott, amely nehezen volt átlátható, a könyvek régiak voltak, és semmiféle szemantikai információt nem tartalmazott, így nehézséget okozott például a többjelentésű szavak esetében, hogy éppen mely jelentéssel szerepelnek a listában. Nem lehetett eldönteni például az angol *narancs* [orange] szóról, hogy mint szín vagy gyümölcs szerepel benne. A kritikák hatására a listát többen összefogva dolgozták át. A méretét 2000 szóra csökkentették, amelyek kiválasztásakor már nemcsak a gyakorisági mutatók adataira támaszkodtak, hanem kvalitatív adatokra is, amely a készítők saját oktatási tapasztalatából származott. Az újonnan készült lista több szótár (ld. Longman Dictionary of Contemporary English) alapjaként is szolgált és még a mai napig szoktak rá hivatkozni (Martinez, 2020: 214).

A fenti példából jól láthatjuk, hogy mennyire fontos a korpuszalapú kvantitatív kutatások során kapott eredmények értelmezése és használata előtt a kvalitatív módszertannal való összevetése. Érdemes tehát a korpuszalapú kutatásokat kevert kutatásmódszertannal [mixed-methods corpus research] végezni. A jelen értekezésben bemutatott kutatás is ilyen jellegű.

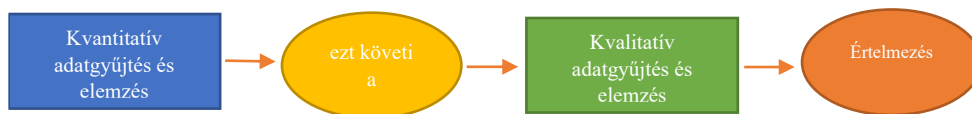
A kevert módszertanú kutatásoknak többféle változata létezik (vö. Dörnyei, 2007: 169), a fenti ábrán (30. ábra) a három legalapvetőbb tipológia látható. A konvergens párhuzamos tervezés [convergent parallel design] során a kvalitatív és a kvantitatív adatgyűjtést és elemzést párhuzamosan zajlik, majd kerül összehasonlításra és végül értelmezésre. A kutatás célja, a kvalitatív és kvantitatív eredmények közötti konvergencia vagy divergencia területeinek kimutatása. A magyarázó szekvenciális tervezés [explanatory sequential design] a kvantitatív adatgyűjtéssel és elemzéssel kezdődik, majd ezt követi a kvalitatív adatgyűjtés és elemzés, amely az értelmezéshez vezet. Ennél a kutatási típusnál meghatározzuk, hogy a kvantitatív kutatási rész eredményei közül melyeket kívánjuk tovább, kvalitatív módszerekkel elemezni. Végül pedig a harmadik tipológia, a feltáró szekvenciális tervezés [exploratory sequential design] a kvalitatív adatgyűjtéssel és elemzéssel kezdődik, majd a kvantitatív adatgyűjtésre és elemzésre épül, amely az értelmezéshez vezet. Ebben a típusban a kvalitatív eredményeket egy a kvantitatív részhez készülő új eszköz vagy taxonómia kidolgozására használják fel (Harvard, é.n.; Dörnyei, 2007: 163–176).

**30. ábra** A kevert módszertanú kutatások alap típusai (Harvard, é.n.) alapján

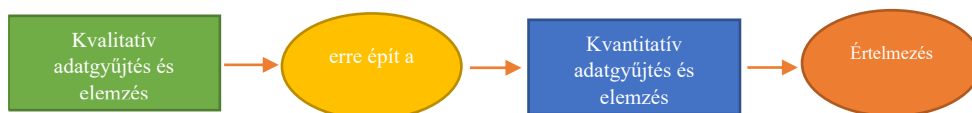
Konvergens párhuzamos tervezés:



Magyarázó szekvenciális tervezés:



Feltáró szekvenciális tervezés:



A jelen kutatás a második típusú, magyarázó szekvenciális tipológia alapján készült, tehát először a kutatás kvantitatív részére, később pedig a kvalitatív szakaszára került sor. A korpusznyelvészeti kutatásokat általában e szerint a tipológia szerint tervezik meg. Először gyakorisági vizsgálatokat alkalmaznak, hogy megállapítsák, mely területeket kell még részletesebben, kvalitatív megközelítéssel megvizsgálni (Hunston, 2022: 47). Ehhez hasonlóan történt a jelen kutatás tervezése is, amelynek kvantitatív és kvalitatív része szorosan összefügg egymással, valamint az is jellemző rá, hogy több szekvenciából áll (ld. IV.1 fejezet). Mivel a célom az adatközlő gyermekek nyelvhasználatában lévő formulaszerű egységek általános feltérképezése és néhány, a nyelvhasználatukra jellemző lexikai egység esetének, használati mintázatainak bemutatása, a KorSzak Gyermeeknyelvi Korpuszon több kvantitatív vizsgálatot is elvégeztem, az így kapott adatokat pedig többféleképpen vettem össze. A vizsgálatokhoz alapvetően két szoftvert alkalmaztam (Sketch Engine és Praat), amelyeknek a használatát a következőkben mutatom be.



## 2.1 Sketch Engine

A jelen doktori értekezésben bemutatott kutatáshoz legfőképpen az Adam Kilgarriff által fejlesztett Sketch Engine<sup>18</sup> elnevezésű korpusznyelvészeti eszközöket tartalmazó online felületet használtam, amelyre nemcsak azért esett a választásom, mert használata egyáltalán nem bonyolult, hanem azért is, mert modern, sokféle funkcióval rendelkezik és meglehetősen gyorsan fejlődik. A következőkben ennek a felületnek a részletes bemutatására kerül sor.

A Brnóban bejegyzett Lexical Computing nevű vállalat legnagyobb volumenű terméke a Sketch Engine, amelynek korpuszaiban összesen körülbelül hatvan milliárd szó található. Amellett, hogy a felhasználók több mint 90 nyelven 600 azonnal használható, írott és beszélt nyelvi szövegekből álló korpuszhoz férhetnek hozzá, saját dokumentumokat is feltölthetnek, vagy saját korpuszt építhetnek interneten található szövegekből. A Sketch Engine egy rendkívül könnyen használható, felhasználóbarát felület, gyorsan áttekinthető leírásokkal, könnyen megérthető videós oktatóanyagokkal, jelenléti és online továbbképzésekkel és e-mailen keresztüli műszaki támogatással. Legfőbb előnye azonban, amely még ennél is jelentősebb az az, hogy egy helyen biztosít szinte minden eszközt, amely egy korpuszalapú kutatáshoz szükséges, amíg korábban az ilyen jellegű kutatások során az egyes vizsgálatok elvégzéséhez különböző programok letöltésére és használatára volt szükség, mert a korpuszokat tároló felületeken a korpuszeszközöknek nem volt ilyen széles választéka elérhető.

### 2.1.1 Korpuszépítés a Sketch Engine felületén

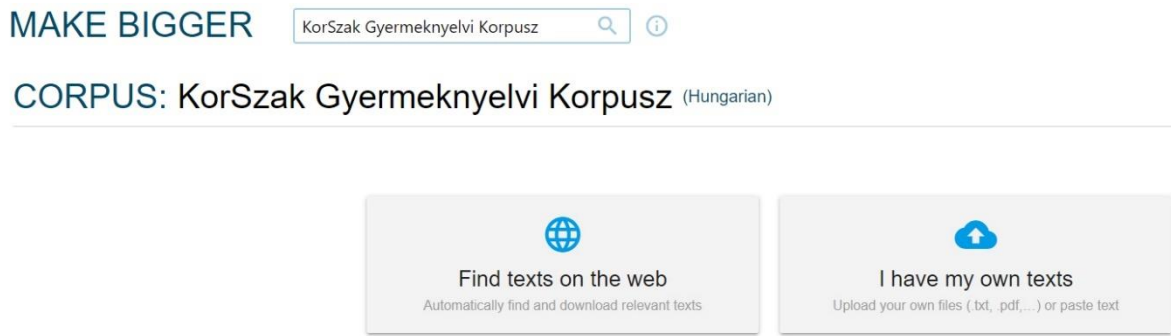
Korpuszalapú kutatásainkhoz egy a Sketch Engine felületén készen hozzáférhető, vagy egy saját magunk által összeállított korpuszt is használhatunk. A jelen értekezésben bemutatott kutatást egy újonnan épített korpuszon végeztem, ezért ebben a részben összefoglalom, hogyan is történt mindez, valamint, hogy milyen lehetőségeket biztosít a Sketch Engine a felhasználóknak egy korpusz építése folyamán.

A saját korpuszunk építésekor két módon járhatunk el (31. ábra): a kiválasztott fájlokat egyszerűen a fogd és vidd [drag and drop] technikával behúzhatjuk a felületre – ahogyan én is tettem – vagy internetes keresés segítségével gyűjthetjük össze.

---

<sup>18</sup> <https://www.sketchengine.eu/>

### 31. ábra Korpuszépítési lehetőségek a Sketch Engine felületén



A korpuszépítéssel kapcsolatban fontos megemlíteni, hogy a saját magunk által összeállított korpuszaink bármikor tovább bővíthetőek, akár egyetlen fájl feltöltésével is kiegészíthetjük őket. Ezzel szemben a már készen elérhető korpuszok bővítésére nincs lehetőségünk, ugyanakkor van mód a korpusz letöltésére, illetve a teljes korpusz egy részéből történő meghatározatlan számú és – az előfizetésünktől függően – méretű alkörpuszok építésére. A saját korpusz építésére visszatérve lényeges megjegyezni, hogy a Sketch Engine további előnye az olyan szoftverekkel szemben, mint például a Mike Scott féle WordSmith Tool, hogy a korpuszt alkotó fájlokat nemcsak text formátumban tudjuk alkalmazni, hanem a PDF formátumúaktól kezdve a táblázatokig szinte bármilyen kiterjesztésű fájlt feltölthetünk<sup>19</sup>. Ha pedig nem rendelkezünk saját fájlokkal, az interneten lévő szövegekből (32. ábra) összeállított korpuszok elkészítése könnyen és – internet szolgáltatásunk sebességétől függően – viszonylag gyorsan megvalósítható a felület keresésfunkcióinak alkalmazásával, amelyek automatikusan képesek számunkra releváns szöveget találni kulcsszavakra való keresés, szövegek weboldalakról történő kinyerése vagy linkek beillesztése után (Sketch Engine, é.n.a).

Miután a korpuszépítés kétféle lehetősége közül valamelyiket vagy éppen mindkettőt használtuk és rákattintottunk a COMPILE, vagyis összeállítás gombra, a Sketch Engine megkezdí a korpusz dokumentumainak feldolgozását, amely nemcsak a szövegek szerverre történő feltöltéséből áll, hanem az automatikus annotációjából (tokenizálás, lemmatizálás, címkézés), valamint a statisztikák elkészítéséből is. A Sketch Engine felületén több magyar korpusz is megtalálható. A magyar nyelvhez készült automatikus annotáció általában az esetek túlnyomó részében megfelelően működik.

<sup>19</sup> Támogatott kiterjesztések továbbá: csv, .doc, .docx, .htm, .html, .ods, .pdf, .tar.bz2, .tar.gz, .tei, .tgz, .tmx, .txt, .vert, .xlf, .xliff, .xml, .zip.

32. ábra Korpuszépítés a Sketch Engine felületén interneten található szövegek segítségével

MAKE BIGGER

CORPUS: KorSzak Gyermeknyelvi Korpusz (Hungarian)

---

← TEXTS FROM WEB

Input type  Web search  URLs  Website

e.g. shop store multiplex hypermarket

Type at least three words or phrases. Hit ENTER after each one.

Folder name

Web search settings ▾  
Denylist settings ▾  
Allowlist settings ▾  
Size restrictions ▾

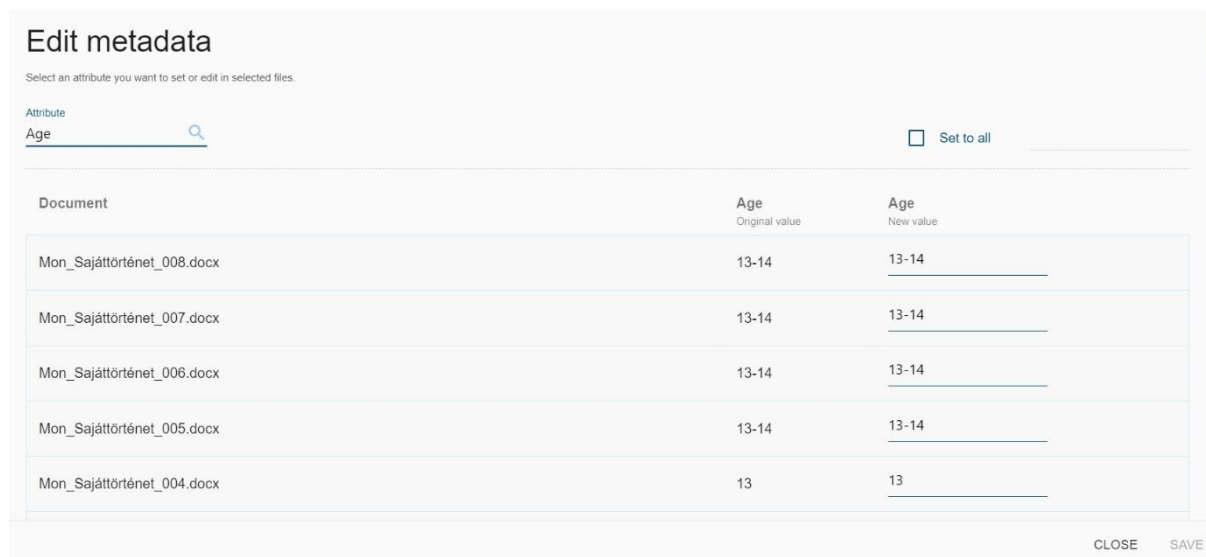
Compile when finished

A szófajok szerinti címkézéshez [PoS tags] a magyar emMorph (Novák, 2016) alapú morfológiai címkékészletet használja a program. A felmerülő hibák nagy része, amelyeket kizárólag a konkordanciasorok vizsgálatával lehetséges észrevenni, az azonos alakú szavak miatt alakulnak ki. Ilyen például a *vagy* kötőszó, amelyet rendszeresen a létige E/2 ragozási formájának tekint a program és a VAN lemmához sorol. Ennek az ellenkezője is előfordul néha a keresések során. Tehát, amikor egy szónak két külön lemmát határoz meg a program. Ilyen például a *jól érzem magam* kifejezésben az ige kétféle szótövének meghatározása (ÉREZ – \*ÉRZIK) vagy a *szerint* és a *szerintem* alakok lemmájának elkülönítése (SZERINT – \*ÉNSZERINT). Ezenkívül a keresések folyamán még két hibalehetőségre kell odafigyelnünk. A *nem* és az *is* szavakat minden esetben a határozószókhöz sorolja a program. A másik pedig, amikor a *hogy* kötőszóra lemmatizált állapotban keresünk rá, akkor kizárólag AHOGY alakban kapjuk meg. Az említett hibák viszonylag könnyen észlelhetők, nem jelentenek jelentős problémát egy korpuszalapú kutatás során, ha több keresési szekvenciában dolgozunk, és várhatóan a program további fejlesztése révén hamarosan kiküszöbölhetők lesznek.

Az automatikus annotációnak köszönhetően tehát nem szükséges a szövegek manuális annotálása, továbbá az automatikus lemmatizálás és címkézés optimális működése érdekében tapasztalataim szerint a legmegfelelőbb, ha szinte teljesen formázatlan szövegek kerülnek feltöltésre. A régi, illetve más módon annotált szövegek feltöltése esetén ugyanis a Sketch

Engine néhány funkciója nem elérhető (vö. II.6.2.1 fejezet: CHILDES korpusz a Sketch Engine-en). Ugyanakkor a könnyebb kezelhetőség és keresés érdekében, érdemes a korpuszunkat alkotó dokumentumokat feltöltés után metaadatokkal ellátni.

33. ábra Metaadatok megadása a Sketch Engine felületén



Edit metadata

Select an attribute you want to set or edit in selected files

Attribute  
Age   Set to all

| Document                   | Age<br>Original value | Age<br>New value |
|----------------------------|-----------------------|------------------|
| Mon_Sajáttörténet_008.docx | 13-14                 | 13-14            |
| Mon_Sajáttörténet_007.docx | 13-14                 | 13-14            |
| Mon_Sajáttörténet_006.docx | 13-14                 | 13-14            |
| Mon_Sajáttörténet_005.docx | 13-14                 | 13-14            |
| Mon_Sajáttörténet_004.docx | 13                    | 13               |

CLOSE SAVE

A metaadat, azaz olyan adat, amely egy másik adatról szolgál információval hasonló ahhoz, amely a számítógépünk használata közben minden fájlról keletkezik. Ilyen adat lehet például, hogy mikor készült egy fájl, vagy ki egy dokumentumnak a szerzője. Ehhez hasonlóan bármilyen információt megadhatunk a korpuszunkban lévő dokumentumokról is, az attribútumokat mi találhatjuk ki. Ezek az attribútumok lehetnek az adatközlők életkorai, a témakörök, a szöveg származási helye vagy bármi a kutatás során általunk hasznosnak tartott információ, amelyet ajánlatos a kutatás megtervezése után, de még a korpuszépítés előtt megtervezni és írásban is feljegyezni magunknak. A Sketch Engine ezeket az adatokat úgynevezett szövegtípusokként [text types] kategorizálja, amely azért hasznos, mert ezáltal lehetőségünk nyílik csak egy adott szövegtípust elemezni, vagy azon belül kereséseket végezni, valamint a teljes korpuszból alkorpuszokat is készíthetünk egy vagy több szövegtípus használatával (Sketch Engine, é.n.b).

A metaadatok megadása során csak két dologra kell nagy figyelmet fordítani: az attribútum és az érték megadásának formátumának egyezésére. A fenti képen (33. ábra) látható, hogy az életkort a szó angol megfelelőjével jelöltem. Mivel az *Age* szót nagybetűvel adtam meg a következő korpuszbővítések során az új dokumentumokban is ugyanilyen formában kell megadnom. Ehhez hasonlóan az attribútumokhoz tartozó értékek formátumának is hasonlóan

kell lennie ahhoz, hogy a szoftver a dokumentumot a megfelelő szövegtípusba tudja kategorizálni. Ha a fenti példát újra megnézzük, láthatjuk, hogy az életkort számmal adtam meg. Amennyiben a korpusz bővítése során az új dokumentumoknál ez máshogy történik (betűvel vagy például két szám esetén nem kötőjellel elválasztva), a program a már meglévő szövegtípus-kategóriák helyett újakba sorolja őket. Ugyanez történik, ha az attribútum eltérő formátumainak használatakor is.

A korpuszépitésével kapcsolatban az előbb említetteket figyelembe véve még annyit érdemes megjegyezni, hogy a metaadatokból származó előnyök kihasználása érdekében ajánlatos a korpuszunkhoz használt egyes szövegeket, a jelen esetben a felvételek leiratát külön dokumentumokba menteni és ugyanígy feltölteni. Egy dokumentumhoz ugyanis többféle attribútum rendelhető, de egy attribútumhoz csak egyféle érték adható meg. Új érték megadása esetén az előző törlődik. Ezért is elengedhetetlen, hogy a korpuszunk építése előtt tisztában legyünk a kutatási céljainkkal, valamint azzal, hogy elérésük érdekében milyen adatokra szeretnénk támaszkodni.

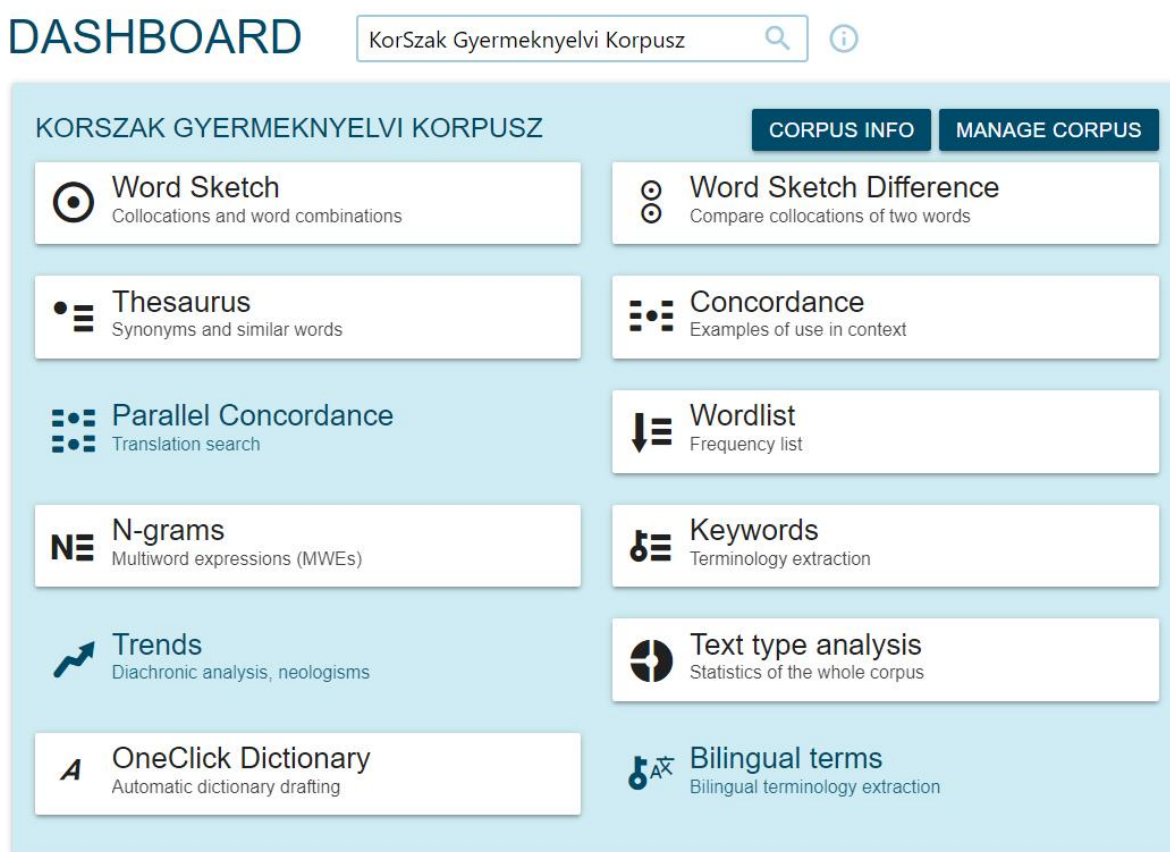
### **2.1.2 Korpuszeszközök a Sketch Engine felületén**

A Sketch Engine felhasználói a saját korpuszaikon a már rendelkezésre álló korpuszokhoz hasonlóan alkalmazhatják a felületen elérhető korpuszeszközök sokaságát. Ahogy korábban említettem ennek a felületnek az egyik nagy előnye, hogy szinte minden eszköz megtalálható rajta, amely egy korpusznyelvészeti kutatáshoz szükséges, valamint rendkívül gyorsan fejlődik és válik egyre könnyebben kezelhetőbbé és kidolgozottabbá.

A Sketch Engine kezdőképernyőjén (34. ábra) összesen tizenkét korpuszeszköz található, fehérrel jelölve azok, amelyek az éppen betöltött korpuszon használhatóak. A korpusz típusától függően vannak olyan eszközök, amelyek nem használhatóak bizonyos esetekben. A fenti ábrán látható, hogy a KorSzak Gyermekegyelvi Korpuszon három program nem működik: a paralel konkordancia, amely értelemszerűen csak paralel korpuszok esetén működik, az úgynevezett trendek keresésére szolgáló eszköz, amely egy diakronikus korpuszon működne csak, valamint a kétnyelvű kifejezések megtalálására kifejlesztett funkció, amely azért nem érhető el, mert a korpusz építésekor csak egy nyelvet, a magyart állítottam be. Az elérhető kilenc eszköz közül a jelen kutatás során nem használtam a szinonima szótárt [Thesaurus], ez a funkció még nem is működik tökéletesen, valamint a szótárkészítő [OneClick Dictionary] programot, amely a korpuszból gyakorisági listák, kulcsszavak vagy n-gramok alapján készít a felhasználó által megadott számú címszót és összegyűjti az azokhoz tartozó információt (szófej,

kollokációk, példamondatok, definíciók, szinonimák és fordítás) megalkotva ezzel egy szótár vázlatot, amely megkönnyíti a lexikográfusok munkáját. A kulcsszóvizsgálatok [Keywords] a tananyag szempontjából nagy jelentőséggel bírnak, de a jelen doktori értekezésben bemutatott kutatásban nem relevánsak, ezért nem kerülnek bemutatásra. A további öt korpuszeszköz, valamint azok több funkciójának alkalmazása azonban részét képezte a kutatásnak, ezért a következő fejezetekben, ezek kerülnek részletesebb bemutatásra.

34. ábra A KorSzak Gyermeknyelvi Korpusz kezdőképernyője a Sketch Engine felületén



### 2.1.3 A kutatás kvantitatív részéhez használt eszközök

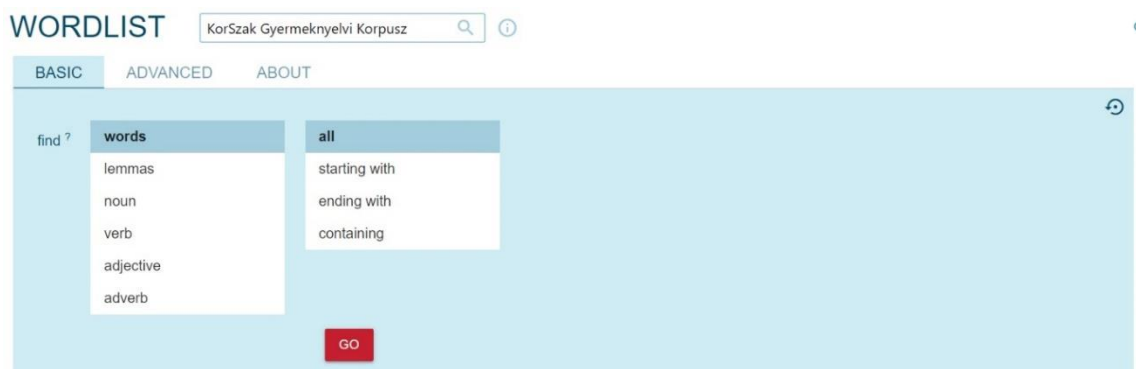
A korpusznyelvészeti kutatások fő kvantitatív vizsgálata a gyakorisági vizsgálat, amely azt mutatja meg, hogy milyen gyakran fordul elő egy lemma, szó vagy szókapcsolat az adott korpuszban. A vizsgálatok során készült listákon láthatjuk a keresett elem összes előfordulási típusát az előfordulási számokkal együtt. A gyakorisági vizsgálatokkal kapcsolatban két lényeges aspektusra kell figyelni. Egyrészt a vizsgálat során kapott gyakorisági mutatóknak

összehasonlíthatóknak kell lenniük. Másrészt pedig fontos megjegyezni, hogy az ilyen elemzések során kapott számadatok önmagukban nem informatívak, kontextusba helyezve értelmezhetőek megfelelően (Hunston, 2022: 82–84). A gyakorisági vizsgálatokat többféleképpen lehet elvégezni. A jelen kutatás során a későbbiekben részletesen is bemutatott N-grams és Wordlist eszközökkel dolgoztam. Az N-grams eszközzel egy adott számú (bigram = 2 elem, trigram = 3 elem, n-gramm = n elem) elemből álló sorozatokat lehet gyakoriságuk alapján megtalálni. Az n-grammokat más néven a Sketch Engine felületén Multiword Expression-nek (MWE) is nevezik, amelynek jelentését magyarul legjobban a *többelemből álló szekvenciák* fordítás adja vissza. Ugyanakkor fontos megjegyezni, hogy az N-grams eszköz segítségével nemcsak szavakra tudunk rákeresni, hanem más egységekre is, például lemmákra. Ez elmondható a másik eszközzel is a Wordlistről, amely szintén nemcsak a korpuszban szereplő leggyakoribb szavak kilistázására alkalmas, hanem például lemmákat, szóalakokat, és egyéb attribútumokat is képes megtalálni. A következő alfejezetekben ezeknek az eszközöknek a bemutatására kerül sor.

### 2.1.3.1 Wordlist – gyakorisági listák készítése

Az első korpuszeszköz a gyakorisági mutatók alapján szólistákat készítő Wordlist funkció. Az eszköz használata közben – mint ahogy általában a Sketch Engine többi eszközének használatakor – kereshetünk egyszerű, illetve összetett keresővel. Az egyszerű kereső alkalmazásakor beállíthatjuk, hogy szófajok, szavak, lemmák szerint keresünk-e, illetve megadhatunk olyan kritériumokat is, mint például, hogy milyen betűvel, betűsorral kezdődjön, végződjön vagy mit tartalmazzon a keresett elem (Sketch Engine, é.n.c).

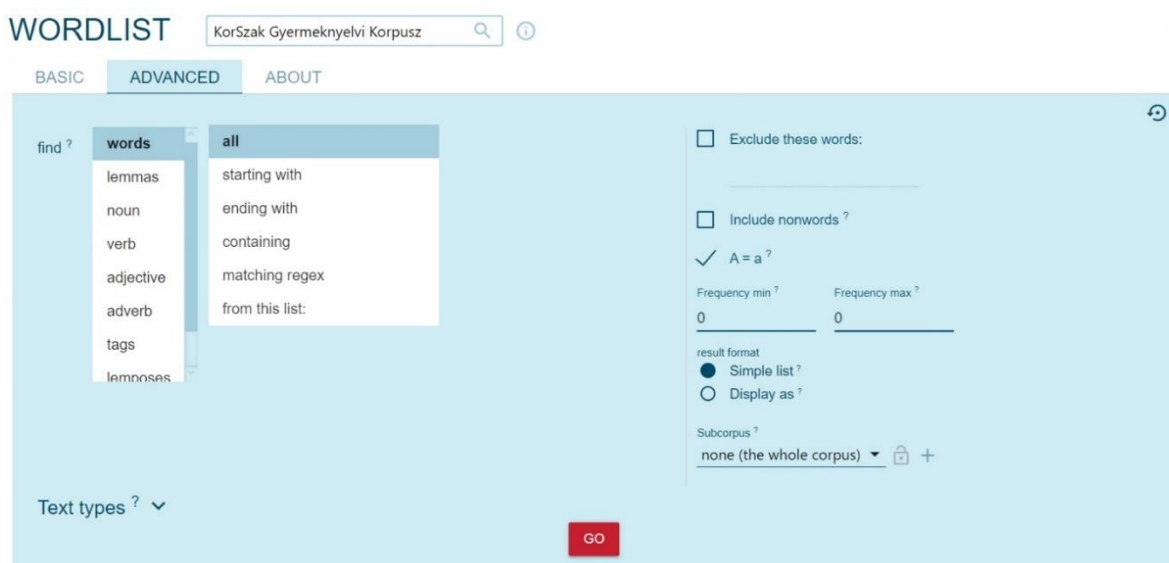
35. ábra A Wordlist egyszerű keresés funkciója



Az összetett keresés kiegészül még a címkék, lemos-ok (a lemma és a szófaj angol megfelelőjének rövidítése) szerinti kereséssel, a szövegtípusok, alkorpuszok meghatározásával, valamint speciális keresőkarakterek, jelek alkalmazhatóságával [Regular Expressions – RegEx], amelyek a konkordanciasorok készítése részben (III.2.1.4.1 fejezet) kerül bemutatásra részletesebben. Az összetett keresésben lehetőség van olyan szólisták beillesztésére is, amelyeknek a korpuszban való gyakoriságára vagyunk kíváncsiak. Ezenkívül meghatározhatjuk, hogy mit zárjon ki és mit foglaljon bele a listába a program, a kis és nagybetűk használatát, a gyakorisági minimumot és maximumot, valamint, hogy mely attribútuma alapján rendeződjön a lista (Sketch Engine, é.n.c).

Mind a Wordlist, mind a következőkben bemutatásra kerülő N-gram eszközzel történő keresés során a magyar nyelvű korpuszok esetében ajánlott a lemmákra való keresés beállításának használata, amikor azt szeretnénk megtudni, hogy egy-egy szó összesen hány előfordulással szerepel a korpuszban, tehát nemcsak egy ragozott szóalak gyakoriságára vagyunk kíváncsiak. Ahogy Durrant (2013: 32) is fogalmaz a török nyelvvel kapcsolatos kutatásában (ld. még III.3.1.2 fejezet), mivel az agglutináló nyelvekben a rendkívül sokféle végződés miatt viszonylag alacsony gyakorisági mutatói vannak egy-egy szóalaknak, az ilyen jellegű korpuszokkal való munka első lépésének a lemmatizálásnak kell lennie. A lemmatizálást a Sketch Engine automatikusan elkészíti a korpusz feltöltése után és a korábban említett (III.2.1.1 fejezet) néhány esetleges hibán kívül a lemmánkénti keresés beállításával megtalálja az összes olyan szóalakot, amely az adott szótöből származik.

36. ábra A Wordlist összetett keresés funkciója





A Wordlist eszköz a tokenek szintjén működik és összesen hat gyakorisági mutatót lehet megismerni a használata által. Megmutatja

1. a listában szereplő elemek gyakoriságát a korpuszban,
2. az elemek gyakoriságát per egymillió szó,
3. a dokumentum gyakoriságot [Document Frequency - DOCF], amely kimutatja, hogy hány feltöltött dokumentumban jelennek meg a listában szereplő elemek,
4. a relatív dokumentum gyakoriságot [Relative Document Frequency], amelyből megtudhatjuk, hogy a dokumentumok hány százalékában jelennek meg a listában szereplő elemek,
5. a redukált gyakoriságok átlagát [Average Reduced Frequency - ARF], amely nem veszi figyelembe a listában szereplő elemek közeli előfordulásait, hanem a teljes korpuszban eloszló megjelenésükből készít egy átlagot,
6. valamint az átlagos logaritmikus távolsági gyakoriságot [Average Logarithmic Distance Frequency - ALDF], amelynek figyelembevétele megakadályozza, hogy az eredményeket túlzottan befolyásolja egy token nagy koncentrációja a korpusz egy vagy több kis részében. Ha a token egyenletesen oszlik el a korpuszban, az ALDF és az abszolút gyakoriság hasonló vagy azonos értékeket kap (Sketch Engine, é.n.c).

37. ábra A Wordlist funkció során alkalmazható gyakorisági mutatók

The screenshot shows the 'WORDLIST' interface for the 'KorSzak Gyermeeknyelvi Korpusz'. The word 'word' is highlighted, with 9,664 items and a total frequency of 55,143. Below this is a 'VIEW OPTIONS' panel with several toggle switches:

| Option                  | Toggle State |
|-------------------------|--------------|
| Single column ?         | Off          |
| Show line numbers ?     | On           |
| Show bars ?             | Off          |
| Frequency ?             | On           |
| Frequency per million ? | Off          |
| Document frequency ?    | Off          |
| Relative DOCF ?         | Off          |
| ARF ?                   | Off          |
| ALDF ?                  | Off          |

A gyakorisági listákat CSV, XLSX, XML, valamint PDF formátumban tölthetjük le. Adatvédelmi okokból a készen elérhető korpuszokból maximum egy 1000 itemből álló listát tölthetünk le. A saját korpuszunk esetében nincs hasonló korlátozás (Sketch Engine, é.n.j).

### 2.1.3.2 N-grams – több elemből álló szekvenciák keresése

Az N-grams eszköz tokenek sorozatából készít gyakorisági listákat (Sketch Engine é.n.d). Ez a korpuszeszköz is lehetőséget biztosít a felhasználónak az egyszerű és az összetett keresésre. Az egyszerű keresés során csak azt szükséges megadnunk, hogy milyen hosszúak legyenek a találatok között szereplő n-gramok. Kettőtől hat elemig választhatjuk ki, hogy hány elemből álljanak a szekvenciák.

38. ábra Az N-grams egyszerű keresés funkciója



Az összetett keresésnél (39. ábra) a találatok között szereplő n-gramok hossza mellett többféle szűrőt is beállíthatunk. N-gramokat bármely attribútum alapján listázhatunk (lemma, szó, szófaj, lepos, címke), beállíthatjuk a minimum és a maximum gyakorisági mutató értékét, kereshetünk szövegtípusokban vagy alkorpuszokban. Megadhatjuk, hogy megjelenítse-e az álszavakat, a beágyazott n-gramokat, vagyis olyan n-gramokat, amelyek hosszabb n-gramok részét képezik, valamint, hogy figyelembe vegye-e a kis és nagybetűk közti különbséget. Bizonyos elemeket módunkban áll ki is zárni a találatok közül. További kritériumként megadhatjuk, hogy mely szóval vagy betűvel, betűkombinációval kezdődjön vagy végződjön a szekvencia, illetve mely szót vagy betűt, betűkombinációt tartalmazza. Ezenkívül ennél a keresésnél is van lehetőségünk [RegEx] jeleket használni, valamint azt az opciót is kiválaszthatjuk, hogy csak a korpuszunkban kulcs elemként funkcionáló szekvenciák kerüljenek a listába (Sketch Engine, é.n.d).

39. ábra Az N-grams összetett keresés funkciója

The screenshot shows the 'N-GRAMS' search interface. At the top, there is a search bar containing 'KorSzak Gyermeknyelvi Korpusz' and a magnifying glass icon. Below the search bar are three tabs: 'BASIC', 'ADVANCED', and 'ABOUT'. The 'ADVANCED' tab is selected. The interface is divided into several sections:

- N-gram length:** A row of buttons for 2, 3, 4, 5, and 6. The '3' button is highlighted.
- Attribute:** A dropdown menu currently showing 'word'.
- Frequency min:** A text input field containing '2'.
- Frequency max:** A text input field containing '0'.
- Subcorpus:** A dropdown menu showing 'none (the whole corpus)' with a lock icon and a plus sign.
- Checkboxes:** Several options are listed with checkboxes: 'Nest n-grams?', 'Include nonwords?', 'A = a?', 'Exclude these words:', and 'Key n-grams?'.
- Additional criteria:** A dropdown menu is open, showing options: 'all', 'starting with letters', 'ending with letters', 'containing letters', 'starting with word', 'containing word', 'ending with word', and 'matching regular expression'.
- Text types:** A dropdown menu with a question mark and a downward arrow.
- GO button:** A red button with the text 'GO' in white.

Az így elkészült tokenalapú szekvenciák listáját tehát gyakorisági mutatók alapján kapjuk meg, amelyek a már korábban említésre kerülő normál gyakoriság, a gyakoriság per egymillió szó, a dokumentum gyakoriság, a relatív dokumentum gyakoriság, a redukált gyakoriságok átlaga, valamint az átlagos logaritmikus gyakoriság átlaga. A kapott listát a Wordlist eszközhöz hasonlóan többféle formátumban tölthetjük le (CSV, XLSX, XML, PDF).

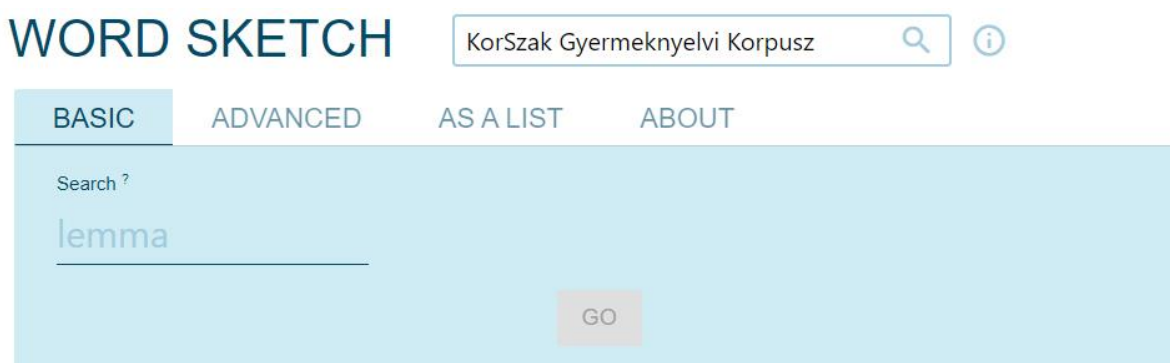
Az N-grams korpuszeszköz használatának kétféle korlátozása van: egyrészt nagy korpuszok esetén csak a korpusz egy részét képes átvizsgálni, amelyet egymilliárd tokenben maximalizáltak, valamint a letöltéssel kapcsolatos adatvédelmi limitálás, amely a készen használható korpuszok esetén egy ezer itemből álló listában került meghatározásra (Sketch Engine, é.n.j).

### 2.1.3.3 Word Sketch – kollokációk és szókombinációk keresése

A Word Sketch elnevezésű eszköz arra használható, hogy megtalálja a felhasználó által kiválasztott szó kollokációit, a környezetében gyakran megjelenő szavakat, valamint azokat a grammatikai mintázatokat, amelyben általában használják azt. A keresés eredményeit ezután grammatikai kapcsolatuk, gyakoriságuk, tipikusságuk alapján kategóriákba rendezi (Sketch Engine, é.n.f). Ha a keresett szó például egy ige, akkor ilyen kategória lehet az ige mellett gyakran megjelenő tárgyak kategóriája a legtipikusabbtól a kevésbé jellemzőkig. A Word Sketch a kollokációkat olyan kritériumok alapján találja meg, mint a szófaji címkék vagy a szavak közötti távolság, amelyben a magyar nyelvre is optimalizált úgynevezett Word Sketch Grammar eszköz nyújt segítséget. Ennek a funkciónak a jelentősége, hogy a használatával egy nagyobb szószámú korpuszban is könnyen megtalálhatjuk a tipikus kollokációkat, majd kiválaszthatjuk azokat, amelyekkel konkordanciasorokban bővebben is szeretnénk foglalkozni (Sketch Engine, é.n.f).

A keresés ennek az eszköznek a használata közben is többféleképpen történhet. Az alap vagy egyszerű keresés (40. ábra) során csak egy lemmát szükséges megadnunk.

40. ábra A Word Sketch egyszerű keresés funkciója



WORD SKETCH

KorSzak Gyermeknyelvi Korpusz

BASIC    ADVANCED    AS A LIST    ABOUT

Search ?

lemma

GO

A haladó, vagy összetett keresés során (41. ábra) az általunk kiválasztott kritériumok alapján szűkíthetjük le a találatokat. Ennél a keresésnél is egy lemmát szükséges megadnunk, de beállíthatjuk annak szófaját, hogy milyen szövegtípusban, illetve melyik alkorpuszban keressen az eszköz, valamint a minimum gyakorisági mutatót, a minimum tipikussági mutatót és egy paralel korpusz kiválasztásával a találatok fordításait is kilistázhatjuk.

41. ábra A Word Sketch összetett keresés funkciója

WORD SKETCH

KorSzak Gyermeeknyelvi Korpusz

BASIC **ADVANCED** AS A LIST ABOUT

Search ?

lemma

Part of speech ?

- auto**
- noun
- verb
- adjective
- adverb
- other

Subcorpus ?

none (the whole corpus) ▾ 🔒 +

Minimum frequency ?

auto

Minimum score ?

0

Translate ?

Select corpus

GO

A harmadik keresési lehetőség [As a list] az előzőtől annyiban tér el, hogy más formátumban közli az eredményeket, illetve új szűrő megadásával tovább szűkíthetjük a találatokat. Ez a funkció ugyanis szűrhető a vizsgált csomópont [node, keyword] vagy annak kollokációja alapján. Ilyen szűrés lehet, hogy a kollokáció milyen betűvel, betűkombinációval kezdődjön vagy végződjön, mit tartalmazzon, valamint megadhatunk kifejezéseket, amelyeket tartalmaznia kell vagy éppen ki kell zárnia a programnak, illetve akár egy saját szólistát is beírhatunk, amelynek elemeire rá kell keresnie a korpuszban.

## 42. ábra A Word Sketch második összetett keresés funkciója

**WORD SKETCH** KorSzak Gyermeknyelvi Korpusz

BASIC ADVANCED **AS A LIST** ABOUT

Frequency min:  Frequency max:  Subcorpus: none (the whole corpus)

A = a  Exclude these words:

Key collocations ?

all  
starting with  
ending with  
containing  
matching regex

**from this list:**

[About collocations as list](#)

⚠ This may take several minutes, especially on large corpora of billions of words.

GO

A Sketch Engine tartalmazza ennek a korpuszeszköznek egy kiterjesztését [Multiword Sketch] is, amellyel azokban a mondatokban szereplő szavak kollokációra tudunk rákeresni, amelyek kapcsolatban állnak a keresett szavakkal (Sketch Engine, é.n.f). A lenti két ábrán látható, hogy a KUTYA lemma milyen kollokációkban jelenik meg a KorSzak Gyermeknyelvi Korpuszban, valamint az, hogy a KUTYA és a NAGY kollokáció hogyan található meg együtt a korpuszban.

## 43. ábra A Word Sketch keresési eredményei (keresett lemma: KUTYA)

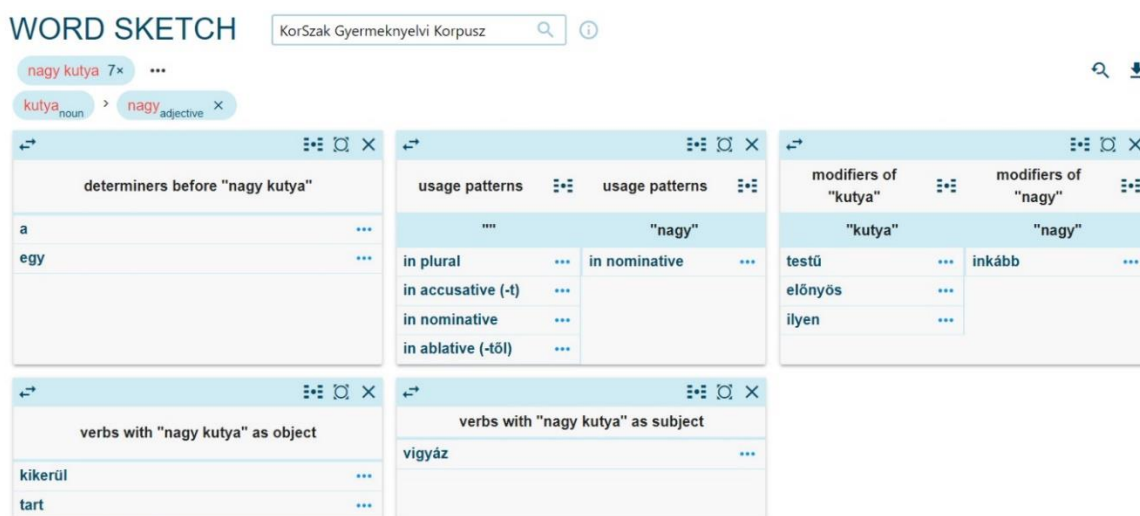
**WORD SKETCH** KorSzak Gyermeknyelvi Korpusz

kutya as noun 266x

determiners before "kutya" usage patterns

| modifiers of "kutya"  | verbs with "kutya" as object   | verbs with "kutya" as subject  | "kutya" and/or ...                             | "kutya" is a ...  | possessors of "kutya"                  |
|---|--|--|--|---|--|
| <p><b>hires</b><br/>híresebb kutya</p> <p><b>nagy</b><br/>nagy kutyát</p> <p><b>olyan</b><br/>olyan kutya</p> <p><b>jellel</b> Multiword Sketch<br/>kutya jellemző</p> <p><b>testű</b><br/>testű kutyákat</p> <p><b>fajta</b><br/>fajta kutya</p> <p><b>ilyen</b><br/>ilyen nagy kutyát</p> <p><b>kicsi</b><br/>kisebb testű kutyákat</p> | <p><b>szeret</b><br/>szeretnék kutyát</p> <p>kutya + nagy</p> <p>kutya + nagy</p> <p>kutya + nagy</p> <p>nagy<br/>amikor meglát egy kutyát, akkor</p> <p><b>gondoz</b><br/>gondozni egy kutyát</p> <p><b>elvisz</b><br/>a kutyát elvisszük</p> <p><b>veszik</b><br/>veszünk egy kutyát</p> | <p><b>reagál</b><br/>kutya úgy reagált</p> <p><b>gondol</b><br/>gondolhat a kutya</p> <p><b>van</b><br/>kutya volt</p> | <p><b>macska</b><br/>macskám és két kutyám</p> | <p><b>háziállat</b><br/>kutya egy ideális háziállat</p> | <p><b>falu</b><br/>falu kutyajának</p> |

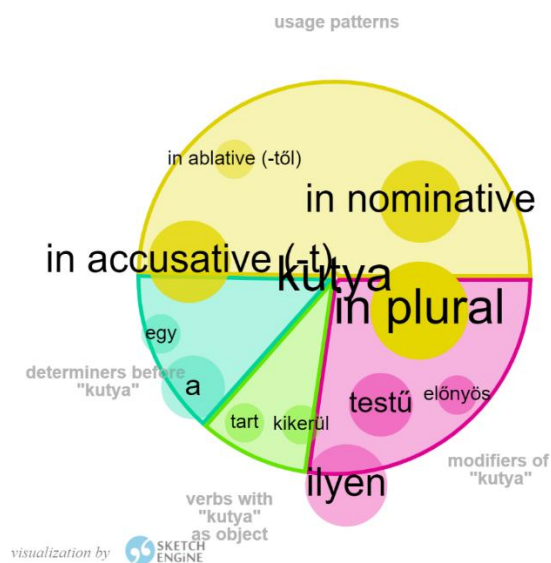
**44. ábra** A Word Sketch keresési eredményei (keresett lemmák: KUTYA + NAGY)



A fenti, második ábrán (46. ábra) látható, hogy a keresett kollokáció grammatikai mintázataival kapcsolatban több információt is megtudhatunk. Ilyen például, hogy milyen szóalakban jelenik meg (egyes, többes szám) vagy milyen igék társaságában található a korpuszban. A kapott eredményeket tovább szűkíthetjük az egyes táblázatok elrejtésével, illetve a három pontra kattintva további lehetőségeket is találhatunk, ilyen a konkordanciasorok mutatása.

A Sketch Engine funkciói között megtalálható több vizualizációs eszköz is, amelyek segítségével a találati eredményekből jól áttekinthető, a felhasználók által alakítható ábrákat készíthetünk. Az így szerkesztett ábrák SVG, PNG és PDF formátumban le is tölthetők.

**45. ábra** A Word Sketch találatainak vizualizációja (keresett lemma: KUTYA)



A Word Sketch eszköznél egy kördiagramhoz hasonló, bár annál összetettebb vizualizáció készítésére van lehetőségünk (45.ábra), amelynél a középponttól való távolság a tipikussági mutatót jelenti. A középponthez képest minél bentebb található a megjelenített elem, annál tipikusabb a kollokáció vagy a grammatikai mintázat. A belső kis körök nagysága a gyakoriságra utal, a színe pedig a grammatikai kapcsolatra. A betűméret a kollokációk számát jelöli. Minél nagyobb, annál több kollokációt tartalmaz a korpusz.

### 2.1.3.4 Word Sketch Difference – két nyelvi elem kollokációk általi összehasonlítása

A Word Sketch Difference nevű korpuszeszközzel lehetőségünk nyílik arra, hogy két nyelvi elemet a kollokációik alapján hasonlítsunk össze (Sketch Engine, é.n.g).

46. ábra A Word Sketch Difference egyszerű kereső funkciója



WORD SKETCH DIFFERENCE

KorSzak Gyermekegyelvi Korpusz

BASIC ADVANCED ABOUT

First lemma? Second lemma?

jön megy

GO

Háromféleképpen tehetjük meg ezt: egyrészt megnézhetjük, hogy mi a különbség két lemma között, másrészt összehasonlíthatjuk egy lemma két szóalakjának használatát, harmadrészt pedig egy lemma két külön alkorpuszban történő megjelenítésének jellemzőit is megismerhetjük (Sketch Engine, é.n.g).

A korpuszeszköz, amely az előző részben bemutatott Word Sketch funkcióhoz hasonlóan működik, eredményeinek kilistázásakor könnyen értelmezhetően, színekkel jelölve jeleníti meg a különbségeket (48. ábra). Ennek jelentősége, hogy a Word Sketch Difference alkalmazásával megtudhatjuk a kollokációk közti, valamint a jelentés és a használat közötti különbségeket különböző regiszterekben, témakörökkel kapcsolatosan és a kutatásunk céljának megfelelő egyéb aspektusok alapján (Sketch Engine, é.n.g).



47. ábra A Word Sketch Difference összetett kereső funkciója

## WORD SKETCH DIFFERENCE

KorSzak Gyermeknyelvi Korpusz

BASIC **ADVANCED** ABOUT

compare ?

Lemmas  
 Word forms  
 Subcorpora

First lemma ? \_\_\_\_\_ Second lemma ? \_\_\_\_\_

Part of speech ?

**auto**

noun  
 verb  
 adjective  
 adverb  
 other

Minimum frequency ?  
 auto \_\_\_\_\_

GO

48. ábra A Word Sketch Difference keresési eredményei (keresett lemmák: JÖN – MEGY)

## WORD SKETCH DIFFERENCE

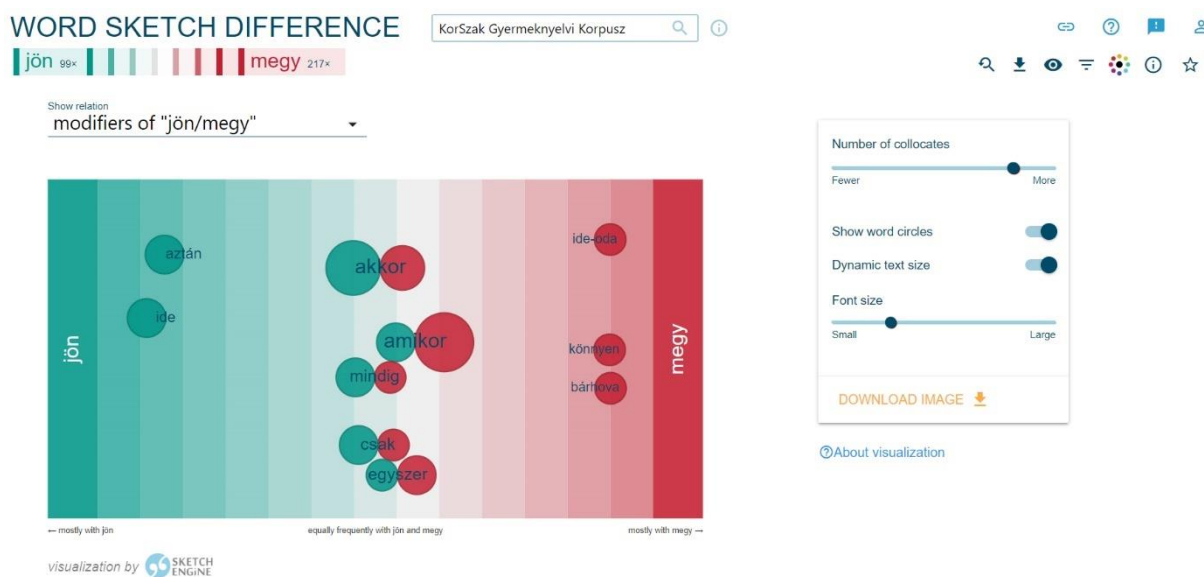
KorSzak Gyermeknyelvi Korpusz

jön 99x | megy 217x

pronominal subjects of "jön/megy"

| objects of "jön/megy"     | subjects of "jön/megy" | pronominal objects of "jön/megy" | modifiers of "jön/megy" |
|---------------------------|------------------------|----------------------------------|-------------------------|
| család 0 2 ...            | őv 4 0 ...             | mi 4 0 ...                       | ide 3 0 ...             |
| út 0 2 ...                | csilla 2 0 ...         | az 2 4 ...                       | aztán 3 0 ...           |
| kérdés 0 2 ...            | felvétel 0 2 ...       | ő 0 3 ...                        | mindig 3 2 ...          |
| tó 0 2 ...                | tévé 0 2 ...           | ők 0 4 ...                       | csak 3 2 ...            |
| osztálykirándulás 0 2 ... |                        |                                  | akkor 6 4 ...           |
| verseny 0 2 ...           |                        |                                  | egyszer 2 3 ...         |
| hátér 0 2 ...             |                        |                                  | amikor 3 7 ...          |
| Budapest 0 2 ...          |                        |                                  | könnyen 0 2 ...         |
| bozót 0 2 ...             |                        |                                  | ide-oda 0 2 ...         |
| ideg 0 2 ...              |                        |                                  | bárhova 0 2 ...         |
| iskola 0 3 ...            |                        |                                  | oda 0 4 ...             |
| erdő 0 4 ...              |                        |                                  | valahova 0 5 ...        |

**49. ábra** A Word Sketch Difference találatainak vizualizációja (keresett lemmák: JÖN – MEGY)



Az előző eszközhöz hasonlóan a Word Sketch Difference is tartalmaz vizualizációs funkciót (49. ábra), amelyet a felhasználók személyre szabhatnak. A skálához hasonlító diagram készülhet minden, vagy bármelyik eredménytáblázat alapján, amelyiket kiválasztjuk. A belső körök mérete ennél a vizualizációnál is a gyakorisági mutatókra utal, a horizontális elrendezés szerint pedig a kollokáció ahhoz a csomóponthoz van közelebb, amellyel erősebb és tipikusabb a kapcsolata.

A korábban bemutatott Word Sketch és a Word Sketch Difference is egy az együttes előfordulások azonosítására szolgáló statisztikai mérésére használt számítás elvégzésével (50. ábra), a logDice (vö. Rychlý, 2008) segítségével állapítja meg, hogy mely elemek ezek (Kilgarriff, 2015: 2).

**50. ábra** A Word Sketch és a Word Sketch Difference korpuszeszközök által használt logDice képlet (Kilgarriff, 2015: 2)

### logDice

$$14 + \log_2 \text{Dice} \left( \frac{\|w_1, R, w_2\|}{\|w_1, R, *\|}, \frac{\|w_1, R, w_2\|}{\|*, *, w_2\|} \right) = 14 + \log_2 \frac{2 \cdot \|w_1, R, w_2\|}{\|w_1, R, *\| + \|*, *, w_2\|}$$

A számítás során egy háromrészből álló szekvencia (első szó – grammatikai kapcsolat – második szó) gyakoriságát vizsgálják. A fenti képletben lévő elemek a következők:

|                      |   |
|----------------------|---|
| <b>  w1, R, w2  </b> | A háromrészből álló szekvencia előfordulásainak száma.  |
| <b>  w1, R, *  </b>  | Az első szó bármely második szóval való nyelvtani kapcsolatban lévő előfordulásának száma.  |
| <b>  *, *, w2  </b>  | A második szó bármely első szóval való bármely nyelvtani kapcsolatban való előfordulásának száma.   |
| <b>  *, *, *  </b>   | bármely első szó bármely második szóval való bármely nyelvtani kapcsolatban való előfordulásainak száma, vagyis az összes hármas szekvencia száma, amely a korpuszban megtalálható. |

(Kilgarriff, 2015: 1)

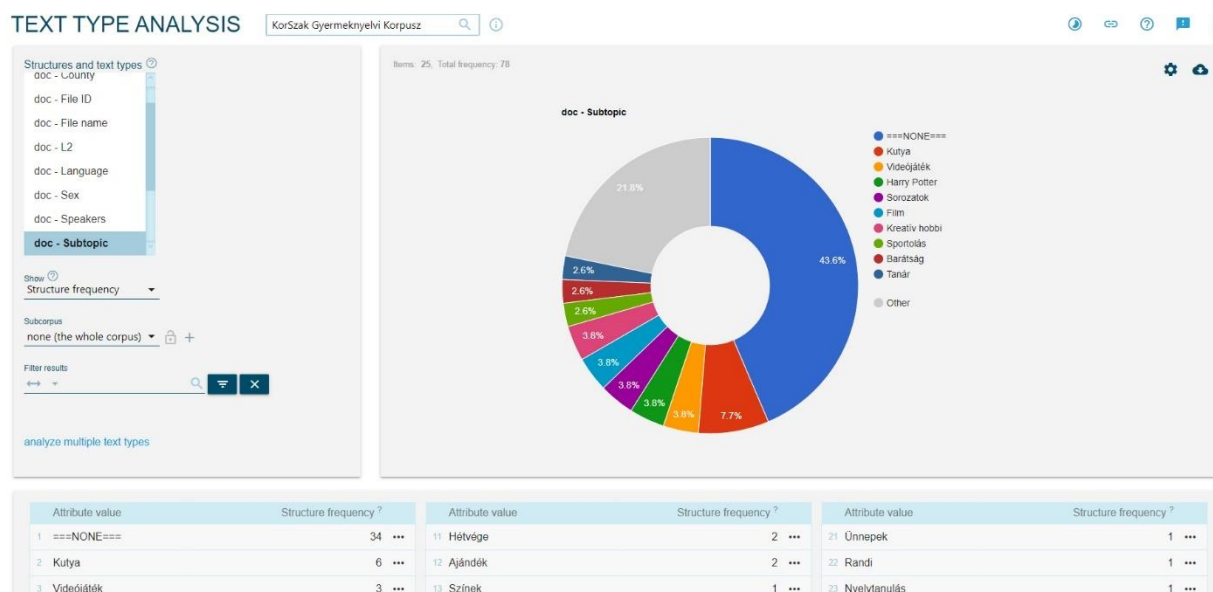
Az algoritmus először a csomóponthoz kapcsolódó lemmákat keresi meg, majd az összes kapcsolódó lemmát megvizsgálja. Azok a lemmák, amelyek a legtöbb esetben vagy kizárólag csak az adott csomóponttal kombinálódnak, bekerülnek a listába, míg azok, amelyek ennél rugalmasabbak, kizárásra kerülnek (Sketch Engine, é.n.g).

A Word Sketch és a Word Sketch Difference eredményeit CSV, XLSX, XML, valamint PDF formátumban menthetjük le a számítógépünkre. Adatvédelmi okokból saját korpusz esetén minden grammatikai kapcsolatnál 200 elem jeleníthető meg és tölthető le. Készen elérhető korpusznál ez a szám maximum 100 megjelenítés kapcsolatonként, valamint összesen 1000 kollokáció tölthető le belőlük (Sketch Engine, é.n.j).

### **2.1.3.5 Text type analysis – a teljes korpusz statisztikájának áttekintése a szövegtípusok alapján**

Az ebben a kategóriában utolsóként bemutatásra kerülő korpuszeszköz, a Text type analysis, amely a korpuszban lévő dokumentumok metaadatainak megadásakor összegyűjtött információkat felhasználva készít statisztikákat a teljes korpuszunkról. A megadott metaadatok szerint képes megadni a struktúrák (mondatok, bekezdések, dokumentumok) szerinti gyakoriságot (Sketch Engine, é.n.h), valamint a token lefedettséget, amelyet kördiagram, valamint lista formájában jelenít meg. Alkorpuszokat és egyszerre több szövegtípust is elemezhetünk vele. A kördiagramot pedig CSV, XLSX, XML, valamint PNG formátumban tölthetjük le.

51. ábra A Teyt type analysis kezdőoldal



A fenti korpuszeszközök segítségével kapott elsődleges eredményeket kvalitatív módon vizsgáltam tovább. A kvalitatív vizsgálatokhoz használt eszközökről a következő részben számolok be.

### 2.1.4 A kutatás kvalitatív részéhez használt eszközök

A kutatás kvantitatív első szakasza után következett a kvalitatív módszereket alkalmazó második része, amelyhez főként a konkordanciasorok elkészítéséhez használható Concordance elnevezésű – szintén a Sketch Engine-en megtalálható – eszközt alkalmaztam. Egy konkordanciakeresés segítségével megtalálhatjuk egy adott szó, kifejezés összes előfordulását egy korpuszban, illetve láthatjuk annak minden egyes előfordulását a közvetlen kontextusában. A konkordanciakeresés során nemcsak teljes szavakra, hanem a gyakorisági vizsgálatokhoz hasonlóan lemmákra, szófajokra és egyéb attribútumokra is rákereshetünk. Az így nyert konkordanciasorok vizuális elrendezésükben is támogatják a nyelvészek munkáját: a keresett elem [node, keyword] mindig középen egymás alatt, a Sketch Engine felületén piros színnel jelölve jelenik meg. A középen lévő elemtől jobbra és balra meghatározott számú karakter áll, ezek általában nem teljes mondatok (Szirmai, 2005: 102; Hunston, 2022: 47).

A konkordanciasorok vizsgálata két szempontból meghatározó a korpuszalapú kutatásokban. Egyrészt azért, mert ahogy fent említésre került a segítségükkel megismerhetjük

a vizsgált elemekhez tartozó használati mintázatokat. Másrészt, ezzel szoros összefüggésben, megfigyelhetjük a vizsgált elem szemantikai asszociációit (Hoey, 2005). A hasonló jelentésű szavak ugyanis hasonló mintázatban jelennek meg, a mintázatban való eltérés pedig gyakran jelenti a jelentésben való eltérést is (Hunston, 2022: 62–63). Sinclair (1991) is ehhez hasonlókat fogalmazott meg korábban. Az általa szemantikai prozódianak nevezett jelenség megfigyeléséhez véleménye szerint fontos látni, hogy a vizsgált szót tipikusan hogyan és milyen kontextusban használják a kompetens nyelvhasználók. Szerinte (1991: 104) a szavak jelentését többnyire azok a minták vagy frazeológiák különböztetik meg, amelyekben jellemzően előfordulnak, ezért érdemes a konkordanciasorokat jelentésenként csoportosítani (Hunston, 2022: 51).

#### 2.1.4.1 Concordance – konkordanciasorok készítése

Sinclair (2004: 30) szerint a nyelv magasan szervezett, ezért egy minta jelentőségét úgy tudjuk megállapítani, hogy ha összehasonlítjuk a gyakoriságát a megjelenésének valószínűségével, ehhez pedig fontos eszköz a konkordanciasorok vizsgálata, amellyel kiszűrhetjük a véletleneket. A Sketch Engine Concordance nevű korpuseszközének használata számos keresési, szűrési és megjelenítési lehetőséget nyújt a felhasználóknak. Ahogy a korábbi eszközöknél itt is elérhető az egyszerű keresés, amely egyetlen funkcióban a szövegtípus beállításának lehetőségében tér el a korábban említett eszközöktől.

52. ábra A Concordance egyszerű keresés funkciója

CONCORDANCE

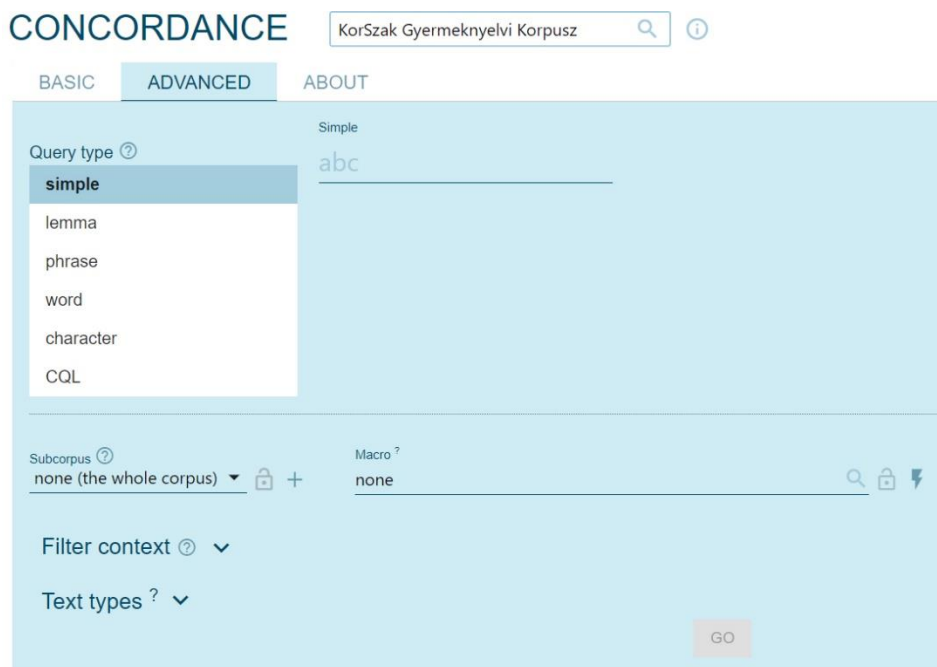
BASIC | ADVANCED | ABOUT

Simple search

Text types

Az egyszerű keresés során is módunkban áll több szóra, szókapcsolatokra keresni, de a találatainkat tovább finomíthatjuk az összetett keresés opciójával.

53. ábra A Concordance összetett keresés funkciója



A lekérdezés típusánál [Query type] hatféle módot állíthatunk be: egyszerű keresés, lemmákra, kifejezésekre, szavakra, karakterekre történő keresés, valamint CQL [Corpus Query Language], amely egy, a korpusznyelvészeten használt kódrendszer. Az egyszerű keresési módnál is használhatunk bizonyos karaktereket, a már sokszor említett RegEx jeleket, amelyek segíthetik a munkánkat. Ezeket a következő ábrán (54. ábra) foglalom össze (Sketch Engine, é.n.i).

54. ábra a Sketch Engine keresőeszközeivel használható jelek [RegEx] (Sketch Engine, é.n.i)

| Jel [RegEx]  | Találat   |
|--------------|---|
| <b>k*</b>    | lemmák, szavak, amelyek <i>k</i> betűvel kezdődnek      |
| <b>*ott</b>  | lemmák, szavak, amelyek <i>-ott</i> végződésűek         |
| <b>*</b>     | nem meghatározott lemma vagy szó egy kifejezésben       |
| <b>* * *</b> | három nem meghatározott lemma vagy szó egy kifejezésben |
| <b>???t</b>  | <i>-t</i> betűvel végződő négy betűs lemma vagy szó     |

| Jel [RegEx]        | Találat   |
|--------------------|---|
| --                 | szóösszetételek keresése attól függetlenül, hogy egyben, külön vagy kötőjellel írják őket       |
| színházba jár megy | megtalálja az <i>színházba jár</i> és az <i>színházba megy</i> összes megjelenését a korpuszban |

A lemmákra való keresésnél két módon járhatunk el. Egyszerűen beírjuk a lemmát, így megkapjuk az összes szóalakot, amely ahhoz a lemmához tartozik, vagy egy pont és egy csillag segítségével rákereshetünk bizonyos lemmákra, amelyek a meghatározott betűvel kezdődnek. A keresésünket a szófaj megadásával tovább szűkíthetjük. Az alábbi, konkordanciasort mutató ábrán a *k.\**, vagyis a *k* betűvel kezdődő lemmák egy része látható (Sketch Engine, é.n.i).

55. ábra A *k* betűvel kezdődő lemmák a KorSzak Gyermeknyelvi Korpuszban

A szófaj megadására ezen kívül még a szavakra történő keresés során van lehetőségünk. A lemmára történő és a szóra, kifejezésre való lekérdezés között egy különbség van. Ahogy fent említettem a lemmára történő lekérdezés során az eredmények között az összes szóalak megtalálható. Abban az esetben, ha kifejezésre vagy szóra keresünk, az eredmények között csak azokat jeleníti meg a rendszer, amelyek pontosan ugyanolyan szóalakban vannak, mint ahogyan beírtuk őket (Sketch Engine, é.n.i).

56. ábra Példa a kapott eredményekre lekérdezési típusonként (Sketch Engine, é.n.i)

| Amit beírtunk | Lekérdezés típusa | Kapott eredmény  |
|---------------|-------------------|--|
| megy          | lemma             | Minden szóalakot mutat: mennek, menni, megy, mentek, megyünk...                                    |
| megy          | szó               | Csak a <i>megy</i> szóalakokat mutatja.  |
| megy          | kifejezés         | Csak a <i>megy</i> szóalakhhoz tartozó kifejezéseket mutatja: megy át az úton, megy át az erdőn... |

A karakterekre történő keresés során olyan tokeneket listáz ki a Concordance eszköz, amelyekben megtalálható a megadott karakterkombináció (Sketch Engine, é.n.i). Ha olyan tokeneket keresünk például, amelyekben két *b* betű található, akkor az alábbi eredményeket kapjuk (57. ábra). A lenti konkordanciasor jól mutatja, hogy a karakterekre történő keresés nem minden esetben hozhatja a kívánt eredmény. Ha valóban csak arra voltunk kíváncsiak, hogy a korpuszunkban hogyan jelennek meg azok a tokenek, amelyekben két *b* betű található, akkor a keresésünk eredményesnek mondható. Abban az esetben azonban, ha azokra a mintázatokra, mondatokra lettünk volna kíváncsiak, amelyekben fokozott melléknevek szerepelnek, már kevésbé nevezhetnénk a találatokat megfelelőnek. Erre a problémára jelenthet megoldást az utolsó lekérdezési típus a CQL, amelyet összetett grammatikai és lexikai mintázatok vagy bonyolultabb keresési feltételek megadására alkalmazhatunk.

57. ábra Találatok a két b betűt tartalmazó tokenekre a KorSzak Gyermekegyelvi Korpuszban

|    | Details                        | Left context   | KWIC                                | Right context   |
|----|--------------------------------|--|-------------------------------------|---|
| 1  | <input type="checkbox"/> doc#0 | atem őket. - És minden állatot, vagy csak a vadon élő, vagy a háziállatokat  | <b>inkább</b><br>Adv                | ? - Mindegyiket, mindegyik ugyanolyan fontos, és mindegyik ugyanolyan ai      |
| 2  | <input type="checkbox"/> doc#0 | tól mást.</s><s>Én az elefántokat pont az érzelmekről ugye beszéltünk az     | <b>előbb</b><br>Adv                 | .</s><s>Az elefántok nagyon összetett állatok, például hogyan ugye mindi      |
| 3  | <input type="checkbox"/> doc#0 | lenvonás az, hogy valaki hogyan bánik az állatokkal? - Hát persze, mert a    | <b>legtöbbben</b><br>Adv            | ügy bánnak az emberekkel is.</s><s>Például, hogyha van egy kutya, akít l      |
| 4  | <input type="checkbox"/> doc#0 | lik a lelkekkel. - Meg nem is feltétlenül jellemvonásnak mondanám, hanem     | <b>inkább</b><br>Adv                | sok mindent megmutat.</s><s>Tehát, hogy aki képes büntudat nélkül, vag        |
| 5  | <input type="checkbox"/> doc#0 | t tudom én, egy hőrcsög nem feltétlenül igényli, hogy foglalkozz vele, vagy  | <b>legalábbis</b><br>Adv            | , hogyha mégis úgy döntött, hogy fényes nappal kivakarod a vackából, ugy      |
| 6  | <input type="checkbox"/> doc#0 | iatják, és te is szereted simogatni.</s><s>Ezért is hívják az emberek egyik  | <b>legjobb</b><br>Adj Nom           | barátjának. - És ha már itt a kutyánál tartunk, te kutyás vagy inkább, vagy r |
| 7  | <input type="checkbox"/> doc#0 | egyik legjobb barátjának. - És ha már itt a kutyánál tartunk, te kutyás vagy | <b>inkább</b><br>Adv                | , vagy macskás, esetleg hőrcsögös? - Én kutyás vagyok, nekünk is van eg       |
| 8  | <input type="checkbox"/> doc#0 | 'mást, csak már szegény kicsit öreg és gyorsan elfárad, úgyhogy már ilyen    | <b>nagyobb</b><br>Adj Nom           | sétákra nem tudunk vele menni. - Én is inkább kutyás vagyok, mondtam is,      |
| 9  | <input type="checkbox"/> doc#0 | lfárad, úgyhogy már ilyen nagyobb sétákra nem tudunk vele menni. - Én is     | <b>inkább</b><br>Adv                | kutyás vagyok, mondtam is, hogy van kutyám.</s><s>Ebben az is közrejárt       |
| 10 | <input type="checkbox"/> doc#0 | enni. - Én is inkább kutyás vagyok, mondtam is, hogy van kutyám.</s><s>      | <b>Ebben</b><br>N_Pro line          | az is közrejártszott, hogy nekem sokáig volt macskám, kisebb-nagyobb me       |
| 11 | <input type="checkbox"/> doc#0 | yám.</s><s>Ebben az is közrejártszott, hogy nekem sokáig volt macskám,       | <b>kisebb-nagyobb</b><br>Adj Nom    | megszakításokkal, és mind a három macskám hasonlóképpen végezte.</s           |
| 12 | <input type="checkbox"/> doc#0 | térve?</s><s>Te mit gondolsz a vadászatról mint, a vadászatról? - Hát így    | <b>hobbiból</b><br>N_Ela            | annyira nem támogatom, de hát valahogy meg kell akadályozni, hogy elsz        |
| 13 | <input type="checkbox"/> doc#0 | Je hát valahogy meg kell akadályozni, hogy elszaporodjanak, mert hogyha      | <b>abbahagyják</b><br>V_Prs_Dat 3Pl | , akkor be fognak jönni a falvakba, meg így borul az egész rendszer. - Meg    |

A CQL kódot, amelyet a Sketch Engine fejlesztői is átvettek, az 1990-es években fejlesztették ki a Stuttgarter Egyetem korpusznyelvészeti műhelyében (IMS, é.n.). Ezt a kódot a konkordancia keresések alatt a tokenekre vonatkozó feltételek beállításánál használhatjuk. A fent említett példára visszatérve abban az esetben, ha a fokozott melléknevek használatát szeretnénk megfigyelni a korpuszunkban és ki szeretnénk zárni a többi tokent, mint például a két *b* betűt tartalmazó hobbiszót, a kilistázott tokenek csak melléknevek lehetnek. Ezt a következő CQL kód megadásával tehetjük meg:



[word= ".\*bb" & tag = "Adj.\*"]

### 58. ábra Találatok a két *b* betűt tartalmazó melléknevekre a KorSzak Gyermeknyelvi Korpuszban

CONCORDANCE KorSzak Gyermeknyelvi Korpusz

CQL [word= ".\*bb" & tag = "Adj.\*"] • 245  
3,424.37 per million tokens • 0.34%

|    | Details                        | Left context   | KWIC                             | Right context   |
|----|--------------------------------|--|----------------------------------|---|
| 1  | <input type="checkbox"/> doc#0 | latják, és te is szeretted simogatni.</s><s>Ezért is hívják az emberek egyik | <b>legjobb</b><br>Adj Nom        | barátjának. - És ha már itt a kutyánál tartunk, te kutyás vagy inkább, vagy r   |
| 2  | <input type="checkbox"/> doc#0 | mást, csak már szegény kicsit öreg és gyorsan elfárad, úgyhogy már ilyen     | <b>nagyobb</b><br>Adj Nom        | sétákra nem tudunk vele menni. - Én is inkább kutyás vagyok, mondtam is,        |
| 3  | <input type="checkbox"/> doc#0 | yám.</s><s>Ebben az is közrejátszott, hogy nekem sokáig volt macskám,        | <b>kisebb-nagyobb</b><br>Adj Nom | megszakításokkal, és mind a három macskám hasonlóképpen végezte.</s>            |
| 4  | <input type="checkbox"/> doc#0 | ul az állatot, hogyha épp nem eszel minden nap húst attól nem lesz sokkal    | <b>kevesebb</b><br>Adj Nom       | , de ennek is... - Pláne, ha nem eszed meg, akkor majd más megeszi. - Ige       |
| 5  | <input type="checkbox"/> doc#0 | bb éljen. -Igen. -Ez effektíve a természet törvénye, hogy amelyik állat nem  | <b>gyorsabb</b><br>Adj Nom       | a ragadozónál, vagy valamit elbaltáz, akkor az el fog pusztulni.</s><s>Me       |
| 6  | <input type="checkbox"/> doc#0 | iskodni úgy, hogy az élőhelyeiket próbáljuk megőrizni, mert szerintem ez a   | <b>legnagyobb</b><br>Adj Nom     | segítség. - Hát igen.</s><s>És hogy nem vágjuk csonkára az erdőcskéket          |
| 7  | <input type="checkbox"/> doc#1 | d is. - A kutya nagyon bűdös, tényleg mert hatalmas bundája van, és minél    | <b>nagyobb</b><br>Adj Nom        | a bundája, annál bűdösebb szerintem, de nyáron fürdetjük rendszeresen.<         |
| 8  | <input type="checkbox"/> doc#1 | is, tényleg mert hatalmas bundája van, és minél nagyobb a bundája, annál     | <b>bűdösebb</b><br>Adj Nom       | szerintem, de nyáron fürdetjük rendszeresen.</s><s>Nyáron jó illata van, c      |
| 9  | <input type="checkbox"/> doc#1 | z, meg a róka. - Róka, azért miért? - Mert nagyon bűdös.</s><s>Ott van a     | <b>legbűdösebb</b><br>Adj Nom    | az állatkerben, ahol róka van, és azért nem, mert az nagyon bűdös. - Hát i      |
| 10 | <input type="checkbox"/> doc#1 | latom lenne, hogy szegénykémet lehet, hogy lehet, hogy lehetne most egy      | <b>jobb</b><br>Adj Nom           | gazdája is, ha nem hoztam el volna mit tudom én, a menhelyről, vagy bárm        |
| 11 | <input type="checkbox"/> doc#1 | am.</s><s>De attól még ők ugyanúgy játszottak vele, de szerintem ez így      | <b>jobb</b><br>Adj Nom           | volt, mert ha ők gondozták volna, akkor lehet, hogy szegényke még előbb i       |
| 12 | <input type="checkbox"/> doc#1 | <s>Nem érek rá.</s><s>Most pont nem jó.</s><s>De majd megcsinálom            | <b>később</b><br>Adj Nom         | , nem csináltam meg. - Hát nálunk muszáj volt, mert olyan bűdös volt a ket      |
| 13 | <input type="checkbox"/> doc#1 | ából.</s><s>Ha valaki például bent lakik a városban, ott lehet, hogy nem a   | <b>leghíresebb</b><br>Adj Nom    | ilyen nagy kutyát tartani.</s><s>Ilyen kis kicsi kutyát.</s><s>Ilyen, mit tudoi |

A fenti kódot alkalmazva a találatok között minden fokozott melléknév, tehát a közép- és felsőfokú alakok is megjelennek (58. ábra). Amennyiben a felsőfokú melléknevekkel nem szeretnénk foglalkozni, a CQL kódunkat tovább kell bővítenünk egy kizáró feltétellel, amelyet egy felkiáltójel segítségével tehetünk meg. A lenti kódban látható !containing kezdetű rész azt jelenti, hogy a konkordanciasor ne tartalmazza azokat a szavakat, amelyek *leg-* előtaggal kezdődnek.

[word= ".\*bb" & tag = "Adj.\*"] !containing [word= "leg.\*"]

Az így kapott eredményekben már csak a középfokú melléknevek szerepelnek (59. ábra). Abban az esetben, ha csak a felsőfokban lévő melléknevek érdekelnek bennünket, a képletből ki kell szednünk a felkiáltójelét. Az így kapott konkordanciasorokban már csak a felsősokú alakok szerepelnek (60. ábra).

## 59. ábra Találatok a középfokú melléknevekre a KorSzak Gyermeke nyelvi Korpuszban

CONCORDANCE

CQL [word="\*" & tag="Adj.\*"] containing [word="leg.\*"] • 188  
2,627.68 per million tokens • 0.26%

|    | Left context  | KWIC                             | Right context  |
|----|---|----------------------------------|--|
| 1  | gymást, csak már szegény kicsit öreg és gyorsan elfárad, úgyhogy már ilyen    | <b>nagyobb</b><br>Adj.Nom        | sétákra nem tudunk vele menni. - Én is inkább kutyás vagyok, mondtam is, h     |
| 2  | utyám.</s></s>Ebben az is közrejátszott, hogy nekem sokáig volt macskám,      | <b>kisebb-nagyobb</b><br>Adj.Nom | megszakításokkal, és mind a három macskám hasonlóképpen végezte.</s>           |
| 3  | lálul az állatot, hogyha épp nem eszel minden nap húst attól nem lesz sokkal  | <b>kevesebb</b><br>Adj.Nom       | , de ennek is... - Pláne, ha nem eszed meg, akkor majd más megeszi. - Igen     |
| 4  | vább éljen. -Igen. -Ez effektíve a természet törvénye, hogy amelyek állat nem | <b>gyorsabb</b><br>Adj.Nom       | a ragadozónál, vagy valamit elbaltáz, akkor az el fog pusztulni.</s></s>Meg é  |
| 5  | ád is. - A kutya nagyon bűdös, tényleg mert hatalmas bundája van, és minél    | <b>nagyobb</b><br>Adj.Nom        | a bundája, annál bűdösebb szerintem, de nyáron fürdetjük rendszeresen.</s>     |
| 6  | dös, tényleg mert hatalmas bundája van, és minél nagyobb a bundája, annál     | <b>bűdösebb</b><br>Adj.Nom       | szerintem, de nyáron fürdetjük rendszeresen.</s></s>Nyáron jó illata van, de   |
| 7  | udatom lenne, hogy szegénykémet lehet, hogy lehet, hogy lehetne most egy      | <b>jobb</b><br>Adj.Nom           | gazdája is, ha nem hoztam el volna mit tudom én, a menhelyről, vagy bármi,     |
| 8  | áltam.</s></s>De attól még ők ugyanúgy játszottak vele, de szerintem ez így   | <b>jobb</b><br>Adj.Nom           | volt, mert ha ők gondozták volna, akkor lehet, hogy szegényke még előbb m      |
| 9  | 's></s>Nem érek rá.</s></s>Most pont nem jó.</s></s>De majd megcsinálom       | <b>később</b><br>Adj.Nom         | , nem csináltam meg. - Hát nálunk muszáj volt, mert olyan bűdös volt a ketre   |
| 10 | gmondják, hogy attól nagyon érzékenyek, és feltűnik nekik, hogyha mondjuk     | <b>melegebb</b><br>Adj.Nom       | az ember bőre, és abból tudnak levonni következtetéseket, és akkor múltkor     |
| 11 | </s>És ez szerintem elég érdekes. - Ez nagyon érdekes. - Meg elvileg sokkal   | <b>jobb</b><br>Adj.Nom           | a hallásuk, mint nekünk. - Az biztos. - Sokkal jobban látnak a sötétben is, mi |
| 12 | íz olyan szemüvegek, mint az ilyen légszem, hogy ilyen kockásított és ilyen   | <b>homályosabb</b><br>Adj.Nom    | , és így, hogy hogy tudnak azzal élni? - Meg az, hogy a légy az, ha például n  |
| 13 | et okoz nekik?</s></s>Vagy én nem értem. - Én se.</s></s>Mondjunk biztos      | <b>puhább</b><br>Adj.Nom         | , mint az ilyen mű izé, de hát. - Hát amennyi pénzzel szerintem megveszik ez   |

## 60. ábra Találatok a felsőfokú melléknevekre a KorSzak Gyermeke nyelvi Korpuszban

CONCORDANCE

CQL [word="\*" & tag="Adj.\*"] containing [word="leg.\*"] • 57  
796.69 per million tokens • 0.03%

|    | Left context   | KWIC                            | Right context   |
|----|--|---------------------------------|---|
| 1  | eg hát ő is szereti, és simogatják, és te is szereted simogatni.</s></s>Ezért is hívják az emberek egyik | <b>legjobb</b><br>Adj.Nom       | barátjának. - És ha már itt a kutyánál tartunk, te kutyás vagy inkább, vagy macskás, esetleg hörcsögös      |
| 2  | ők, vagy tudunk rólok gondoskodni úgy, hogy az élőhelyeiket próbáljuk megőrizni, mert szerintem ez a     | <b>legnagyobb</b><br>Adj.Nom    | segítség. - Hát igen.</s></s>És hogy nem vájuk csak a erdőcskéket. - Igen, mert rengeteg hely               |
| 3  | gy nem. - A pók, meg a borz, meg a róka. - Róka, azért miért? - Mert nagyon bűdös.</s></s>Ott van a      | <b>legbűdösebb</b><br>Adj.Nom   | az állatkerben, ahol róka van, és azért nem, mert az nagyon bűdös. - Hát nem tudom, az ilyen kigyó r        |
| 4  | </s></s>De nem tudom igazából.</s></s>Ha valaki például bent lakik a városban, ott lehet, hogy nem a     | <b>legelőnyösebb</b><br>Adj.Nom | ilyen nagy kutyát tartani.</s></s>Ilyen kis pici kutyát.</s></s>Ilyen, mit tudom én pomerániai torpe spio   |
| 5  | iskutya. - Amúgy eskü.</s></s>De a gonosz végül megkapta méltóbb büntetését. - Na ja, a karma. - A       | <b>legjobb</b><br>Adj.Nom       | rész, amikor beesett a kakiba. - Következő. - Mit gondolsz a vadászatról? - Szerintem a vadászatról D       |
| 6  | mény ott megnézni az állatokat kint, de nekik, hogy mindig be vannak zárva.</s></s>Az azért nem a        | <b>legjobb</b><br>Adj.Nom       | ézés. - Jó, de mondjuk igazából részben kaptak védelmet is, meg nem is kell nagyon vadászni úgy m           |
| 7  | nem képesek ilyen mértékű direkt gonoszságra.</s></s>Tehát, hogy amikor az oroszán levadássza a          | <b>leggyengébb</b><br>Adj.Nom   | gnút, vagy mit tudom én, zebrát akkor, nem arra törekszik, hogy az állat ott minél többet szenvedje         |
| 8  | szem-szájnak ingere, és Bubuka egy ponttal verte meg a németjuhász keveréket is. - Szóval ő volt a       | <b>legjobb</b><br>Adj.Nom       | . - Igen, ő volt a legjobb. - a magyar. - Ez egy igazi elismerő. - Igen. - Váól - Amilyen luke szegényke, o |
| 9  | és Bubuka egy ponttal verte meg a németjuhász keveréket is. - Szóval ő volt a legjobb. - Igen, ő volt a  | <b>legjobb</b><br>Adj.Nom       | .a magyar. - Ez egy igazi elismerő. - Igen. - Váól - Amilyen luke szegényke, olyan jól teljesített. - Na,   |
| 10 | e lájja, hogy valami nem oké, és akkor elindul befelé. - És érted bemegy. -Szerintem ezért is a kutya a  | <b>legjobb</b><br>Adj.Nom       | , vagy legideálisabb háziállat, mert én egy macskáról egyáltalán nem tudom elképzelni, hogy ő akkor r       |
| 11 | alami nem oké, és akkor elindul befelé. - És érted bemegy. -Szerintem ezért is a kutya a legjobb, vagy   | <b>legideálisabb</b><br>Adj.Nom | háziállat, mert én egy macskáról egyáltalán nem tudom elképzelni, hogy ő akkor most superhős, meg           |
| 12 | és így megszívja magát, mint a ruhaanyag. - Igen.</s></s>Ilyenek a hosszabb szőrű kutyák. - Én, az a     | <b>legjobb</b><br>Adj.Nom       | rész, az az évben a legjobb rész, amikor a Bubu szőre rövid, mert akkor nem kell kiszedni belőle a bo       |
| 13 | it a ruhaanyag. - Igen.</s></s>Ilyenek a hosszabb szőrű kutyák. - Én, az a legjobb rész, az az évben a   | <b>legjobb</b><br>Adj.Nom       | rész, amikor a Bubu szőre rövid, mert akkor nem kell kiszedni belőle a bogáncsot, nem lesz dzsuvas.         |

A Concordance eszközzel való keresés során kapott konkordanciasorok alapbeállításuként a korpuszban található dokumentumok sorrendjében jelennek meg. A sorrenden változtathatunk, ha rákattintunk a kérdőjeles ikonnal jelölt véletlenszerű sorrendre [Get a random sample] vagy a két nyílal jelölt sorok keverése ikonra [Shuffle lines].



A konkordanciasorokat a mellettük balra vagy jobbra álló lemmák szerint alfabetikus sorrendbe is rendezhetjük, amely megkönnyíti a mintázatok keresését. A Concordance eszköz további funkcióinak segítségével megismerhetjük a konkordanciasorok elemeinek gyakorisági mutatóját, a kollokációit, valamint a találatok korpuszon belüli eloszlását is. Az ikonok az iménti felsorolás sorrendjében a következők:



A konkordanciasorok megjelenésén is változtathatunk a különböző nézetek beállításával. Ez egyrészt jelentheti a két megjelenítési formát: KWIC [keyword in context], vagyis kontextusban lévő kulcsszavak, amelynél a keresett elemek egymás alatt szerepelnek, mellettük mindkét irányban néhány szó látható, vagy pedig mondatok [sentence], amely a keresett elemeket teljes mondatokban jeleníti meg. A második esetben a keresett elemek nem egymás alatt szerepelnek, mint ahogy a következő ábrán is látható (61. ábra).

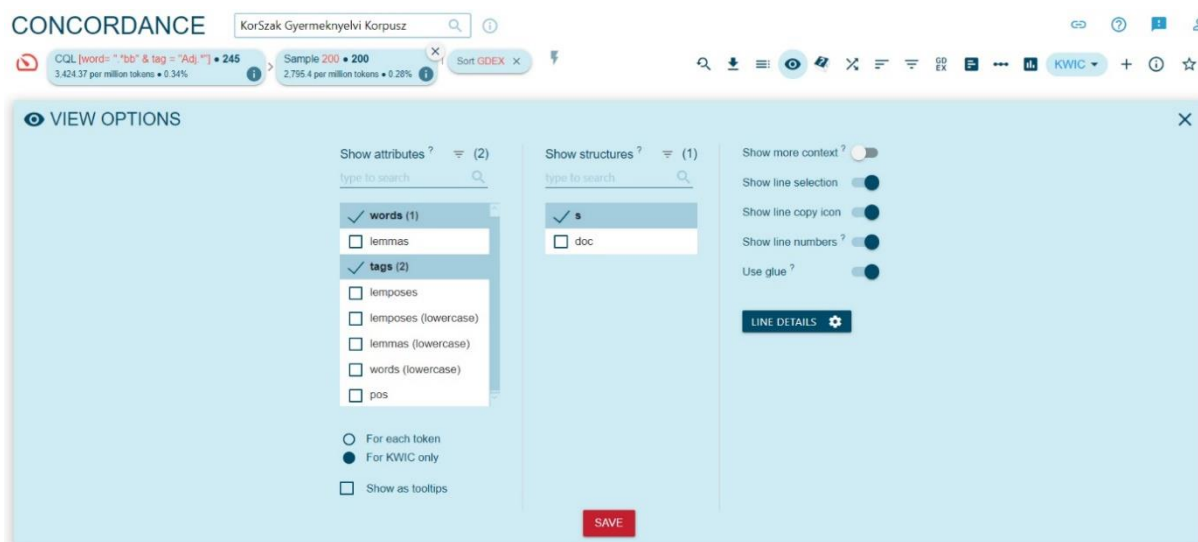
**61. ábra** Találatok a középfokú melléknevekre a KorSzak Gyermekegyelvi Korpuszban mondatonkénti elrendezésben

| Document ID | Search Results  | Sentence   |
|-------------|---|--|
| doc#65      | <s>En ezt nagyon szeretem, meg hát ugye a Mi kis falunk is azért, úgymond így <b>híresebb</b> lett.</s>                   | <s>En ezt nagyon szeretem, meg hát ugye a Mi kis falunk is azért, úgymond így híresebb lett.</s>                   |
| doc#8       | <s>Az egyik nagyon kicsi volt, a másik meg már kicsit <b>nagyobb</b> , de ő is nagyon aranyos volt. - Jaj!</s>            | <s>Az egyik nagyon kicsi volt, a másik meg már kicsit nagyobb, de ő is nagyon aranyos volt. - Jaj!</s>             |
| doc#59      | <s>Hát úgy <b>könnyebb</b> volt.</s>  | <s>Hát úgy könnyebb volt.</s>  |
| doc#52      | <s>Talán <b>jobb</b> .</s>  | <s>Talán jobb.</s>   |
| doc#68      | <s>En attól sokkal <b>fáradtabb</b> leszek, tehát egy ilyen edzés után így hiába elfáradt egy ilyen felfrissülést ad.</s> | <s>En attól sokkal fáradtabb leszek, tehát egy ilyen edzés után így hiába elfáradt egy ilyen felfrissülést ad.</s> |
| doc#2       | <s>Nekik nincs más választásuk csivava, <b>kisebb</b> állat.</s>  | <s>Nekik nincs más választásuk csivava, kisebb állat.</s>  |
| doc#52      | <s>De nem tudom, filmeknél az összesnél <b>jobb</b> .</s>   | <s>De nem tudom, filmeknél az összesnél jobb.</s>  |
| doc#2       | <s>Csak az embereknek sokkal <b>rosszabb</b> tulajdonságaik vannak, mint az állatoknak.</s>                               | <s>Csak az embereknek sokkal rosszabb tulajdonságaik vannak, mint az állatoknak.</s>                               |
| doc#27      | <s>Az a fontos, hogy a <b>nehezebb</b> időkben is számíthass rá.</s>  | <s>Az a fontos, hogy a nehezebb időkben is számíthass rá.</s>  |
| doc#28      | <s>Szerintem ez így <b>személyesebb</b> , és közben én is jól érzem magamat.</s>  | <s>Szerintem ez így személyesebb, és közben én is jól érzem magamat.</s>   |
| doc#29      | <s>Példának azt hozta fel, hogy az egyikőjük csinált valami <b>nagyobb</b> butaságot.</s>                                 | <s>Példának azt hozta fel, hogy az egyikőjük csinált valami nagyobb butaságot.</s>                                 |
| doc#52      | <s>En nem tudom, szívesebben rajzolok helyette, meg ilyen <b>kratívabb</b> dolgokat csinállok.</s>                        | <s>En nem tudom, szívesebben rajzolok helyette, meg ilyen kratívabb dolgokat csinállok.</s>                        |

Ezeket a beállításokat kívül a szem ikonra kattintva azt is megadhatjuk, hogy milyen információ jelenjen meg még a konkordanciasorokon. Azoknál a konkordanciasoroknál,

amelyeket eddig példaként beillesztettem beállításaként a címkék megjelenítése szerepelt, ezért a pirossal jelölt keresett elemek alatt szerepeltek olyan rövidítések, mint például az Adj. Nom., amely a melléknév nominatívusz alakját jelöli. A lepos beállítása esetén minden találat alatt megjelenik a szó lemmatizált változata és a szófajának betűjele. Ilyen lehet a *megy-v*, amelyből a MEGY lemma, a *v* betű pedig az igét jelöli. Az attribútumok közül kiválaszthatjuk még a szavak, a lemmák és a szófajok jelölését is, valamint ahogy a lenti ábrán (62. ábra) is látható ugyanezek megtalálhatók egy zárójelben lévő lowercase, vagyis kisbetűk szóval kiegészítve. Ennek a beállításnak az az előnye, hogy a szavak, lemmák kis és nagybetűs változatai között nem tesz különbséget. Tehát ha a keresésünkben nagy kezdőbetűvel szerepel az adott elem, a keresés során a kisbetűs változatait is kilistázza az eszköz.

62. ábra A Concordance eszköz nézetének beállítási lehetőségei



A fenti ábrán (62. ábra) a nézet beállításokban még az egyes sorok megjelenésének szerkesztésére is van lehetőségünk. Megjeleníthetjük vagy eltüntethetjük a sorszámot, a másolás ikont, a kiválasztásra szolgáló négyzetet is. A keresett elem kerülhet az alapbeállításnál hosszabb kontextusba, illetve bizonyos szavak vagy a központozás között megjeleníthetjük vagy eltörölhetjük a szóközt [Use glue]. A további beállításokra kattintva [Line details] megadhatjuk, hogy pontosan hány elem jelenjen meg a KWIC mellett balra és jobbra, illetve azt is meghatározhatjuk, hogy metaadatok jelenjenek-e meg az egyes sorokban (63. ábra).

### 63. ábra A Concordance eszköz nézetének beállítási lehetőségei, metaadatok megjelenítése

Display and count metadata

Select the metadata to be displayed in the concordance. Click to calculate statistics.

Display above lines ?  Shorten to 30 characters

en? - Persze, hát ők is élőlények is éreznek, és hogyha belegondolsz, a kutyák is szeretnek téged, te is szeretheted a kutyákat, a cicád is szerethet tége

type to search (3)

|   |         |  |
|---|---------|--|
| <input type="checkbox"/> Token number               | 73      |  |
| <input checked="" type="checkbox"/> Document number | 0       |  |
| <input checked="" type="checkbox"/> doc.Age         | 13      |  |
| <input checked="" type="checkbox"/> doc.County      | Baranya |  |

Visszatérve a konkordanciasorok összetett kereséséhez, meg kell említenünk még két lényeges funkciót. Egyrészt a lehetőséget, hogy alkorpuszokban, szűrőket alkalmazva (lemmák, szófajok kiszűrése), illetve bizonyos szövegtípusokban található eredményeket vizsgálhassunk meg. Másrészt pedig, hogy ahogy a legtöbb modern szoftvernél már megszokhattuk, a Sketch Engine fejlesztői is törekedtek arra, hogy a felhasználók minél kisebb idő és energia befektetésével ériék el a lehető legjobb eredményeket. Erre utal, hogy a Concordance eszköz használata közben módunkban áll a keresési folyamatot makrórögzítő segítségével automatizálni. A makrórögzítő (64. ábra) megjegyzi a beállításainkat, így egyetlen kattintással végezhetjük el ugyanazokat a vizsgálatokat egymás után, nem szükséges minden alkalommal előlről kezdeni a folyamatot.

A kapott konkordanciasorokat több fájlkiterjesztésben is lementhetjük: TXT, CSV, XLSX, XML, valamint PDF formátumban. A letöltéseknek adatvédelmi okokból szabtak meg néhány határt. A keresések során tízmillió sor jeleníthető meg, amelyből saját korpuszunk esetén mind, készen elérhető korpuszból csak tízezer sor tölthető le. A keresett elemek mellett saját korpusz esetén maximum 500, készen hozzáférhető korpusznál maximum 100 szomszédos elem menthető le (Sketch Engine, é.n.j).

64. ábra A konkordanciasorok keresési beállításainak makrórögzítése

# CONCORDANCE

KorSzak Gyermeknyelvi Korpusz

BASIC **ADVANCED** ABOUT

Query type ⓘ

- simple**
- lemma
- phrase
- word
- character
- CQL

Simple  
abc

---

Subcorpus ⓘ

none (the whole corpus) 🔒 +

Macro ?  
none 🔍 🔒 ⚡

Filter context ⓘ ▾

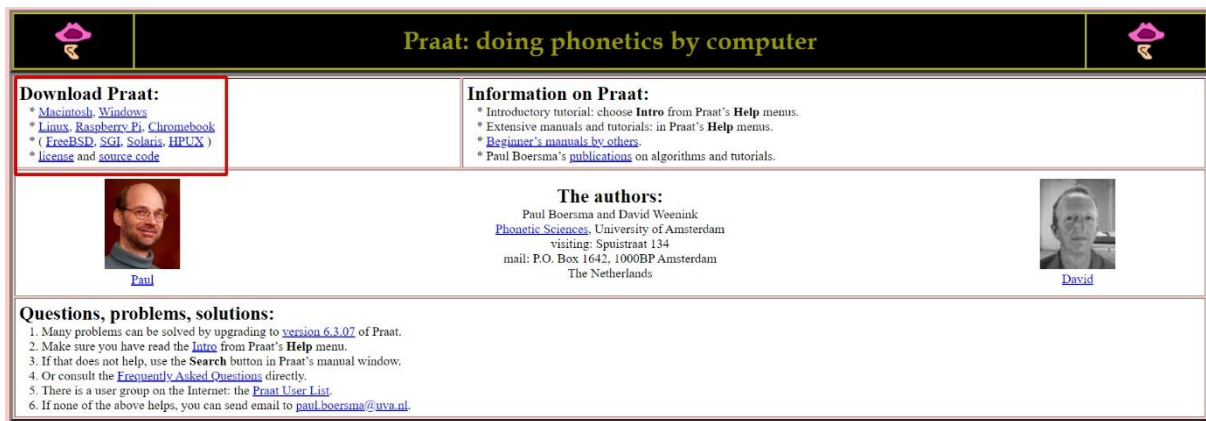
Text types ? ▾

GO

## 2.2 Praat

Ahogy korábban már említésre került (ld. II.5 fejezet) a legújabb kutatások bebizonyították, hogy a szünetek hossza vagy sebessége helyett a szünetek helye a legjelentősebb tényező az érzékelt beszédfolyékonyság [perceived fluency], tehát a beszélő fluensségének hallgatók általi értékelésében. Kimutatták, hogy a kompetens nyelvhasználók az esetek elnyomóan nagy részében nem tartanak szünetet egy-egy formulaszerű egységen belül. Döntően csak a formulaszerű egységek előtt vagy után kerül erre sor (vö. Kahng, 2017). Ebből kifolyólag a beszédszünetek helyének megfigyelése is megfelelő szegmens lehet a formulaszerű egységek meghatározásának folyamatában. Ezért döntöttem úgy, hogy a KorSzak Gyermekegyetemi Korpuszból kiválasztott formulaszerű egységek beazonosításához nemcsak a Sketch Engine szoftvert, hanem azokban az esetekben, amikor beágyazott többelemű lexikai egységekről van szó, az egység határainak megállapításához egy másik programot is alkalmazok.

65. ábra A Praat programhoz tartozó weboldal kezdőlapja



**Praat: doing phonetics by computer**

**Download Praat:**

- \* [Macintosh, Windows](#)
- \* [Linux, Raspberry Pi, Chromebook](#)
- \* ([FreeBSD, SGI, Solaris, HP-UX](#))
- \* [license and source code](#)

**Information on Praat:**

- \* Introductory tutorial: choose **Intro** from Praat's **Help** menus.
- \* Extensive manuals and tutorials: in Praat's **Help** menus.
- \* [Beginner's manuals by others](#).
- \* Paul Boersma's [publications](#) on algorithms and tutorials.

**The authors:**

Paul Boersma and David Weenink  
[Phonetic Sciences](#), University of Amsterdam  
visiting: Spuistraat 134  
mail: P.O. Box 1642, 1000BP Amsterdam  
The Netherlands

**Questions, problems, solutions:**

1. Many problems can be solved by upgrading to [version 6.3.07](#) of Praat.
2. Make sure you have read the [Intro](#) from Praat's **Help** menu.
3. If that does not help, use the **Search** button in Praat's manual window.
4. Or consult the [Frequently Asked Questions](#) directly.
5. There is a user group on the Internet: the [Praat User List](#).
6. If none of the above helps, you can send email to [paul.boersma@uva.nl](mailto:paul.boersma@uva.nl).

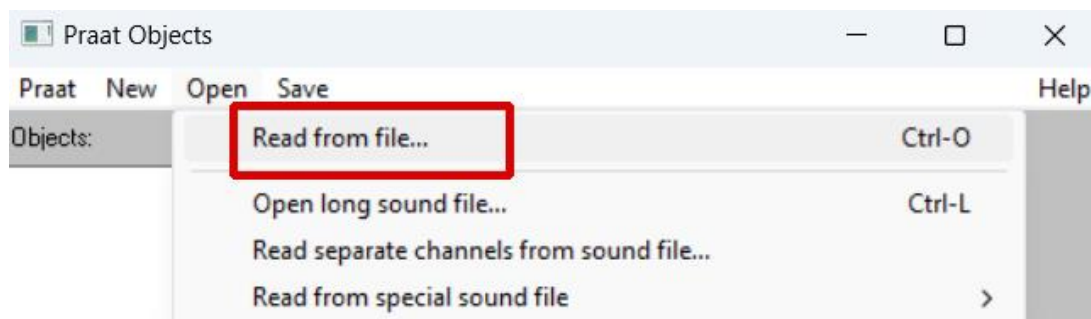
A Praat egy nyílt forráskódú szoftver, amelyet beszéd, zene és egyéb hangelemzésre terveztek. Alkotói Paul Boersma és David Weenink a programot az Amszterdami Egyetem Fonetikai Tudományok Intézetében fejlesztették ki 1993-ban. A Praat lehetővé teszi a felhasználók számára a beszédhangok tanulmányozását, elemzését és manipulálását, valamint a beszéd szintetizálását és hangok generálását is. A szoftver számos eszközt kínál fel a kutatók számára. Ilyenek többek között a spektrális megjelenítés, a paraméterbecslés, a statisztikai elemzés is. Segítségével elemezhető a hangmagasság, az intenzitás, amelyet más akusztikai paraméterek mérésével is kiegészíthetünk. Emellett az artikulációs gesztusok és a beszédsebesség is vizsgálható vele. A felvett és generált hangok szerkesztésére egy grafikus

felületen keresztül is van módunk. Továbbá olyan lehetőségeket is tartalmaz, mint a folyamatok automatizálására vagy testreszabható funkciók létrehozására szolgáló szkriptelés (Boersma – Weeink, 2020). A Praat program alkalmazásával további betekintést nyerjünk a kompetens beszélők természetes nyelvhasználatba.

A program számos felhasználási lehetősége közül a jelen kutatás számára a beszéd közbeni szünetek meghatározásának funkciója volt releváns. A Praat használatával megfigyelhetjük a szüneteket és megszakításokat a beszédfolyamban, ezáltal pedig lehetőségünk nyílik rá, hogy azonosíthassuk azokat a szavakat, amelyek nagyobb valószínűséggel követik egymást (Boersma – Weeink, 2020). Hasonló vizsgálatot végeztek már például lengyel anyanyelvű beszélőkkel (vö. Guz, 2014). Kutatásában Guz (2014) a Praat ezen funkciójának segítségével identifikálta a kísérletben lengyel beszélők által használt formulaszerű egységeket. Az általam bemutatott kutatás azonban nem a Praat programra alapul. Használatával a céloom a Sketch Engine segítségével talált többemű lexikai egységek további vizsgálata volt, hogy egy másik aspektust, a szünetek helyét is megfigyeljem a korpuszban, amely megerősítheti vagy ki is zárhat bizonyos vizsgált egységeket. A kutatás céljainak megvalósításához csak a program alapfunkcióinak használata volt szükséges, amelyek alkalmazását a következőkben mutatom be.

A beszédelemzés a Praat segítségével több lépést foglal magában. Először is, a program – ellentétben az előzőekben bemutatott Sketch Engine szoftverrel – nem használható online. Alkalmazásához le kell töltenünk a számítógépünkön futó operációs rendszerrel kompatibilis változatát. Ha ez megtörtént, a hanganyagot külső forrásból kell a szoftverbe importálni, amelyhez támogatott hangfájl formátumot szükséges használnunk.

**66. ábra** Hangfájl importálása külső forrásból a Praat szoftverben

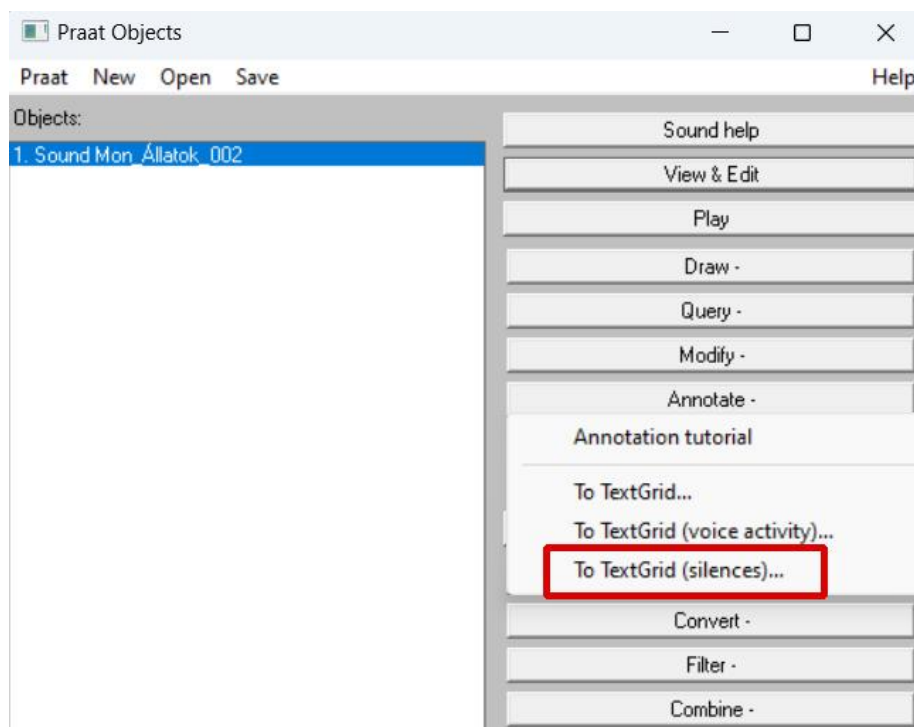




A támogatott formátumok közé tartozik a WAV, MP3, AIFF, AIFC, NeXT/Sun, NIST és a FLAC. A szünetek megfigyeléséhez a legmegfelelőbb fájl típus a WAV, mert az MP3 tömörített fájlként apró csúszásokat okozhat az elemzés során. Ezért a korpuszban lévő videókat és hanganyagokat az elemzés megkezdése előtt ebbe a formátumba konvertáltam.

A Praatban a beszéd szünetek elemzése egyszerű folyamat, amelyhez használhatjuk a programban található eszközt, ahogy a jelen kutatás során is történt, vagy letölthetünk, illetve készíthetünk egy saját szkriptet is erre a célra. A hangfájl betöltése után az annotáció részénél a To TextGrid (silences)... parancsra kattintva (67. ábra) a program létrehoz egy úgynevezett TextGrid típusú objektumot, amelyben automatikusan jelölésre kerülnek a beszédben lévő hangzó [sounding] és csendes [silent] részek. Az így kapott objektumban további elemzésre van lehetőségünk, kivághatunk, kijelölhetünk benne részeket, további egységekre bonthatjuk az automatikusan bejelölt hangzó és csendes sávokat. Ez a beépített funkció alapbeállítása szerint úgy határozza meg a szüneteket, hogy a beszéd közben lévő 0,25 másodperces vagy annál hosszabb csendes szakaszok vagy nonverbális kitöltő elemek határait jelöli be (vö. de Jong, 2012). Ezt a beállítást meg lehet változtatni, így a keresési kritériumokat igazítani tudjuk a hanganyagunkhoz, valamint a kutatásunk céljaihoz.

**67. ábra** A beszéd szünetek automatikus jelölésének módja a Praat szoftverben



Az elemzés a fájl nagyságától függően maximum pár percig tart. Ennek ellenére a formulaszerű egységek gyorsabb vizsgálata érdekében a Sketch Engine programmal talált egységeket a korpuszban megkerestem, majd csak a szükséges részt vágtam ki a hangfájlból, és csak ezt elemeztem a Praat programmal. Az elemzés után az adatok a Praatból különböző formátumokban exportálhatók további értelmezés céljából. A kutatás szempontjából számomra a legmegfelelőbb a vizualizáció elkészítése utána, annak PNG vagy EPS formátumú mentése volt.

### 3. A formulaszerű egységek beazonosítása

Ahogy az elméleti keret szakirodalmi áttekintésében (II. fejezet) említésre került a formulaszerű egységeket „a nyelvhasználó úgy érzi, hogy egyetlen lexikai egységként azonosítja vagy hasznosítja” (Wray, 2019: 267). A Wray (2002) által összegyűjtött több mint ötven aspektusból, valamint a korábbiakban olvasható sokféle tulajdonságból (átláthatóság, gyakoriság, gyors feldolgozás, jelentéstartalom, kiszámíthatóság, konvencionizáltság, megjósolhatóság, pragmatikai funkció, rögzítettség, sematikusság) arra következtethetünk, hogy a formulaszerű egységek beazonosítása sem egyféleképpen történik. A különböző megközelítések egy-egy tulajdonságra koncentrálnak kategorizálják a szekvenciákat, miközben más jellemzőknek nem tulajdonítanak nagy szerepet. A pragmatikai megközelítés inkább a funkcióra helyezi a hangsúlyt, amíg a pszicholingvisztikai a gyakoriságra, a kiszámíthatóságra, illetve a megjósolhatóságra (Siyanova-Chanturia – Pellicer-Sánchez, 2019: 5). Az identifikálásuk Wray szerint vagy empirikus módszerrel például kompetens nyelvhasználók által kitöltött kérdőív segítségével történhet vagy pedig saját magunk által gyűjtött nyelvi anyagokban „vadászhatunk” ezekre az elemekre, szókapcsolatokra. Mindkét módszer alkalmazása során azonban inkább a megérzéseinkre támaszkodunk (Wray, 2002: 19–20), ezért nem lehetünk biztosak a találatok validitásában.

A többszavas formulaszerű szekvenciák beazonosítására korpusznyelvészeti eszközöket is használhatunk. Ugyanakkor meg kell jegyezni, hogy ezekkel az eszközökkel sem egyszerű azonosítani őket. Ha rögzült lexikai egységekről van szó, megtalálásukra az úgynevezett concgram, vagyis variálható végű Ngram keresések alkalmazhatóak. Ha csak részben rögzült lexikai egységekről van szó, egy másik módszert használhatunk: konkordanciasorokból manuálisan találhatjuk meg őket (Hunston, 2022: 102). Ahhoz, hogy meg tudjuk állapítani a gyakorisági alapon kiválasztott szókapcsolatokról, hogy valóban formulaszerű szekvenciák, meg kell határoznunk bizonyos kritériumokat. Biber (2006) például egyetemi oktatási anyagokban és osztálytermi nyelvben keresett formulaszerű lexikai elemeket, amelyeket ő lexikai kötegekként [lexical bundles] említ. Meghatározott feltételrendszere szerint az általa talált lexikai egységek három vagy négy szavasak lehetnek, a kutatásához azonban csak a négy szóból állókat használta fel. Ezen kívül még azt is megszabta, hogy a kiválasztott lexikai egységek legalább negyvenszer jelenjenek meg egy egymillió szóból álló korpuszban (Biber, 2006: 133–134).

A formulaszerű szekvenciáknak egy harmadik beazonosítási módja lehet a spontán

beszédben lévő szünetek, illetve a szünet nélküli egységek identifikálása. Ahogy korábban a beszéd folyékonyaságról szóló résznél (II.5 fejezet) láthattuk, a fluensség egyik meghatározó jellemzője a szünetek, akadozások nélküli beszéd képessége. Meg kell jegyezni azonban, hogy a kompetens nyelvhasználók is tartanak szüneteket, hezitálnak, meg-megállnak a beszédük közben. Ezek az időnkénti hezitációk nem rontják el a beszéd minőségét vagy érthetőségét. Ezzel ellenkezőleg, valójában segítik a megértés folyamatát. A szünetek és az intonáció időnkénti emelkedése és esése támogatja a beszédet hallgató személyt a szintaktikai keret megértésében, valamint a szövegértelmezésben is jelentős szerepet játszanak. A szünetek tekintetében tehát nem az a kérdés, hogy van-e a beszéd közben szünet, hanem inkább az, hogy ez a szünet hol található. Pawley és Syder (2000) vizsgálatából rendelkezésünkre álló eredményekből megtudhatjuk, hogy amikor a beszélők szünetet tartanak, nagyobb valószínűséggel teszik ezt a tagmondatok, többszavas lexikai szekvenciák között, nem pedig azok közepén. Azért nem tűnik fel a kompetens nyelvhasználók számára, amikor a beszélgetőtársuk a diskurzus közben szünetet tart, mert a hallgatók teljes egységek folyékony sorozatát hallják. Ezzel szemben a nyelvtanulók beszédét gyakran a szünetek nem megszokott helye miatt érezzük diszfluensnek. A beszédsebesség lassulása a szekvencia határához közeledve, a szünet a következő szekvencia megkezdése előtt, és a beszédsebesség enyhe gyorsulása a következő szekvencia kimondásakor szintén bizonyítékai egy olyan beszédtervezési modellnek, amelyben részben vagy teljesen előre felépített elemek egyben kerülnek kódolásra. Ezek a szünetek nagyon rövid ideig tartanak. Abban az esetben, amikor a beszélőnek több időre van szüksége, a hezitációra jellemző hangok (őzés) mellett, olyan időkitöltőként funkcionáló formulaszerű egységeket használ, amelyeket szintén egyben szünet nélkül mond ki (Pawley – Syder, 2000: 173). Ilyenek lehetnek többek között magyarul a *hogy is mondjam* vagy a *mit tudom én* kifejezések.

### **3.1 Formulaszerű egységek vizsgálata az agglutináló nyelvekben**

Kijelenthetjük, hogy az alkalmazott nyelvészet és azon belül a korpusz nyelvészet nyelve is az angol. Nemcsak a nagyszámú angol nyelvű szakirodalommal indokolható ez, hanem azzal is, hogy ezt a nyelvet vizsgálják a legtöbbször a korpuszalapú kutatások, beleértve a formulaszerű nyelvhasználat vizsgálatát is. Nyelvünk agglutináló jellege, gazdag morfológiai felépítése miatt azonban máshogy működik, mint az angol nyelv. A legtöbbször említett példa erre, hogy számos esetben, amit az angolul beszélők több szóval fejeznek ki, mi akár egyetlen egy szóval is megtehetjük. Hogyan egyeztetethetők össze és alkalmazhatók akkor az angol nyelvre irányuló

kutatásokhoz kifejlesztett módszerek és paradigmák egy agglutináló nyelv vizsgálatánál? Hogyan lehet kimutatni például a lexikai előfeszítés (Hoey, 2005) jelenségét ebben az esetben? A válasz az, hogy ugyanúgy, ahogyan beszédünk közben eltérően fejezzük ki magunkat, a kutatás során is máshogyan kell gondolkodnunk (Vermeki, 2023b, megjelenés alatt).

A már említett Biber (2006) által kutatott lexikai kötegek [lexical bundles] vizsgálatára a magyar nyelvben Dóla (2014: 20) két megoldást javasol. Az egyik, hogy amíg Biber a négy szóból álló lexikai kötegeket vizsgálta és a lexikai köteget úgy határozta meg, hogy három vagy több szóból álló szekvencia, addig a magyar nyelv vizsgálatánál ezt a számot le kellene csökkenteni. Kettő vagy több szövegszóban – vagyis olyan betűcsoportban, amelynek mindkét oldalán szóköz áll – kellene meghatározni. A másik lehetőség, hogy a lexikai egység hossza maradhatna három vagy több elem, ugyanakkor a nyelv agglutináló jellegének megfelelően a keresés alapegységét, amely a szövegszó, kiegészítenénk az igekötőkkel és a toldalékokkal. Ha megnézzünk más agglutináló nyelvekkel kapcsolatos kutatásokat, akkor láthatjuk, hogy ehhez hasonlóan gondolkodnak az azokat vizsgáló nyelvészek is.

A finn és a török nyelv kutatása kapcsán morfológiai formulaszerűségről [morphological formulacity] írnak a tanulmányokban. Ezt két gondolatra alapozzák. Egyrészt, hogy a formulaszerű egységek a formulaszerű nyelvhasználat definíciói (Wray, 2002: 9) szerint minden szinten megjelennek, másrészt pedig a morfológiailag komplex szavak feldolgozásának a memóriával való kapcsolatára irányuló pszicholingvisztikai vizsgálatokra (vö. Baayen – Schreuder, 2003) (Durrant, 2013: 3).

### **3.1.1 A finn nyelvvel kapcsolatos kutatások**

A finn nyelvben lévő formulaszerű nyelvhasználatot jelenlegi ismereteim szerint csak pszicholingvisztikai keretben vizsgálták, ilyen jellegű korpusznyelvészeti vizsgálatokra még nem került sor. A kutatások elméleti keretükben nem alapoztak a használatalapú nyelvelméletekre. A vizsgálatok során, amelyek főként a főnevek elemzésére irányult, egy úgynevezett kétutas, vagy duális modellt (Pinker, 1991) alkalmaztak, amely segítségével véleményük szerint megállapították, hogy mely szavak tárolódnak holisztikusan és melyek morfémakként a mentális lexikonban. A modell egyik része ahhoz az elmélethez igazodik, amely szerint a morfológiailag összetett szavak felismerése a szavak felbontásán keresztül történik (vö. Taft – Forster, 1975), a másik része (vö. Butterworth, 1983) pedig arra az elméletre, amely szerint minden ismert morfológiailag összetett szó egy egész egységként tárolódik a mentális lexikonban (Lehtonen et al., 2007).

Az említett duális modellt alkalmazó kutatás Lehtonen és Laine (2003) vizsgálata is, amelyben azt próbálták megtudni, hogy a szógyakoriság hogyan befolyásolja az egynyelvű finnek és a finn-svéd kétnyelvűek morfológiai feldolgozását. A lexikai döntés feladatokat [lexical decision task] húsz egyetemistával végezték, akik számára összeállítottak hat darab húsz elemből álló szólistát. A listák közül három csak egy morfémából álló elemeket tartalmazott, a másik három listán pedig több morfémából álló, toldalékolt főnevek szerepeltek. Az elemek gyakorisági mutatójuk szerint is eltértek egymástól. A hallgatóknak ezek különböző variációit mutatták meg. Tehát volt, amikor egy magas gyakorisági mutatójú egymorfémás elem mellé egy alacsony frekvenciaszámú toldalékolt főnév került, vagy fordítva. Az eredmények alapján kijelentették, hogy az egynyelvű finnek az alacsonyabb gyakorisági mutatójú főneveket morfémákként, amíg a magasabb gyakorisági mutatójú főneveket egyben ismerték fel, vagyis azok valószínűleg holisztikusan tárolódnak a memóriájukban. Ezzel szemben a kétnyelvűek minden gyakorisági mutató esetében késleltetett feldolgozást mutattak, ami a szerzők szerint arra utal, hogy a kétnyelvűségük miatt még nem találkoztak elég gyakran egyik főnévvel sem ahhoz, hogy egyben tudják elérni őket a memóriájukban.

Későbbi kutatásokban (vö. Soveri et al., 2007; Vartiainen et al., 2009) szintén arra az eredményre jutottak, hogy a reakcióidő kisebb egymorfémás elemek esetében. Ezt egy afáziás beteg vizsgálatával bizonyították, akinek megfigyelték az olvasási problémáit, a többmorfémás elemek előállításának nehézségeit, valamint azt, hogy a bal felső temporális agykérge erősebben és hosszabb ideig aktiválódott a többmorfémás szavak feldolgozása és produkciója során. Ugyanakkor azt is bebizonyították, hogy a magas gyakorisági mutatójú toldalékolt főnevekre ez nem volt igaz, amely véleményük szerint arra utal, hogy az ilyen nyelvi egységek holisztikusan tárolódnak.

A magyar nyelv vonatkozásában a duális modellt Pléh és Lukács (2001) vizsgálta meg, akik Pléh, Palotás és Lőrík 1994-es gyermeknyelvi kutatását, valamint előfeszítés kísérletek eredményeit is összevetették a kétutas rendszerrel. A gyermeknyelvi kutatás ragozási feladatában három szabályos, produktív főnevet és három szabálytalan, nem produktív főnevet használtak. Arra a következtetésre jutottak, hogy a

„fejlődési kép azt sugallja, hogy nincsen egyetlen egységes vágópont, amely elválasztaná egymástól a szabályos és a szabálytalan alakokat, amint az Pinker (1991) kiinduló elképzelése alapján várható lenne. A magyar alaktan elsajátítása lépcsőzetesebb viszonyt sugall, és azt mutatja, hogy tétel- és típusgyakorisági hatások kölcsönhatása fontos befolyásoló tényező lehet” (Pléh – Lukács, 2001).

Erre abból következtettek, hogy a *madár* szó, amely a nem produktív típushoz tartozik, nem maradt el sokkal a produktív típust képviselő többi főnévtől. Ezt a *madár* – *madarat* típusú paradigma gyakoriságával magyarázták. Az előfeszítés-vizsgálatokból is hasonló következtetésre jutottak. Ahogy fogalmaztak ezek „szintén a pusztá dichotómiánál komplexebb képet támasztanak alá” (Pléh – Lukács, 2001).

Habár a finn kutatások az agglutináló nyelvek elsajátításának, a beszédfeldolgozásnak és beszédprodukciónak érdekes új megközelítését mutatják be, a duális modell használata Durrant (2013) szerint túlságosan leegyszerűsíti és limitálja is azt. Ezt két okkal is indokolja. Egyrészt a duális modell nem egyeztethető össze a használat alapú nyelvelméleti modellekkel, mert többen kimutatták, hogy a formulaszerű egységek között gyakoribbak a nem teljesen rögzült egységek (vö. Biber, 2009; Cheng et al., 2006; Moon, 1998; Sinclair, 2004), ezzel szemben a fenti kutatásokból azt láthatjuk, hogy az eredményeik alapján csak a teljesen rögzült lexikai egységeket tartják a mentális lexikonban holisztikusan tároltnak. Durrant még azt is megemlíti, hogy a pszicholingvisztikai kutatások (vö. Gibbs et al., 1989) az idiómákkal kapcsolatban kimutatták, hogy a formulaszerű egységeken belüli rugalmasság nagyban hozzájárul azok feldolgozásához (Durrant, 2013: 6). A másik indokként a konnekcionalista elméleteket és a priming jelenségének figyelembe nem vételét jegyzi meg. Ehhez kapcsolódóan Hay és Baayen (2005) kutatását említi, akik vizsgálataik során kimutatták, hogy a részlegesen egyező szavaknál is végbemegy a priming folyamata. A szavak nem csak a különböző toldalékaikat hívják elő, mint például a múlt idő jele, hanem bizonyos fonémaelemek kapcsolódása is előhívja a hasonló alakú és jelentésű szavakat. Példájuk a folyékonyssággal, vízzel kapcsolatos *fl* kezdetű szavak az angolban [*flow, float, flood*] (Durrant, 2013: 6).

### 3.1.2 A török nyelvvel kapcsolatos kutatások

Bár Pierce már az 1960-as években vizsgálta a török nyelvet korpusznyelvészeti eszközökkel (vö. Pierce, 1961), a *Türkçe Ulusal Derlemi* (a továbbiakban TUD), vagyis a török nemzeti korpusz megjelenésével indultak el igazán a korpuszalapú kutatások. A TUD a brit nemzeti korpusz (BNC) mintájára készült, 50 millió szót tartalmazó általános korpusz. A korpuszba 1990 és 2009 között készült írott nyelvi anyagokat (98%) gyűjtötték össze, amelyek műfajukat illetően sokszínűek, valamint található benne néhány beszélt nyelvi anyag (2%) leírata is (TUD, 2013). A török többszavas kifejezéseket [multi-world expressions] többen is vizsgálták ezen a korpuszon. Ilyen Mersinli (2015) kutatása, aki a többszavas kifejezések nyelvi relevanciájával foglalkozott. Aksan és Aksan (2015) a többszavas kifejezések diskurzusban betöltött funkcióit

elemezték, valamint megállapították, hogy a különböző regiszterekben mely egységek a legjellemzőbbek. Egy másik kutatás, amelyet érdemes megemlíteni, az Mersinli és Aksan (2016) kutatása, akik a török nyelv agglutináló mivoltát figyelembe véve írták le módszertani ajánlásaikat a többszavas kifejezések korpuszból történő kinyerésére (Aksan – Aksan, 2014: 4). Mersinli, Aksan és Aksan a többszavas kifejezések keresésénél két és három szóból álló egységeket kerestek, tehát a Dóla (2014: 20) által is javasolt első megoldást választották. A meghatározott kritériumaik alapján a kételemű egységeknek egymillió szóból százszor, míg a háromelemű egységeknek legalább tizennégyszer kellett megjelennie. Az így talált többszavas kifejezéseket funkciójuk szerint is csoportosították, amelynek megvalósításához Biber (2006) – a későbbiekben a jelen értekezésben is ismertetett – kategóriáit (ld. III.3.1.2 fejezet) vették alapul és konkordanciasorokat használtak (Mersinli – Aksan, 2016: 27–31).

A formulaszerű nyelvhasználatot használatalapú nyelvelméleti keretben, Hoey lexikai előfeszítését is alapul véve egy, a fent említetteknel korábbi kutatásban Durrant (2013) vizsgálta. A török nyelv tanulása, a hosszú, szinte megjegyezhetetlen szavak ösztönözték arra, hogy korpusznyelvészeti eszközökkel mintázatokot találjon a nyelvben. A kutatáshoz nem a TUD korpuszt használta, hanem saját korpuszt épített, amely hét török napilap több mint 700 cikkét tartalmazta. Durrant azért választotta a saját korpusz építését a TUD helyett, mert fontos volt számára, hogy az általa valóban elolvasott cikkeket vizsgálja meg annak érdekében, hogy kimutassa, milyen formulaszerű egységekkel találkozik egy nyelvtanuló, aki fél éven keresztül minden nap elolvas egy újságcikket. A korpuszból különböző gyakorisági mutatókkal rendelkező húsz igét választott ki, amelyeket inflexiós morfémaikkal együtt vizsgált meg. Az eredményeiből látható, hogy bár a nagy gyakorisági mutatókkal rendelkező végződések mind az újonnan alkotott, mind a formulaszerű szerkezetekben megjelennek, ez utóbbi dominanciája egy használatalapú nyelvi rendszerre utal. A leggyakoribb kollokációk két elemből állnak, de vannak olyan lexikai elemek (esetragot tartalmazók vagy a kettős tagadás a *değil* segítségével), amelyek ennél több részből tevődnek össze. Ezeket a három vagy négy morfémaiból álló egységeket Durrant morfémakötegeknek [morphemic bundles] nevezte el. A kutatásából az is kiderül, hogy a beágyazódás [nesting] jelensége a morfémaiknál is megfigyelhető, tehát ezek a három- és négyelemű morfémakötegek összekapcsolódva hosszabb kötegeket is alkothatnak. A vizsgálatából azt is megállapította, hogy a török nyelvű korpuszában a legnagyobb gyakorisági mutatóval a háromelemű morfémakötegek rendelkeznek, a leggyakoribb tíz egység a szöveg igéinek 12,38%-át tették ki. Ezzel szemben a négy szóból állók csak 2,22%-át jelentették (Durrant, 2013: 20). Vizsgálatával Durrant arra mutatott rá, hogy amíg az angol nyelvben a formulaszerű egységek gyakori együttes előfordulása a szavak szintjén valósul meg,



addig az agglutináló nyelvekben a szavak mellett a morfológiai szintet is tekintetbe kell venni. Tehát Dóla (2014: 20) második megoldásához hasonlóan gondolkodott. Következtetései szerint a morféma erős kollokációs kapcsolatba kerülnek a közvetlenül, illetve gyakran nemcsak a közvetlenül mellettük álló elemmel. A morfológiai kombinációk között rengeteg a hasonlóság, amely annak a jele, hogy ezek egy-egy típus különböző variációi lehetnek. A nagy gyakorisági mutatókkal rendelkező morféma-összetevők a lexikával erős kapcsolatban állnak. Ezek a morféma-összetevők formájukban és funkciójukban is átfedésben vannak és az ismétlődő találkozások által, analógiai módon új elemekre terjednek ki. „Az egyes ismétlődő minták bizonyos esetekben önállóan reprezentált egységekként jelenhetnek meg, de valószínűleg más morfológiai mintákkal és bizonyos lexikai gyökökkel való asszociációs hálózatokban léteznek” (Durrant, 2013: 31).

### **3.1.3 A magyar nyelvvel kapcsolatos kutatások**

A magyar nyelvvel kapcsolatban a formulaszerű nyelvhasználatának a kutatása viszonylag korlátozott, a legtöbb tanulmány egyetlen típusára, például a kollokációkra összpontosít (ld. Reményi, 2010). Jelenlegi ismereteim szerint kevés olyan átfogó tanulmány létezik, amely a formulaszerű nyelvhasználat több aspektusát vizsgálja. Továbbá ezt a jelenséget a kutatások többnyire a fordítástudomány területén említik (ld. Heltai, 2004; Heltai – Gósy, 2005). A jelen kutatás kimeneti céljának, vagyis az eredmények tananyagkészítés során történő felhasználása okán a magyar mint idegen nyelv területén végzett vizsgálatokat is lényeges áttekinteni. A magyar mint idegen nyelvvel összefüggésben elsősorban Dóla Mónika foglalkozott ezzel a területtel. Dóla (2018: 28) említi, hogy bár egy 1990-ben megjelent írásban Murvai Olga kiemeli a formulaszerű egységek tanításának hasznosságát, azóta csak néhány kutatás és tanulmány foglalkozott ezzel a témával. Az angol szakirodalomban megjelenő terminológiát is Dóla vezeti be a magyar nyelvtudományba (vö. Dóla, 2006a). Habár ő a KER szerinti B1-es nyelvi szinten lévő magyarul tanuló diákok nyelvhasználatában megjelenő formulaszerű egységek analízisét, vizsgálata mégis kapcsolódik a jelen értekezésben bemutatásra kerülő kutatáshoz. Ahogy fentebb említésre került például ő vetette fel, hogy az angol nyelv eltéréseiből fakadóan, hogyan tudjuk Biber (2006) a formulaszerű egységek beazonosítására szolgáló feltételrendszerét úgy átalakítani, hogy az a magyar nyelvre is érvényes lehessen (vö. Dóla, 2016: 98).

2016-os kutatásában Dóla a formulaszerű egységeknek egy csoportját, a többmorfémás interakciós rutinokat kutatta, amelyeket magyarul tanuló diákok korpuszából kigyűjtve leíró

nyelvészeti, pragmatikai, korpusznyelvészeti, pszicholingvisztikai, szociolingvisztikai, valamint nyelvpedagógiai elméleti keretben vizsgált meg. A beazonosításukra Wray – Namba (2003) és Wray (2009) kritériumrendszerét összedolgozva készítette el saját szempontrendszerét (Dóla, 2016: 181). Hozzá hasonlóan a már korábban – a Praat szoftver bemutatásánál (III.2.2 fejezet) – röviden megemlített lengyel anyanyelvi beszélőkkel folytatott kutatásában Guz (2014) szintén Wray és Namba feltételrendszerét alkalmazta a formulaszerű egységek beazonosításához. A Dóla általa felállított szempontrendszer több részből állt. Egyrészt meghatározta az elemek hosszát (legalább két szövegszó vagy legalább három morféma), másrészt a gyakoriságát (legalább kétszer használja ugyanaz a beszélő ugyanabban a formában és jelentésben, valamint legalább két szövegben előfordul). Ezenkívül a jelentés szempontjából az idiomatikusság, illetve az határozta meg az általa kiválasztott elemeket, hogy a lexikai jelentésük eltér a teljes szerkezet jelentésétől. Részletesen kidolgozott szempontrendszerében kitér még az elem formájára is, amelyet három részre bont. A fonológiai szempontokra (például szünetek hiánya vagy a beszédtempó), a morfofonológiai szempontokra (például a hangrendi illeszkedés), valamint a morfológiai és szintaktikai szempontokra (például egyeztetés, szabályos vagy szabálytalan, produktív vagy nem produktív alakok) (Dóla, 2016: 111, 180). A többmorfémás interakciós rutinok tulajdonságainak meghatározásakor említi, hogy ezeket az elemeket a kompetens nyelvhasználók gyakori beszédhelyzetekben, rutinszerűen alkalmazzák (Dóla, 2006b: 42). A beszélő a használatukkal a társas-interakciós célját szeretné elérni, diskurzus-szervezésre, illetve egy „üzenet vagy reláció kifejezésére” használja őket (Dóla, 2016: 179).

Dóla kutatását egy 20 adatközlővel ECL nyelvvizsgákon készült, 11 hangfelvételtől álló, több mint 21 ezer szövegszót tartalmazó korpuszon végezte el, amelyet nemcsak ő analizált a felállított szempontrendszer segítségével, hanem a validitás érdekében, két független elemzőt is felkért erre a feladatra (Dóla, 2016: 337). A többmorfémás interakciós rutinok beazonosításának a folyamatának első lépéseként a három elemző egymástól függetlenül, a meghatározott szempontrendszer szerint kiválasztotta az általa többmorfémás interakciós rutinoknak gondolt elemsorokat a korpuszból. Ezután együtt megtárgyalták a találatukat, amely egy újabb, letisztázott lista kialakulásához vezetett. A következő lépésben az elemzők két vizsgainterjú anyagán közösen dolgozva elkülönítették a formulaszerű és nem formulaszerű egységeket, majd ezt a munkát Dóla folytatta a fennmaradó interjúkon. Az eddig leírtakból látszik, hogy a kutatás először kvalitatív módszereket használt, hiszen az elemzők saját nyelvtanári tapasztalataikra is építettek a munkájuk során. Az így felsorolásra került többmorfémás interakciós elemsorok standardizációja – amely során Dóla a beszédaktusokat

tekintette alapvetően rendszerező elvnek – után került sor a kvantitatív szakaszra, vagyis a gyakorisági vizsgálatok elvégzésére (Dóla, 2016: 202–204).

## 4. A jelen kutatásban meghatározott feltételrendszer

Ahhoz, hogy meg tudjuk állapítani, hogy az eredményeinkben felsorolt elemek valóban formulaszerű szekvenciák, pontos kritériumokat kell meghatároznunk. A jelen doktori értekezésben bemutatott kutatás folyamán a formulaszerűséget elsősorban a korpuszban való nagy gyakoriság alapján határozzuk meg (vö. Hoey, 2005; Sinclair, 2004; Stubbs, 1996). Ugyanakkor a formulaszerű egységek kiválasztására szolgáló feltételrendszer azok tulajdonságain is alapul (átláthatóság, gyakoriság, gyors feldolgozás, jelentéstartalom, kiszámíthatóság, konvencionalizáltság, megjósolhatóság, pragmatikai funkció, rögzítettség, sematikusság), valamint az agglutináló nyelvek jellegzetességeit figyelembe véve került meghatározásra (Vermeki, 2023b, megjelenés alatt). A következőkben a feltételrendszer pontjait fogom bemutatni részletesen.

### 4.1 Gyakoriság

A feltételrendszer első eleme tehát a gyakoriság. Ahogy már korábban említésre került (II. 4 fejezet), Biber (2006) egyetemi tananyagokban és osztálytermi nyelvben keresett lexikai egységeket, amelyeket ő lexikai kötegeknek [lexical bundles] nevezett. Az általa talált lexikai kötegek három-négy szóból álltak, de kutatásaihoz csak négyszavas egységeket használt. Emellett azt is meghatározta, hogy a kiválasztott lexikai kötegek legalább negyvenszer szerepeljenek az egymillió szóból álló korpuszban (Biber 2006: 133–134). Biber vizsgálatát, Dóla (2016) magyar nyelvvel és Mersinli – Aksan (2016) török nyelvvel kapcsolatos kutatását figyelembe véve határoztam meg a jelen kutatás kiválasztott lexikai egységeire vonatkozó gyakorisági mutatókat. Az abszolút gyakorisági mutatónál kételemű egységek esetén legalább 100 előfordulás per egymillió szóra, háromelemű egységeknél 14 előfordulás per egymillió szóra, a négyelemű egységeknél pedig legalább 4 előfordulás per egymillió szóra volt szükség ahhoz, hogy kiválasztásra kerüljenek (Vermeki, 2023b, megjelenés alatt).

| Lexikai egység típusa      | Minimum gyakorisági mutató     |
|----------------------------|--------------------------------|
| kételemű egység [2-gram]   | 100 előfordulás / 1 millió szó |
| háromelemű egység [3-gram] | 14 előfordulás / 1 millió szó  |
| négyelemű egység [4-gram]  | 4 előfordulás / 1 millió szó   |

A kutatás során a lexikai egységek kiválasztásához a fent említett abszolút gyakorisági mutatón kívül még két mutatót vettem figyelembe: a dokumentum gyakoriságot és a relatív dokumentum gyakoriságot. Az első azt mutatja meg, hogy hány feltöltött dokumentum tartalmazza a listában szereplő elemeket, amíg a relatív dokumentum gyakoriság a dokumentumok százalékos arányát mutatja, amelyekben a listában szereplő elemek szerepelnek (Vermeki, 2023b, megjelenés alatt). Ezzel az volt a célom, hogy az eredmények különböző adatközlők nyelvhasználatát tükrözzék és ne egyetlen beszélő nyelvhasználatára korlátozódjanak. Az abszolút gyakorisági mutatók alapján kiválasztott elemek számát így a következőképpen csökkentettem le. Egyelemű lexémák esetében a kiválasztott elemeknek a dokumentumok legalább felében meg kellett jelenniük. Kételemű egységnél már ennek a számnak a felét határoztam meg, tehát a dokumentumok legalább egy negyedében kellett szerepelniük. A háromelemű egységek esetében úgy adtam meg, hogy legalább öt dokumentumban jelenjenek meg, amíg a négyelemű egységeknél ezt a számot háromban határoztam meg. Az ötös és a hármas szám nemcsak a feltöltött dokumentumokra utal, hanem arra is, hogy legalább ennyi gyermek szövegében jelenjenek meg a kiválasztott egységek.

| Lexikai egység típusa      | Minimum dokumentum gyakoriság | Minimum relatív dokumentum gyakoriság (%) |
|----------------------------|-------------------------------|---|
| egyelemű lexéma            | 39                            | 50  |
| kételemű egység [2-gram]   | 20                            | 25  |
| háromelemű egység [3-gram] | 5                             | 6,4                                       |
| négyelemű egység [4-gram]  | 3                             | 3,8                                       |

## 4.2 A formulaszerű egységek mérete

A Dóla (2016: 98) által említett két módszerből – kisebb lexikai egységek, illetve az igekötők és toldalékok definícióba történő bevonása – az elsőt választottam a formulaszerű egységek méretének meghatározásánál. A jelen kutatás folyamán keresett többelemű lexikai egységek ezért két vagy több szóból állnak. Ugyanakkor Siyanova-Chanturia és Pellicer-Sánchez (2019: 2) javaslata alapján a formulaszerű egységek közül nem zárom ki az egyelemű lexémákat sem.

### 4.3 Funkció és jelentéstartalom

A gyakoriság alapján kiválasztott lexikai egységeket a Biber által meghatározott, szövegben betöltött szerepük szerinti kategóriákba kerülnek besorolásra (18. táblázat). A legtöbb lexikai egység a szövegek építőelemeként működik [building blocks] (Hunston, 2022: 103–104), Biber (2006) ezeket funkcionális szempontból három nagyobb és több alkategóriára osztotta. Az első az álláspont kifejezésére szolgáló lexikai egységek [stance bundles], amelyek keretet adnak az utána következő állításnak. A beszélő segítségével kifejezheti, hogy valami biztos, bizonytalan vagy lehetséges, illetve a következő részben leírt eseményhez vagy cselekményhez való viszonyát is meghatározhatja általuk. Ezek kifejezhetnek kívánságokat, kötelezettségeket, szándékokat, jóslatokat, képességeket. Alkategóriái az episztemikus kifejezések és a szemléletet, modalitást kifejező lexikai egységek (vágy, óhaj, utasítás, szándék, előfeszítés, képesség, kötelezettség). A második a diskurzus-szervező kifejezések [discourse organising bundles], amelyeket a téma bevezetése, a témára való fókuszálás elérése, valamint a téma pontosítása és kidolgozása céljából használunk. Az utolsó kategória a hivatkozás kifejezésére szolgáló lexikai egységek [referential bundles], amelyek használatával a beszélő beazonosít valamit vagy valaminek a szerinte fontos tulajdonságát. Alkategóriái a pontatlanság indikálása, a tulajdonság megadása, az idő és hely megadása, valamint a multifunkcionális hivatkozás.

18. táblázat A lexikai egységek szövegben betöltött szerepük szerint Biber (2006: 139-146) alapján (Vermekei, 2023b, megjelenés alatt)

| Fő kategória                           | Funkció   | Alkategória   | Biber példái   | A KorSzak Gyermeknyelvi Korpusz példái                        |
|--|---|---|--|---|
| Álláspont kifejezések [stance bundles] | keretet adása, a beszélő kifejezheti, hogy valami biztos, bizonytalan vagy lehetséges, a cselekményhez való viszonyát, kifejezhetnek kívánságokat, kötelezettségeket, szándékokat, jóslatokat, képességeket | episztemikus kifejezések  | <i>I don't know what, I think it was, are more likely to</i>                                   | <i>nem tudom, nem is tudom</i>                                |
|  |   | szemléletet, modalitást kifejező lexikai egységek (vágy, óhaj, utasítás, szándék, előfeszítés, képesség, kötelezettség) | <i>you need to know, I don't want to, you might want to, it is important to, to be able to</i> | <i>nagyon jó lenne, nekem az a kedvencem, én is szeretnék</i> |
| Diskurzus-szervező kifejezések         | a téma bevezetése, a témára való fókuszálás elérése,  | témabevezetés   | <i>want to talk about, going to</i>  | <i>az lenne a kérdésem, akkor ott van a</i>                   |

| Főkategória                                 | Funkció   | Alkategória                  | Biber példái  | A KorSzak Gyermeknyelvi Korpusz példái                         |
|---|---|------------------------------|---|--|
| [discourse organising bundles]              | valamint a téma pontosítása és kidolgozása                                    |                              | <i>talk about, if you look at</i>   |  |
|   |   | a téma pontosítása           | <i>has to do with the, you know I mean, as well as the</i>                | <i>ez a, erről a, volt az az</i>                               |
|   |   | azonosítás, fókusz           | <i>those of you who, that's one of the, one of the things</i>             | <i>az volt a, neki volt egy</i>                                |
| Hivatkozás kifejezése [referential bundles] | a beszélő beazonosít valamit, vagy valaminek a szerinte fontos tulajdonságát. | pontatlanság indikálása      | <i>or something like that, and things like that</i>                       | <i>vagy hogy mondjam, vagy nem is tudom, vagy mit tudom én</i> |
|   |   | tulajdonság megadása         | <i>have a lot of, the size of the, the nature of the, in terms of the</i> | <i>az ilyen nagy(obb), de ő is nagyon +Adj.</i>                |
|   |   | idő és hely megadása         | <i>at the end of the year, as shown in figure</i>                         | <i>egy idő után, amikor kicsi voltam, akkor ott volt</i>       |
|   |   | multifunkcionális hivatkozás | <i>at the end of, the end of the</i>                                      | <i>az volt az első, és akkor így</i>                           |

#### 4.4 Feldolgozás és produkció

A kutatás szempontjából fontos megvizsgálni, hogy a különböző lexikai egységek hogyan kapcsolódnak egymáshoz a beszédprodukció során. A kiválasztott többelemű lexikai egységek között számos olyan található, amelyre a beágyazottság [nesting] jellemző (vö. Hoey, 2005). Tehát egy kisebb lexikai egység egy nagyobb részét képezi. Ezeknek az egységeknek az esetében, valamint a részben rögzült formulaszerű egységeknél, amelyeknél a korpuszban való előfordulásuknál még a szórend is különböző, megvizsgáltam, hogy a beszédprodukció közben hol tartanak szünetet a kutatásban részt vevő gyermekek. Közülük azok az elemek kerültek kiválasztásra, amelyeknél az adott egység kimondása közben nem tartottak szünetet a gyermekek, a szünetre az egység előtt és után került sor (Vermekei, 2023b, megjelenés alatt).

## 5. Kutatásetika

A jelen kutatás az általánosan elfogadott etikai normák betartásával, az adatközlők személyiségi jogainak tiszteletben tartásával jött létre. Minden kutatás során, illetve egy kutatás minden szakaszában – a témaválasztástól a kutatás megtervezésén, az adatgyűjtésen át egészen az eredmények értékeléséig és publikussá tételéig – megfogalmazódhatnak etikai kérdések, amelyek vonatkozhatnak az egyén magatartására, vagy a kutatás eredményeinek társadalmi felhasználással kapcsolatos elvárásokra (Fóris, 2013: 208). A korpusznyelvészetről szóló szakirodalomban kevesen tárgyalják a kifejezetten erre a területre vonatkozó kutatásetikai kérdéseket. McEnery és Hardie (2012: 60) arról számol be, hogy ennek az az oka, hogy a korpusznyelvészet területén kutatók már létező irányvonalakat alkalmaznak erre a célra, ilyen a Brit Alkalmazott Nyelvészeti Egyesület [British Association of Applied Linguistics] által kidolgozott dokumentum (BAAL, 2021), amelyben ajánlásokat fogalmaztak meg az alkalmazott nyelvészeti kutatások különböző szereplőinek felelősségére vonatkozóan. Ezek mellett az általános elvek mellett, a korpusznyelvészeti kutatásokkal kapcsolatban néhány sajátos kutatásetikai kérdés is felmerül. Ezek négy fő csoportba oszthatók:

1. a korpuszépítőkre vonatkozó,
2. a beszélt nyelvi korpuszok adatközlőire vonatkozó,
3. a korpusz megosztására, hozzáférhetőségére vonatkozó,
4. valamint a korpuszban összegyűjtött adatokat felhasználókra vonatkozó kérdések (McEnery – Hardie, 2012: 61–69).

A fent felsorolt négy pont alapján, valamint az általános alkalmazott nyelvészeti kutatásokra vonatkoztatható irányelveket is figyelembe véve mutatom be, hogy a jelen kutatás során, milyen kutatásetikai irányelvek betartását tartottam szem előtt.

### 5.1 A korpuszépítéssel, a korpuszépítőkkal kapcsolatos etikai irányelvek

A témaválasztás, a kutatás megtervezése, valamint az adatgyűjtési protokoll minden részletében alkalmazkodik ahhoz a tényhez, hogy a kutatásban résztvevő adatközlők kiskorúak. A beszélgetések során csak a gyermekek életkorának megfelelő témák feldolgozására került sor. Az adatközlők szülei, az iskolaigazgatók a kutatásról és az adatgyűjtés menetéről a felvételek



rögzítését megelőzően írásos formában kaptak általános tájékoztatást (ld. Mellékletek). A tájékoztatásban szerepelt a korpuszépítés céljának leírása, hogyan zajlik a felvételek rögzítése, valamint, hogy milyen adatokat tárolunk és azok milyen módon kerülnek felhasználásra (O’Keeffe – McCarthy, 2010: 42–43).

Az adatgyűjtésben csak olyan gyermekek vettek részt, akiknek a szülei aláírták a személyes adataik felhasználásáról az Európai Unió általános adatvédelmi rendelete (General Data Protection Regulation, 2018, röviden GDPR) alapján készült hozzájáruló nyilatkozatot. A GDPR alapelvei, amelyeket a korpuszépítés során is szem előtt tartottunk a következők:

- jogszerűség, tisztesség és átláthatóság,
- célhoz kötöttség,
- adatminimalizálás,
- pontosság,
- tárolási korlátozás,
- integritás és titoktartás (biztonság),
- elszámoltathatóság (GDPR, 2018).

Ezenkívül kitöltötték és aláírták a kutatásban való részvételről szóló beleegyező nyilatkozatot, amelyben több lehetőség közül is választhattak felhasználásra (videó vagy hangfelvétel rögzítése; a kutatás során a felvétel vagy kizárólag csak a felvétel leiratának felhasználása) (vö. O’Keeffe – McCarthy, 2010: 43). A szülők mellett természetesen a gyermekek is beleegyeztek a felvételek elkészítésébe, a rögzítés során bármikor leállíthatták a felvételt, illetve, ha nem ők kezelték a rögzítő eszközt, kérhették annak leállítását. Így az adatgyűjtés során nem merült fel obtruzív eljárási forma, az adatközlők önként vettek részt a folyamatban. Az adatközlőknek továbbá jogában áll bármikor visszalépni az adatgyűjtés során vagy azt követően (BAAL, 2021).

## **5.2 A beszélt nyelvi korpuszok adatközlőivel kapcsolatos etikai irányelvek**

Az adatgyűjtés során tehát a GDPR alapelveinek megfelelően csak a kutatáshoz szükséges személyes adatok kerültek begyűjtésre (életkor, nem, megye, tanult idegen nyelv, felvétel készítésének éve). A beszélt nyelvi korpuszok építésénél három tényezőre kell figyelni kutatásetikai szempontból:

1. a megnyilatkozók anonimitása, nem beazonosíthatósága

2. a hallgatók anonimitása, nem beazonosíthatósága
3. a beszélgetés során felmerülő, megemlített személyek anonimitása, nem beazonosíthatósága (McEnery – Hardie, 2012: 61–64).

A videó és hanganyagok leiratainak elkészítése után a fenti kérdéseket figyelembe véve kicseréltem az olyan személyes adatokat, amelyekből az adatközlők vagy azok a nem közszereplő személyek, akiket a beszélgetések során megemlítenek felismerhetőek lennének. Más személyneveket, illetve néhány alkalommal más településneveket írtam az eredeti adatok helyére. Fontosnak tartottam, hogy ne számokkal vagy kódokkal jelöljem a személyeket, mert a rögzített szövegekben előfordulnak olyan megnyilatkozások, ahol az illetők megszólítása vagy ahogyan beszélnek róluk érdekes lehet egy későbbi kutatás során (becézés, kitalált nevek). A települések neveit pedig csak abban az esetben változtattam meg, ha abból beazonosítható a személy, aki éppen beszél, akinek beszélnek vagy akiről szó van. Az adatközlők neveit kódoltam, így a korpusz adatait tartalmazó összefoglaló adatbázisban már csak a kódok szerepelnek. A leiratokban lévő adatok így váltak teljesen anonimá. A videó és hangfelvételek anonimizálása már egy kicsit bonyolultabb feladat. A videón, illetve a hangjuk alapján könnyen felismerhetőek az adatközlők. A videó képének elmosása, valamint a hangtorzítás lehet erre egy megoldás, de így az anyag alkalmatlanná válik bizonyos vizsgálatok elvégzésére (prozódia, fonetikai kutatások), ezért O’Keeffe és McCarthy (2010: 43) azt ajánlja, hogy mindenképpen ismertessük ezt a problémát az adatközlőkkel, jelen esetben az adatközlők szüleivel is, és csak ezután járuljanak hozzá a felvételek felhasználásához. Ezért került bele a már említett beleegyező nyilatkozatba két lehetőség. Első lehetőségként a videó és/vagy hanganyag felhasználásának engedélyezése vagy másodikként csak a leiratok felhasználásának engedélyezése.

### **5.3 A korpusz megosztásával, hozzáférhetőségével kapcsolatos etikai irányelvek**

A korpusz hozzáférhetőségével kapcsolatban két etikai kérdés merülhet fel. Az egyik, hogy a korpusz építőinek célja eltérhet a korpuszt használók céljától, tehát a korpusz bármilyen célra felhasználható lehet, amit mindenképpen végig kell gondolni a korpusz megosztása előtt. A másik pedig, hogy a korpusz adatai épen és elérhetően maradjanak, hiszen a korpusznyelvészet egyik fontos célja a megismételhetőségre való törekvés. A korpusz készítőinek biztosítaniuk kell, hogy más kutatók is hozzáférjenek az általuk gyűjtött adatokhoz (McEnery – Hardie, 2012:

65–66). Jelenleg a KorSzak Gyermeknyelvi Korpusz videó és hangfelvételekhez a KorSzak Munkacsoport azon tagjai férhetnek hozzá, akik a Gyermeknyelvi kutatócsoportban dolgoznak. A kutatócsoportba való belépés bárki számára nyitott. A felvételeket egy jelszóval védett mappában, valamint egy biometrikus azonosítóval (ujjlenyomat) ellátott tárolóegységen tartom. A leiratok nyilvánosságra hozásába minden szülő beleegyezett, a korpusz a Sketch Engine felületére került feltöltésre. A KorSzak munkacsoportba bárkinek lehetősége van belépni, munkájával támogatni a korpuszépítés folyamatát. A munkacsoport tagjai a közösen épített korpuszokon végezhetnek további korpuszalapú kutatásokat vagy megismételhetnek egy már elvégzett kutatást.

#### **5.4 A korpuszban összegyűjtött adatokat felhasználókkal kapcsolatos etikai irányelvek**

A korpuszok felhasználóival kapcsolatban is számos etikai kérdés merül fel (vö. McEnery – Hardie, 2012: 66), amelyeket szem előtt tartva a jelen kutatás során az adatok elemzését minél részletesebben szándékozom közölni, hogy egyrészt a kutatás maga megismételhető legyen másrészt pedig, hogy az eredmények egyértelműek legyenek. A Sketch Engine automatikus szófaji, lemmák szerinti, valamint szintaktikai címkézésén kívül más módon nem annotálom a korpuszt, hogy azt bárki egyszerűen felhasználhassa. Ugyanakkor a jövőbeli kutatások megkönnyítése érdekében metaadatokkal (a már az előző részben említett életkor, lakóhely, nem, felvétel készítésének éve, témakör, altémakör, beszélők száma, a szöveg típusa, tanult idegen nyelv) bővítettem, amely által a korpuszban való keresés könnyebbé válhat mások számára is.

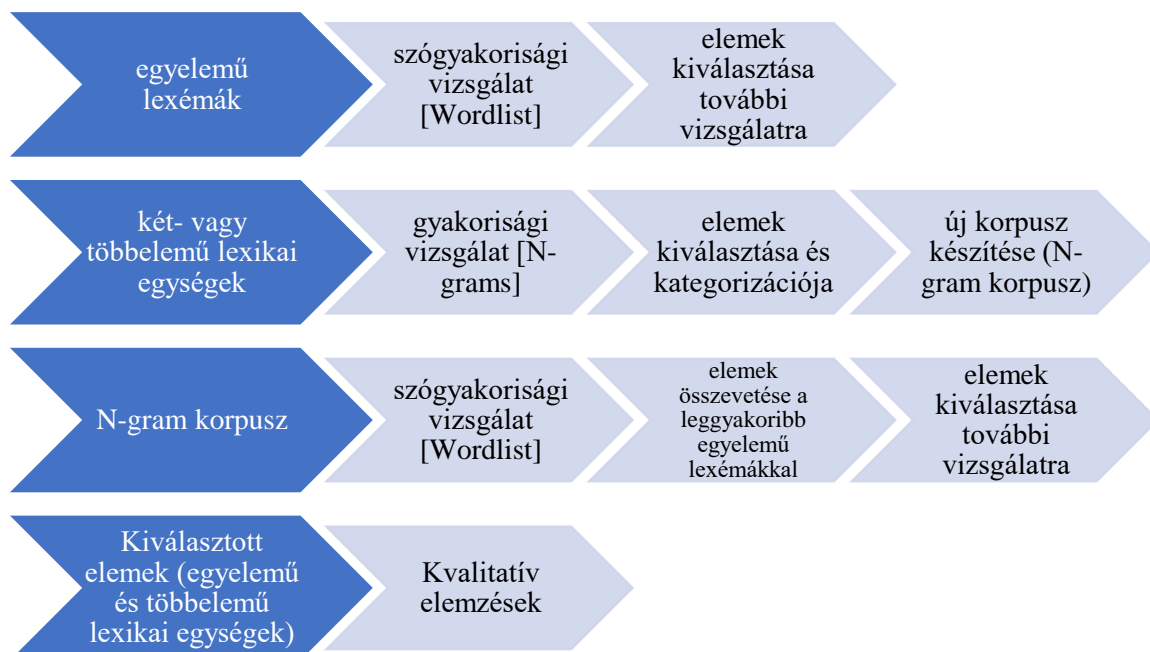
## IV. A kutatás menete és eredményei

### 1. Bevezetés

Az előző fejezetben kiemeltük, hogy a bemutatásra kerülő eredmények egy magyarázó szekvenciális tipológia alapján készült kutatás eredményei, amelynek során a kvantitatív vizsgálatokat kvalitatív elemzések követik. Annak érdekében, hogy minél átfogóbb képet nyújtson a kutatás a vizsgált jelenségekről, több szekvenciából tevődött össze, amelyek alapvetően két meghatározott pontból indultak ki:

- egyelemű lexémák
- két vagy több elemből álló lexikai egységek

68. ábra A kutatás kvantitatív és kvalitatív vizsgálatok szekvenciái



Ebben a fejezetben részletesen ismertetésre kerülnek a 68. ábrán bemutatott kutatási szekvenciák során kapott eredmények. Az alfejezetek első részében a kvantitatív jellegű vizsgálatokra fókuszálunk, amelyek célja, hogy pontosabb képet kapjunk a korpuszban található lexikai egységek eloszlásáról és gyakoriságáról. Az alfejezetek második részében a

kvalitatív vizsgálatok kerülnek előtérbe, amelyek segítenek mélyebb betekintést nyerni a korpuszban található gyakori lexikai egységek használati mintázataiba, kolligációiba, a korpuszban szereplő gyermekek beszélgetéseiben betöltött szerepeibe és jelentésébe. Elsőként tehát a korpuszban talált egyelemű lexémák bemutatására kerül sor, amelyet kétféle megközelítéssel teszünk. A szógyakorisági vizsgálatok után az egyes szófajokon belüli gyakorisági vizsgálatok szintén ismertetésre kerülnek. Ezután a korpuszban talált többelemű lexikai egységek ismertetése következik. A fejezet elején felsorolt gyakori lexikai egységek közül szintén kiemelésre kerül néhány a tananyagkészítés szempontjából is hasznos elem, amelyeknek a részletesebb elemzésére a gyakorisági vizsgálatok eredményeit bemutató alfejezet után kerül sor. A fejezet végén kerül sor a kiválasztott lexikai egységeknek a szövegben betöltött szerepe szerinti funkcióinak bemutatására.

Az eredmények értelmezésével kapcsolatosan lényeges megemlíteni, hogy a korpuszban szereplő gyermekek beszéde nem reprezentatív az összes nyelvhasználó gyermek beszédére nézve (vö. Hoey, 2005: 14). Ebből adódóan nem vonhatunk le a magyarul beszélő teljes 11–15 éves korosztályra vonatkozó következtetéseket, ahogy azt sem állapíthatjuk meg, hogy egy-egy beszélő az élete során, 11–15 éves korában pontosan ugyanilyen nyelvhasználati példákkal találkozhat. Az eredményekből levont következtetések tehát csak a fókuszkorpuszban szereplő adatközlő gyermekek adott beszédprodukcióira vonatkoznak. Ugyanakkor miként Hoey (2005: 14) is kijelenti, a korpuszok jelezhetik, hogy milyen típusú adatokkal találkozhat egy nyelvhasználó, valamint a szavak és szókapcsolatok tulajdonságait is indikálhatja. A korpusz adatai lehetővé teszik, hogy részletesen megvizsgáljuk a benne szereplő gyermekek nyelvhasználatának különböző aspektusait. A vizsgálati eredmények által számos információ gyűjthetünk, amely jelentős szerepet játszhat a hatékony oktatási anyagok készítésében.

## 2. Egyelemű lexémák

### 2.1 A szógyakorisági vizsgálatok eredményei

*1. kutatási kérdés: Melyek a leggyakoribb egyelemű lexikai egységek a KorSzak Gyermeknyelvi Korpuszban?*

Az elvégzett szógyakorisági vizsgálatokkal – amely az előző fejezetben említett Sketch Engine Wordlist eszköz segítségével történt – sikerült részletesen feltérképezni a KorSzak Gyermeknyelvi Korpuszban található leggyakrabban előforduló lemmákat, illetve szófajonként osztályozva: a főneveket, mellékneveket és igéket. Ennek az elemzésnek az elvégzése két okból volt szükséges. Egyrészt hozzájárult az ezután bemutatásra kerülő formulaszerű lexikai elemek kiválasztásához. Másrészt pedig ezek az adatok kiválóan alkalmasak a készülő korpuszinformált tananyag lexikájának összeállítására, hiszen a tanulók számára releváns, gyakran előforduló szavakat tartalmazzák.

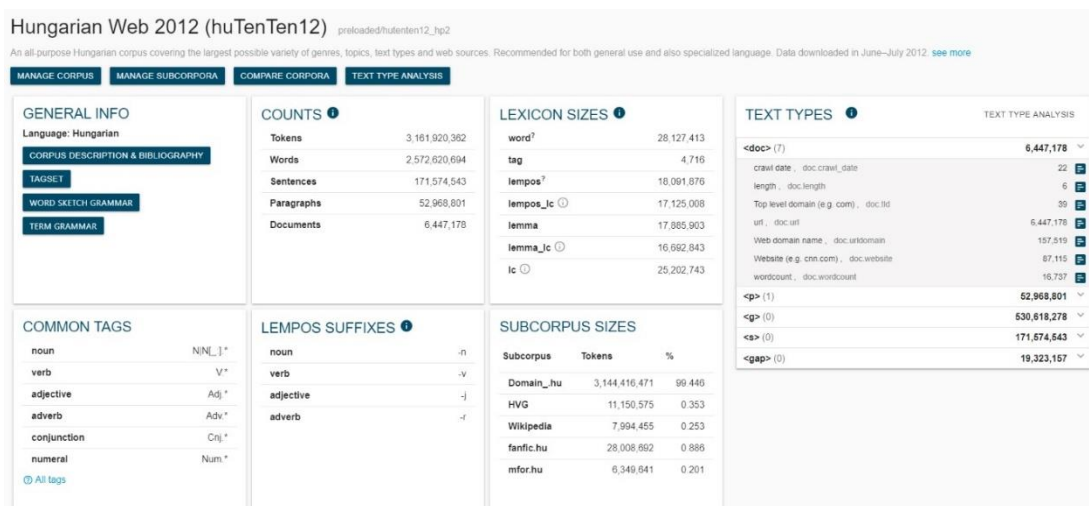
A KorSzak Gyermeknyelvi Korpuszból<sup>20</sup> kapott eredményeket a Hungarian Web 2012 (69. ábra) korpuszal hasonlítottam össze (Vermekei, 2023c, megjelenés alatt). A referenciakorpuszként használt Hungarian Web 2012, vagyis a huTenTen2012 egy magyar nyelvű, a világhálóról összeállított szövegtörzs, amely a TenTen korpuszcsoport része, és amelynek célkitűzése egy olyan több mint 10 milliárd szóból álló korpusz létrehozása, amelynek a különböző nyelveken gyűjtött részei azonos módszertant alkalmazva épülnek fel. A Sketch Engine jelenleg több mint 30 nyelven kínál hozzáférést ezekhez a korpuszokhoz. A magyar változat adatait 2012 júniusában gyűjtötték össze a SpiderLing szoftver segítségével, így egy körülbelül 2,5 milliárd szavas korpuszt kaptak (huTenTen2012, é.n.).

Mivel a referenciakorpusz írott nyelvi szövegeket, elsősorban cikket tartalmaz, elsődleges különbségek az írott és beszélt nyelv közötti eltérésekből adódnak. Ugyanakkor arra megfelelő eszköz ez a referenciakorpusz, hogy a segítségével kizárjuk azokat az elemeket, amelyekkel nem érdemes foglalkozni a kutatás során, illetve kiemeljük azokat, amelyeket részletesebben is érdekes lehet megvizsgálni.

---

<sup>20</sup> A következőkben közölt eredmények a KorSzak Gyermeknyelvi Korpusz 2022. októberi állapotára vonatkoznak.

## 69. ábra A huTenTen12 korpusz adatai



Az első vizsgálat során a korpuszban található leggyakoribb 100 lemmát kerestem meg a Wordlist funkció segítségével. Az így elkészült listát a könnyebb elemzés érdekében három szakaszra bontottam. Az első határt az első nyolc, vagyis a KorSzak Gyermekegyelvi Korpuszban megjelenő ezer fölötti abszolút gyakorisági mutatóval rendelkező lemma után húztam meg. A másodikat pedig az ötszáz előfordulás fölötti, vagyis az első húsz leggyakoribb lemma után (Vermekei, 2023c, megjelenés alatt).

19. táblázat A KorSzak Gyermekegyelvi Korpusz és a huTenTen12 első nyolc leggyakoribb lemmája

| KorSzak Gyermekegyelvi Korpusz |       |       | huTenTen12 |       |             |
|--------------------------------|-------|-------|------------|-------|-------------|
| Sz.                            | Lemma | Gyak. | Sz.        | Lemma | Gyak.       |
| 1.                             | A     | 2903  | 1.         | A     | 215.513.554 |
| 2.                             | VAN   | 2309  | 2.         | AZ    | 104.271.723 |
| 3.                             | AZ    | 2142  | 3.         | ÉS    | 56.483.438  |
| 4.                             | ÉS    | 1687  | 4.         | VAN   | 49.151.567  |
| 5.                             | NEM   | 1412  | 5.         | NEM   | 39.532.965  |
| 6.                             | AHOGY | 1325  | 6.         | IS    | 36.030.809  |
| 7.                             | IS    | 1218  | 7.         | EZ    | 28.283.027  |
| 8.                             | ÉN    | 1132  | 8.         | EGY   | 28.016.414  |

Az első szakaszt ezután összehasonlítottam a huTenTen12 első nyolc lemmájával (19. táblázat). Két lemma kivételével az eredmények ennél a szakasznál még nagyon hasonlóak, ahogy várható is volt, hiszen a nyelvekben általában a leggyakoribb szavak a névelők – ha megtalálható az adott nyelvben – és a létige, a kötőszavak, illetve a személyes névmások, mert ezek alapvető nyelvi elemek, amelyek nélkül nehéz volna a kommunikáció. A két korpusz között a különbség csak annyi volt ebben a szakaszban, hogy a KorSzak Gyermeknyelvi Korpusz első nyolc lemmája között szerepel az AHOGY, amely valójában a HOGY lemma (ld. III.2.1 fejezet) és az ÉN lemma. Ezzel szemben a referenciakorpuszban az AHOGY a kilencedik és a tizedik helyen áll (itt a Sketch Engine két lemmára szedte szét, de mindkettő a *hogy* kötőszó szótöveként szerepel a listában), míg az ÉN csak jóval később, a huszonkettedik helyen jelenik meg. A személyes névmás korai megjelenése minden bizonnyal a beszélt nyelv egyik sajátosságának tudható be. Ezenkívül még egy apró, de érdekes különbséget figyelhetünk meg: a fókuszkorpuszban a VAN lemma az A és AZ névelők közé ékelődött be, amíg a referenciakorpuszban a negyedik helyen jelenik meg.

**20. táblázat** A KorSzak Gyermeknyelvi Korpusz és a huTenTen12 9-20. leggyakoribb lemmája

| KorSzak Gyermeknyelvi Korpusz |       |       |                        | huTenTen12 |       |            |                        |
|-------------------------------|-------|-------|------------------------|------------|-------|------------|------------------------|
| Sz.                           | Lemma | Gyak. | Gyakoriság<br>/ millió | Sz.        | Lemma | Gyak.      | Gyakoriság<br>/ millió |
| 9.                            | MEG   | 863   | 12.062,17              | 9.         | AHOGY | 27.668.793 | 8750,63                |
| 10.                           | EZ    | 827   | 11.559,00              | 10.        | HOGY  | 17.204.243 | 5441,07                |
| 11.                           | EGY   | 812   | 11.349,34              | 11.        | DE    | 15.606.151 | 4935,66                |
| 12.                           | DE    | 769   | 10.748,33              | 12.        | MEG   | 13.984.050 | 4422,64                |
| 13.                           | TUD   | 650   | 9085,06                | 13.        | HA    | 12.399.513 | 3921,51                |
| 14.                           | ILYEN | 597   | 8344,28                | 14.        | CSAK  | 11.786.478 | 3727,63                |
| 15.                           | AKKOR | 592   | 8274,40                | 15.        | MÁR   | 11.301.444 | 3574,23                |
| 16.                           | JÓ    | 583   | 8148,60                | 16.        | AMI   | 11.295.300 | 3572,29                |
| 17.                           | IGEN  | 561   | 7841,11                | 17.        | JÓ    | 10.781.347 | 3409,75                |
| 18.                           | HÁT   | 559   | 7813,16                | 18.        | MÉG   | 10.000.704 | 3162,86                |



| KorSzak GyermeKNyelvi Korpusz |        |       |                     | huTenTen12 |       |           |                     |
|-------------------------------|--------|-------|---------------------|------------|-------|-----------|---------------------|
| Sz.                           | Lemma  | Gyak. | Gyakoriság / millió | Sz.        | Lemma | Gyak.     | Gyakoriság / millió |
| 19.                           | NAGYON | 556   | 7771,22             | 19.        | KELL  | 9.962.040 | 3150,63             |
| 20.                           | Ő      | 515   | 7198,17             | 20.        | Ő     | 9.844.353 | 3113,41             |

Bár a legnagyobb eltéréseket majd csak a harmadik szakaszban (20–100. hely) találjuk, a második részben (9–20. hely) a huTenTen12 korpuszhoz képest már több különbség is megfigyelhető (20. táblázat). A második szakaszban két lemma is előfordul (HÁT, IGEN), amely egyáltalán nem szerepel a referenciakorpusz első száz lemmájának listáján. Ez szintén a beszélt és az írott nyelv különbségének következménye, illetve a diskurzusnak a jellemzője. Ezenkívül a fókuszkorpuszban szerepel több olyan lemma is, amelyek jóval nagyobb gyakorisági mutatóval (gyakoriság per egymillió szót) rendelkeznek (21. táblázat), mint a referenciakorpuszban lévő párjuk (Vermekei, 2023c, megjelenés alatt).

**21. táblázat** A szignifikánsan nagyobb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező itemek a 9–20. leggyakoribb lemma közül

| Lemma  | KorSzak GyermeKNyelvi Korpusz |                     | huTenTen12 |                     | Eltérés (%) |
|--------|-------------------------------|---------------------|------------|---------------------|-------------|
|        | Sz.                           | Gyakoriság / millió | Sz.        | Gyakoriság / millió |             |
|        | ILYEN                         | 14.                 | 8344,28    | 66.                 |             |
| NAGYON | 19.                           | 7771,22             | 41.        | 1614,00             | <b>80</b>   |
| AKKOR  | 15.                           | 8274,40             | 32.        | 2079,59             | <b>75</b>   |
| TUD    | 13.                           | 9085,06             | 23.        | 2899,36             | <b>71</b>   |
| MEG    | 9.                            | 1062,17             | 12.        | 4422,64             | <b>64</b>   |
| JÓ     | 16.                           | 8148,60             | 17.        | 3409,75             | <b>59</b>   |
| DE     | 12.                           | 10748,33            | 11.        | 4935,66             | <b>55</b>   |

| Lemma | KorSzak       |                     |            |                     | Eltérés (%) |
|-------|---------------|---------------------|------------|---------------------|-------------|
|       | Gyermeknyelvi |                     | huTenTen12 |                     |             |
|       | Korpusz       |                     |            |                     |             |
|       | Sz.           | Gyakoriság / millió | Sz.        | Gyakoriság / millió |             |
| EGY   | 11.           | 11349,34            | 8.         | 8860,57             | 22          |

A négy legnagyobb eltérést mutató lemma közül az ILYEN 88%-kal, a NAGYON 80%-kal, az AKKOR 75%-kal és a TUD 71%-kal magasabb értéket mutat. Ezek mindegyike a fókuszkorpusz többemű lexikai egységeiben is több százszor megjelenik (22. táblázat), de ezekről majd a későbbi fejezetekben (ld. IV.3 fejezet) lesz szó részletesebben (Vermeki, 2023c, megjelenés alatt).

**22. táblázat** A TUD, NAGYON, ILYEN és az AKKOR lemmák gyakorisága a többemű lexikai egységekben

| Sz. | Lemma  | Gyakoriság az N-gramokban |
|-----|--------|---------------------------|
| 13. | TUD    | 591                       |
| 19. | NAGYON | 406                       |
| 14. | ILYEN  | 266                       |
| 15. | AKKOR  | 266                       |

A harmadik szakaszban (20–100. hely) több olyan elem is található, amely a referenciakorpusz első száz lemmája között egyáltalán nem jelenik meg (23. táblázat). Ilyen például a dokumentumok 50%-ban szereplő IGAZÁBÓL, amelynek használatára a későbbiekben még szó kerül (IV.4.6 fejezet). Ennél a szakasznál is megjelennek a beszéltnyelvi fordulatok, mint az UGYE, TÖK, ÚGYHOGY, TÉNYLEG, NA, SZÓVAL, valamint az AMÚGY (Vermeki, 2023c, megjelenés alatt).

**23. táblázat** A KorSzak GyermeKNyelvi Korpusz első száz lemmája között (21–100. hely) szereplő, de a huTenTen12 első száz lemmája között nem megjelenő itemek

| Sz.  | Lemma    | Gyak. | Gyakoriság<br>/ millió | DOKgy | Relatív<br>DOKgy % |
|------|----------|-------|------------------------|-------|--------------------|
| 39.  | UGYE     | 231   | 3228,69                | 41    | 52,56              |
| 40.  | SZOKIK   | 226   | 3158,81                | 44    | 56,41              |
| 49.  | IGAZÁBÓL | 170   | 2376,09                | 39    | 50,00              |
| 59.  | INKÁBB   | 138   | 1928,83                | 33    | 42,31              |
| 80.  | TÖK      | 88    | 1229,98                | 19    | 24,36              |
| 84.  | ÚGYHOGY  | 81    | 1132,14                | 26    | 33,33              |
| 87.  | TÉNYLEG  | 78    | 1090,21                | 25    | 32,05              |
| 90.  | GONDOL   | 76    | 1062,25                | 29    | 37,18              |
| 91.  | EGYSZER  | 73    | 1020,32                | 32    | 41,03              |
| 92.  | NA       | 73    | 1020,32                | 26    | 33,33              |
| 93.  | HISZ     | 73    | 1020,32                | 24    | 30,77              |
| 94.  | SZÓVAL   | 73    | 1020,32                | 23    | 29,49              |
| 95.  | AMÚGY    | 72    | 1006,35                | 21    | 26,92              |
| 99.  | AZTÁN    | 69    | 964,41                 | 24    | 30,77              |
| 100. | BESZÉL   | 68    | 950,44                 | 17    | 21,79              |

A harmadik szakaszban (21–100. hely) lévő és a referenciakorpusz első száz leggyakoribb lemmája között nem szereplő elemek a fenti táblázatban (23. táblázat) láthatóak. Ezeket, valamint a nagyobb gyakorisági mutatóval rendelkező elemek használatát érdemes ismét a többemű lexikai egységekkel is összevetni. A fenti táblázatban lévő elemek közül a kritériumok alapján a fókuszkorpuszból kiválasztott többemű lexikai egységekben nem szerepel az UGYE, INKÁBB, ÚGYHOGY, TÉNYLEG, NA, HISZ, SZÓVAL, AMÚGY, AZTÁN, valamint a BESZÉL lemma. Ezt azt jelenti, hogy a korpuszban bár megjelenik a többeműlexikai egységek részeként, de nem szignifikáns számban, illetve többségük önmagában alkot egy-egy formulaszerű egységet, amely főként diskurzusjelölőként funkcionál (Vermekei, 2023c,

megjelenés alatt). A kutatás feltételrendszere alapján a felsoroltak közül az *ugye* felel meg annak a kritériumnak, hogy egyelemű lexémaként a dokumentumok legalább 50%-ában szerepel. Ebből kifolyólag ennek használata egy későbbi részben (IV.4.1 fejezet) részletesebben is bemutatásra kerül.

A többi lemma – SZOKIK, IGAZÁBÓL, TÖK, GONDOL, EGYSZER – mind egy-egy, gyakori többelemű lexikai egység részét is képezik (24. táblázat).

**24. táblázat** A TÖK, GONDOL, EGYSZER, IGAZÁBÓL és a SZOKIK lemmák gyakorisága a többelemű lexikai egységekben

| Sz. | Lemma    | Gyakoriság az N-gramokban |
|-----|----------|---------------------------|
| 42. | TÖK      | 11                        |
| 44. | GONDOL   | 10                        |
| 50. | EGYSZER  | 8                         |
| 56. | IGAZÁBÓL | 6                         |
| 55. | SZOKIK   | 6                         |

A harmadik szakaszban mindkét korpuszban megtalálható, de eltéréseket mutató lemmák közül a SZERET (79%-kal magasabb), az OLYAN (63%-kal magasabb) az ÍGY (76%-kal magasabb), a MOND (76%-kal magasabb), a SZERINT (63%-kal magasabb), a MEGY (72%-kal magasabb) lemmák, amelyek mind nagyobb gyakorisági mutatóval rendelkeznek, mint a referenciakorpuszban lévő párjuk (25. táblázat) (Vermeki, 2023c, megjelenés alatt).

**25. táblázat** A szignifikánsan nagyobb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező itemek a 21–100. leggyakoribb lemma közül

| Lemma  | KorSzak               |                     | huTenTen2012 |                     | Eltérés (%) |
|--------|-----------------------|---------------------|--------------|---------------------|-------------|
|        | Gyermeknyelvi Korpusz |                     |              |                     |             |
|        | Sz.                   | Gyakoriság / millió | Sz.          | Gyakoriság / millió |             |
| SZERET | 27.                   | 4626,39             | 68.          | 995,92              | <b>79</b>   |

| Lemma   | KorSzak       |                     | huTenTen2012 |                     | Eltérés (%) |
|---------|---------------|---------------------|--------------|---------------------|-------------|
|         | Gyermeknyelvi |                     |              |                     |             |
|         | Korpusz       |                     |              |                     |             |
|         | Sz.           | Gyakoriság / millió | Sz.          | Gyakoriság / millió |             |
| ÍGY     | 23.           | 6862,72             | 38.          | 1767,44             | <b>76</b>   |
| MOND    | 24.           | 6639,09             | 39.          | 1682,10             | <b>76</b>   |
| MEGY    | 41.           | 3033,01             | 77.          | 866,55              | <b>72</b>   |
| OLYAN   | 28.           | 4570,49             | 37.          | 1787,94             | <b>63</b>   |
| SZERINT | 29.           | 4318,90             | 40.          | 1652,95             | <b>63</b>   |

Ezek közül a MEGY kivételével mind részét képezik a fókuszkorpuszban lévő többelemű lexikai egységeknek is (26. táblázat).

**26. táblázat** Az ÍGY, SZERET, OLYAN, SZERINT és a MOND lemmák gyakorisága a többelemű lexikai egységekben

| Sz. | Lemma   | Gyakoriság az N-gramokban |
|-----|---------|---------------------------|
| 20. | ÍGY     | 152                       |
| 22. | SZERET  | 144                       |
| 25. | OLYAN   | 89                        |
| 28. | SZERINT | 62                        |
| 33. | MOND    | 29                        |

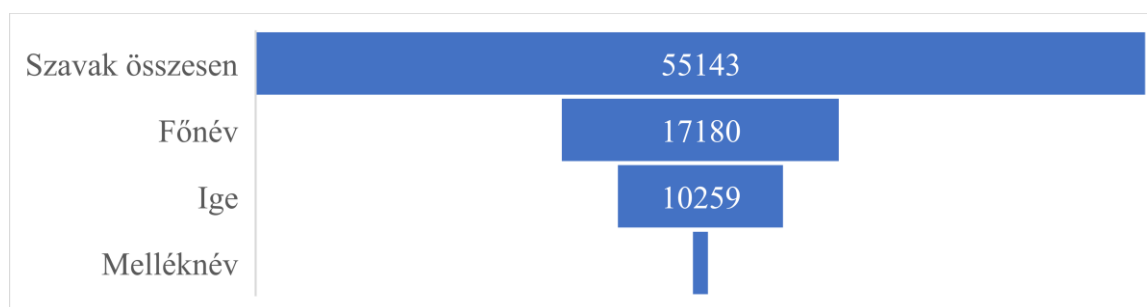
A fent közölt eredmények közül, azokra a lemmákra, amelyek a gyakoriságuk alapján kiválasztott többelemű lexikai egységek részét képezik még a IV.3 fejezetben kitérünk. A többi, gyakorisági számaikban eltéréseket mutató lemma közül (IGEN, UGYE, MEGY) egy kerül kiemelésre. Az *ugye* használatáról pedig – ahogy már említettük – a IV.4.1 fejezetben lesz szó bővebben. Mielőtt erre rátérnénk, megvizsgáljuk a szógyakorisági listákat egy másfajta megközelítésből is, amely a tananyagkészítés szempontjából jelentős szereppel bír.

## 2.2 Szófajonkénti (főnév, melléknév, ige) eredmények

2. kutatási kérdés: Milyen a leggyakoribb egyelemű lexémák szófaji (főnév, melléknév, ige) eloszlása a KorSzak Gyermekeyelvi Korpuszban?

A gyakorisági vizsgálatok szófajonkénti eredményei röviden kerülnek bemutatásra, kizárólag a legfontosabb jelenségekre koncentrálna a figyelmet. A kutatás során kizárólag három szófaj: a főnevek, a melléknévek és az igék gyakoriságának vizsgálatára került sor a kutatás keretei miatt. A KorSzak Gyermekeyelvi Korpuszban összesen 55.143 szó található, amelyből 17.180 főnév, 10.259 ige és 933 melléknév.

70. ábra A szavak szófajonkénti előfordulása a KorSzak Gyermekeyelvi Korpuszban (2020. október)



A főnevek tekintetében (27. táblázat) a legnagyobb gyakorisági mutatóval rendelkező lemmák a beszélgetések témaköreit reprezentálják (például a KUTYA, ÁLLAT, SOROZAT, SPORT, FILM, BARÁT, CSALÁD, VIDEÓ, HOBBI, vagy egy kutya neve: BUBU). Ezeket kizárva, valamint az első ötven elemet áttekintve az ÉSZ és az ÉRZÉS lemmák, amelyek egyáltalán nem szerepelnek a huTenTen12 első száz találata között, valamint a némileg eltérő gyakorisági mutatókkal rendelkező RÉSZ, BAJ, KÉRDÉS lemmák. A felsorolt lemmák közül az ÉSZ, ÉRZÉS és RÉSZ nem képezi részét egyik a fókuszkorpuszban található többelemű lexikai egységnek sem és önmagukban egyelemű lexémaként nem jelennek meg a dokumentumok felében, ezért a jelen kutatásban nem kerül bemutatásra a használatuk. Ezzel szemben a BAJ és a KÉRDÉS lemmák kapcsolódnak többelemű lexikai egységekhez, ezért a későbbiekben (IV.3 fejezet) még röviden kitérünk rájuk (Vermekei, 2023c, megjelenés alatt).

27. táblázat Az 50 leggyakoribb főnév a KorSzak Gyermeknyelvi Korpuszban

| Sz. | Lemma     | Gyak. | Sz. | Lemma    | Gyak. |
|-----|-----------|-------|-----|----------|-------|
| 1.  | KUTYA     | 266   | 26. | APA      | 33    |
| 2.  | ÁLLAT     | 123   | 27. | AJÁNDÉK  | 32    |
| 3.  | DOLOG     | 112   | 28. | KARAKTER | 32    |
| 4.  | SOROZAT   | 109   | 29. | NÉV      | 32    |
| 5.  | IDŐ       | 91    | 30. | TANÁR    | 30    |
| 6.  | NAP       | 78    | 31. | ÉLET     | 29    |
| 7.  | RÉSZ      | 67    | 32. | TÖRTÉNET | 29    |
| 8.  | MACSKA    | 60    | 33. | KÉZ      | 27    |
| 9.  | SPORT     | 57    | 34. | JÁTÉK    | 26    |
| 10. | FILM      | 52    | 35. | ÓRA      | 26    |
| 11. | KEDVENC   | 51    | 36. | LÁNY     | 26    |
| 12. | ANYA      | 49    | 37. | ERDŐ     | 26    |
| 13. | BARÁT     | 49    | 38. | GYEREK   | 25    |
| 14. | CSALÁD    | 48    | 39. | NÉNI     | 25    |
| 15. | HARRY     | 44    | 40. | VÍZ      | 25    |
| 16. | VIDEÓ     | 42    | 41. | OSZTÁLY  | 25    |
| 17. | BAJ       | 42    | 42. | POTTER   | 25    |
| 18. | HELY      | 41    | 43. | HÁZ      | 25    |
| 19. | ÉV        | 39    | 44. | ANYUKA   | 24    |
| 20. | HÁZIÁLLAT | 38    | 45. | ÉRZÉS    | 24    |
| 21. | ISKOLA    | 36    | 46. | ÚT       | 23    |
| 22. | KÉRDÉS    | 36    | 47. | FEJ      | 21    |
| 23. | KÖNYV     | 34    | 48. | PATAK    | 21    |

| Sz. | Lemma | Gyak. | Sz. | Lemma | Gyak. |
|-----|-------|-------|-----|-------|-------|
| 24. | HOBBI | 33    | 49. | FIÚ   | 21    |
| 25. | ÉSZ   | 33    | 50. | BUBU  | 21    |

A melléknevek közül (28. táblázat) a KorSzak Gyermeknyelvi Korpusz leggyakrabban megjelenő lemmája a JÓ, amely a referenciakorpuszban a harmadik helyen szerepel. Ez a melléknév számos többemű lexikai egység részét képezi, ezért még a későbbiekben (IV.3 fejezet) lesz róla szó. A melléknevek gyakorisági listáiban található a legnagyobb különbség a fókusz és a referenciakorpusz között. Az adatközlő gyermekek által használt melléknevek nagy része nem szerepel a huTenTen12 első száz lemmája között (Vermekei, 2023c, megjelenés alatt). Ilyenek az

ARANYOS, BOLDOG, BÜDÖS, CUKI, CSODÁLATOS, CSODÁS, CSÚNYA, FURA, HÍRES, IDEÁLIS, IDŐS, IZGALMAS, KEDVENC, KREATÍV, MENŐ, NORMÁLIS, PICI, SZERENCSETLEN, SZÖRNYŰ, VICCES.

Ezenkívül a huTenTen12 korpuszban szereplő elemek közül a fókuszkorpuszban lévők eltérő, legtöbbször sokkal alacsonyabb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkeznek.

28. táblázat Az 50 leggyakoribb melléknév a KorSzak Gyermeknyelvi Korpuszban

| Sz. | Lemma   | Gyak. | Sz. | Lemma      | Gyak. |
|-----|---------|-------|-----|------------|-------|
| 1.  | JÓ      | 239   | 26. | KÜLÖNBÖZŐ  | 14    |
| 2.  | NAGY    | 111   | 27. | CSODÁLATOS | 13    |
| 3.  | KICSI   | 95    | 28. | FEKETE     | 13    |
| 4.  | KIS     | 82    | 29. | IDŐS       | 13    |
| 5.  | KEDVENC | 69    | 30. | CUKI       | 13    |
| 6.  | BIZTOS  | 45    | 31. | IDEÁLIS    | 13    |
| 7.  | MAGYAR  | 41    | 32. | KEDVES     | 12    |
| 8.  | ÉRDEKES | 39    | 33. | KÖNNYŰ     | 12    |



| Sz. | Lemma    | Gyak. | Sz. | Lemma          | Gyak. |
|-----|----------|-------|-----|----------------|-------|
| 9.  | RENGETEG | 38    | 34. | FEHÉR          | 11    |
| 10. | VICCES   | 36    | 35. | IZGALMAS       | 11    |
| 11. | ARANYOS  | 32    | 36. | BOLDOG         | 11    |
| 12. | ROSSZ    | 30    | 37. | NÉMET          | 11    |
| 13. | KREATÍV  | 29    | 38. | FURA           | 10    |
| 14. | ANGOL    | 26    | 39. | CSODÁS         | 10    |
| 15. | HOSSZÚ   | 24    | 40. | NORMÁLIS       | 10    |
| 16. | SZÉP     | 23    | 41. | BÜDÖS          | 10    |
| 17. | NEHÉZ    | 23    | 42. | IGAZI          | 9     |
| 18. | ÚJ       | 22    | 43. | HASONLÓ        | 9     |
| 19. | FONTOS   | 20    | 44. | CSÚNYA         | 9     |
| 20. | HÍRES    | 20    | 45. | SZÖRNYŰ        | 9     |
| 21. | PICI     | 20    | 46. | UTOLSÓ         | 9     |
| 22. | MENŐ     | 20    | 47. | BARNA          | 9     |
| 23. | KEVÉS    | 19    | 48. | RÉGI           | 9     |
| 24. | DRÁGA    | 17    | 49. | MAI            | 9     |
| 25. | NYÁRI    | 15    | 50. | SZERENCSÉTLLEN | 8     |

Ilyenek például az ÚJ vagy az UTOLSÓ lemmák (29. táblázat), amelyek a szövegekben szignifikánsan ritkábban fordulnak elő. Az ÚJ lemmára a Korszak Gyermeeknyelvi Korpuszban csaknem 80%-kal kevesebb előfordulást találtam, míg az UTOLSÓ esetében 60%-kal kevesebb előfordulást jegyeztem fel.

**29. táblázat** A szignifikánsan alacsonyabb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező melléknevek

| Lemma  | KorSzak       |                     | huTenTen2012 |                     | Eltérés (%) |
|--------|---------------|---------------------|--------------|---------------------|-------------|
|        | Gyermeknyelvi |                     |              |                     |             |
|        | Sz.           | Gyakoriság / millió | Sz.          | Gyakoriság / millió |             |
| ÚJ     | 28.           | 307,49              | 4.           | 1434,89             | <b>79</b>   |
| UTOLSÓ | 72.           | 125,79              | 29.          | 307,05              | <b>60</b>   |

Az igék kategóriájában (30. táblázat) a főnevekhez hasonlóan – de nem olyan nagy mértékben – megjelennek azok a lemmák, amelyek a témakörökhöz szorosan kapcsolódnak. Ilyenek például a szabadidős tevékenységekhez köthetőek közül a SPORTOL, RAJZOL, SÉTÁL, ÉLVEZ, amelyek nem szerepelnek a referenciakorpusz leggyakoribb száz igéje között, a tananyagkészítés szempontjából, a témakörökhöz tartozó lexika kiválasztásához viszont rendkívüli hasznosak. A referenciakorpuszban egyáltalán nem jelennek meg olyan igekötős igék, mint az ELMEGY, ELKEZD, MEGNÉZ vagy a KIMEGY, amelyet a gyermek adatközlők számos alkalommal használnak a párbeszéd során, valamint nincs benne a referenciakorpusz 100 leggyakoribb igéje között a fókuszkorpusz 41. helyén álló ESZIK sem (Vermeki, 2023c, megjelenés alatt).

**30. táblázat** Az 50 leggyakoribb ige a KorSzak Gyermeknyelvi Korpuszban

| Sz. | Lemma  | Gyak. | Sz. | Lemma   | Gyak. |
|-----|--------|-------|-----|---------|-------|
| 1.  | VAN    | 2,309 | 26. | ISMER   | 46    |
| 2.  | TUD    | 650   | 27. | ÉL      | 42    |
| 3.  | MOND   | 474   | 28. | VESZ    | 42    |
| 4.  | SZERET | 331   | 29. | JÁTSZIK | 41    |
| 5.  | SZOKIK | 225   | 30. | TANUL   | 41    |
| 6.  | MEGY   | 217   | 31. | OLVAS   | 41    |

| Sz. | Lemma     | Gyak. | Sz. | Lemma    | Gyak. |
|-----|-----------|-------|-----|----------|-------|
| 7.  | NÉZ       | 176   | 32. | JUT      | 40    |
| 8.  | LESZ      | 152   | 33. | TÖRTÉNIK | 39    |
| 9.  | CSINÁL    | 140   | 34. | AD       | 38    |
| 10. | LÁT       | 112   | 35. | TETSZIK  | 38    |
| 11. | JÖN       | 99    | 36. | MEGNÉZ   | 36    |
| 12. | NINCS     | 92    | 37. | SÉTÁL    | 36    |
| 13. | FOG       | 79    | 38. | TALÁL    | 32    |
| 14. | AKAR      | 78    | 39. | MESÉL    | 32    |
| 15. | GONDOL    | 76    | 40. | ÉREZ     | 31    |
| 16. | ELMEGY    | 71    | 41. | ESZIK    | 30    |
| 17. | HISZ      | 70    | 42. | KIMEGY   | 30    |
| 18. | BESZÉL    | 68    | 43. | KEZD     | 30    |
| 19. | EMLÉKSZIK | 62    | 44. | SPORTOL  | 29    |
| 20. | KAP       | 56    | 45. | RAJZOL   | 29    |
| 21. | ELKEZD    | 54    | 46. | FÜGG     | 29    |
| 22. | TART      | 51    | 47. | ÉLVEZ    | 28    |
| 23. | JÁR       | 51    | 48. | ÁLL      | 26    |
| 24. | ÉRT       | 48    | 49. | PRÓBÁL   | 25    |
| 25. | TESZ      | 47    | 50. | SZÓL     | 25    |

Erre a kategóriára a melléknevekkel ellentétben inkább az jellemző, hogy azok az igék, amelyek mindkét korpuszban megjelennek, és nem azonos helyen állnak a szógyakorisági listában, a fókuszkorpuszban nagyobb gyakorisági mutatóval rendelkeznek (31. táblázat). Ilyen a már az előző fejezetben említett MEGY (72%-kal magasabb), a CSINÁL (81%-kal magasabb), valamint az EMLÉKSZIK (78%-kal magasabb) igék (Vermekei, 2023c, megjelenés alatt).

**31. táblázat** A szignifikánsan magasabb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező igék

| Lemma     | KorSzak       |                     | huTenTen2012 |                     | Eltérés (%) |
|-----------|---------------|---------------------|--------------|---------------------|-------------|
|           | Gyermeknyelvi |                     |              |                     |             |
|           | Korpusz       |                     |              |                     |             |
|           | Sz.           | Gyakoriság / millió | Sz.          | Gyakoriság / millió |             |
| CSINÁL    | 10.           | 1956,78             | 40.          | 376,89              | <b>81</b>   |
| EMLÉKSZIK | 20.           | 866,58              | 85.          | 196,61              | <b>78</b>   |
| MEGY      | 6.            | 3033,01             | 13.          | 861,26              | <b>72</b>   |

Bár a jelen kutatás nem tudott teljes körű, részletes képet adni a KorSzak Gyermeknyelvi Korpuszban megjelenő szófajokról, érdemes lenne további vizsgálatokat végezni ezzel az aspektussal kapcsolatosan is. A gyermekek nyelvhasználatában kimutatható eltérések részletesebb elemzése, illetve a szófajok bővebb vizsgálata további segítséget nyújthat a tananyagkészítés során.

### 3. A két- vagy többelemű lexikai egységek

3. kutatási kérdés: Melyek a leggyakoribb két- vagy többelemű lexikai egységek a KorSzak Gyermeknyelvi Korpuszban?

A két vagy több elemből álló lexikai egységek az egyelemű lexémákhoz hasonlóan szintén gyakorisági mutatójuk alapján kerültek kigyűjtésre a Sketch Engine korábban bemutatott N-gram eszközének segítségével. Mielőtt az így kapott lista véglegesült, konkordanciasorokban is megvizsgáltam a kapott eredményeket (Vermeki, 2022b, megjelenés alatt). Erre azért volt szükség, mert a magyar nyelvre jellemző szórend miatt a találatok között előfordulhatnak olyan kifejezések, amelyeket a korpuszeszköz második keresésével találhatunk csak meg, mert például az N-gram eszköz külön csoportokba sorolta őket. Íme egy példa erre a problémára:

(3a) **ezzel** én is így vagyok

(3b) én is így vagyok **ezzel**

(3c) én is így vagyok **vele**

(KorSzak Gyermeknyelvi Korpusz, 2020)

Nézzük meg a fenti példákat (3a–3c) tágabb szövegekörnyezetben, hogy jobban megértsük a használatuk jellemzőit.

(doc#2 Mon\_Állatok\_003: A megnyilatkozó gyermekek az állatkísérletek kegyetlenségeiről beszélnek.)

(4) A: Igen, de hogyha beadnak nekik ilyen-olyan szereket, mit tudom én, és attól valami bajuk lesz, vagy fájdalmaik lesznek, az szerintem nagyon embertelen, vagy hogy mondjam.

B: Igen, **ezzel én is így vagyok**, hogyha már egyszer valamin ki kell próbálni, akkor muszáj, de annyira belegondolok, hogy ők mit élhetnek ilyenkor át [...].

(doc#54 Mon\_Kirándulás\_001: A gyermekek arról beszélnek, hogy kivel – barátok vagy család – szeretnek kirándulni menni.)

(5) A: [...] családdal, ismerősökkel, barátokkal, szóval úgymond bárkivel, akivel jól érzem magam, igen.

B: **Én is így vagyok ezzel** [...].

(doc#73 Mon\_Utazás\_001: A gyermekek különböző úticélokról beszélgetnek.)

(6) A: De igazából, hogy ha mondjuk van egy ilyen érdekes város, meg jó történetek fűződnek hozzá, akkor az is érdekel igazából, de mondjuk az ilyen természeti látnivalók inkább jobban vonzanak.

B: **Én is így vagyok vele** [...].

(KorSzak Gyermekegyetemesi Korpusz, 2020)

A fenti példákban (4–6. példa) található *ezzel* névmás előfordulhat a szekvencia elején (4. példa) vagy a végén (5. példa), amíg a másik, hasonló jelentésű *vele* névmásunk, amely önmagában személyre is utalhat, csak a szekvencia végén jelenik meg (6. példa). Ezért elengedhetetlen a keresési eredmények részletes vizsgálata és egy második, esteleg rövidített kifejezéssel történő konkordanciasor készítése. Az így kapott eredmények összesítésével, valamint az esetleges duplikációk kiszűrésével, pontosabb eredményeket kaphatunk (Vermekei, 2022b, megjelenés alatt).

### 71. ábra Példa az én is így vagyok ezzel/vele rövidített formulaszerű szekvencia konkordanciasorára

| Details  | Left context                                     | KWIC  | Right context                                   |
|----------|--|---|---|
| ① doc#2  | ambertelen, vagy hogy mondjam. - Igen, ezzel     | <b>én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Adv_Pro V:Prs:NDef:1Sg | , hogyha már egyszer valamin ki kell próbálni,  |
| ① doc#8  | kitalálni, mert nem ismered a személyiséget. -   | <b>Én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Cnj V:Prs:NDef:1Sg     | vele.</s><s>És milyen színű golden retriever l  |
| ① doc#54 | nond bárkivel, akivel jól érzem magam, igen. -   | <b>Én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Adv_Pro V:Prs:NDef:1Sg | ezzel, mint az Amira, hogy én is bárkivel kimeg |
| ① doc#54 | ik, akkor csak egyedül is szeretek sétálgatni. - | <b>Én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Adv_Pro V:Prs:NDef:1Sg | ezzel a dologgal, hogyha családdal vagyunk, a   |
| ① doc#65 | t se nézem, én inkább az angol sorozatokat. -    | <b>Én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Cnj V:Prs:NDef:1Sg     | vele, mint az Eszter.</s><s>Igazából nem nag    |
| ① doc#66 | tudom abbahagyni, mivel izgalmas. - Hü ezzel     | <b>én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Adv_Pro V:Prs:NDef:1Sg | .</s><s>Én függővé válok és ez nagyon durva     |
| ① doc#73 | természeti látnivalók inkább jobban vonzanak. -  | <b>Én is így vagyok</b><br>N_Pro:1Sg:Nom Adv Cnj V:Prs:NDef:1Sg     | vele, de például, amikor mentünk Firenzébe, a   |

A fenti példák (3a–3c), amelyek mindegyike hasonló jelentéssel bír, jól mutatják a magyar szórenddel történő korpusznyelvészeti elemzések kihívásait. Miközben számunkra egyértelműen azonos vagy közel hasonló szerkezeteket használtak a megnyilatkozást tévő gyermekek egyetértésüket kifejezve egy adott témával kapcsolatban, a keresőeszköz az első keresés során az *ezzel* névmás különböző helyzetei, illetve a *vele* névmás használata miatt, három különálló n-gramként kezelte ezeket a példákat (32. táblázat). Két változatát pedig nem is találta meg az eszköz:

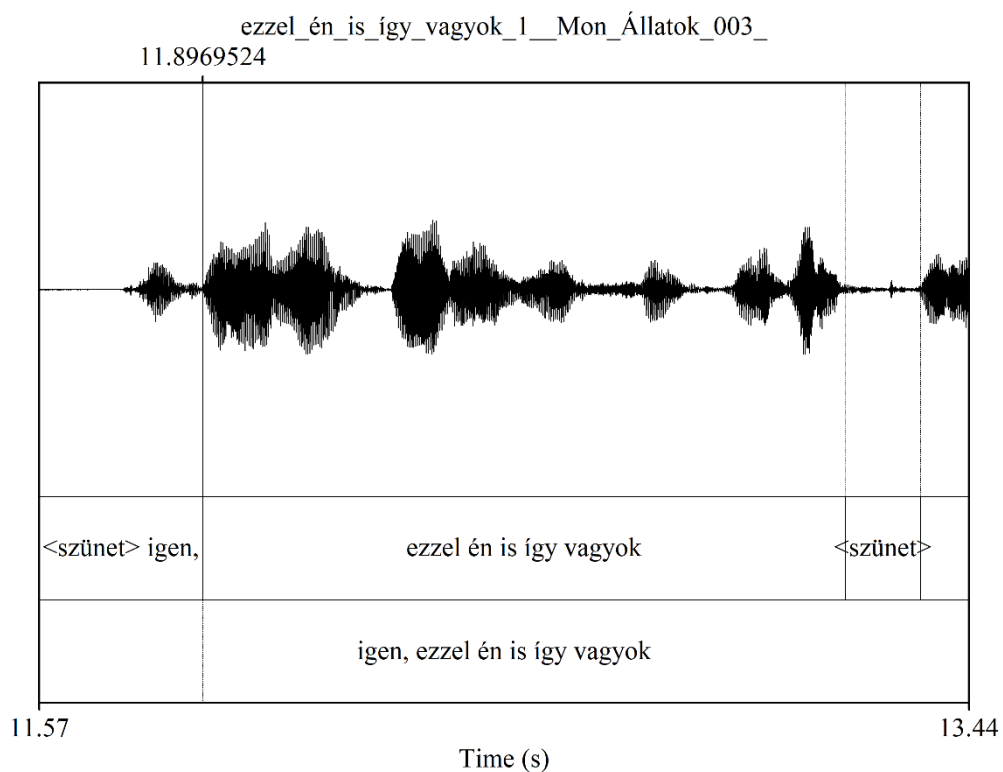
- *én is pontosan így vagyok vele* (doc#50 Mon\_Hp\_001),
- *huu, ezzel én is nagyon így vagyok* (doc#66 Mon\_Sorozatok\_002).

**32. táblázat** Az *én is így vagyok* + *vele/ezzel* előfordulásai a KorSzak Gyermekegyelvi Korpuszban

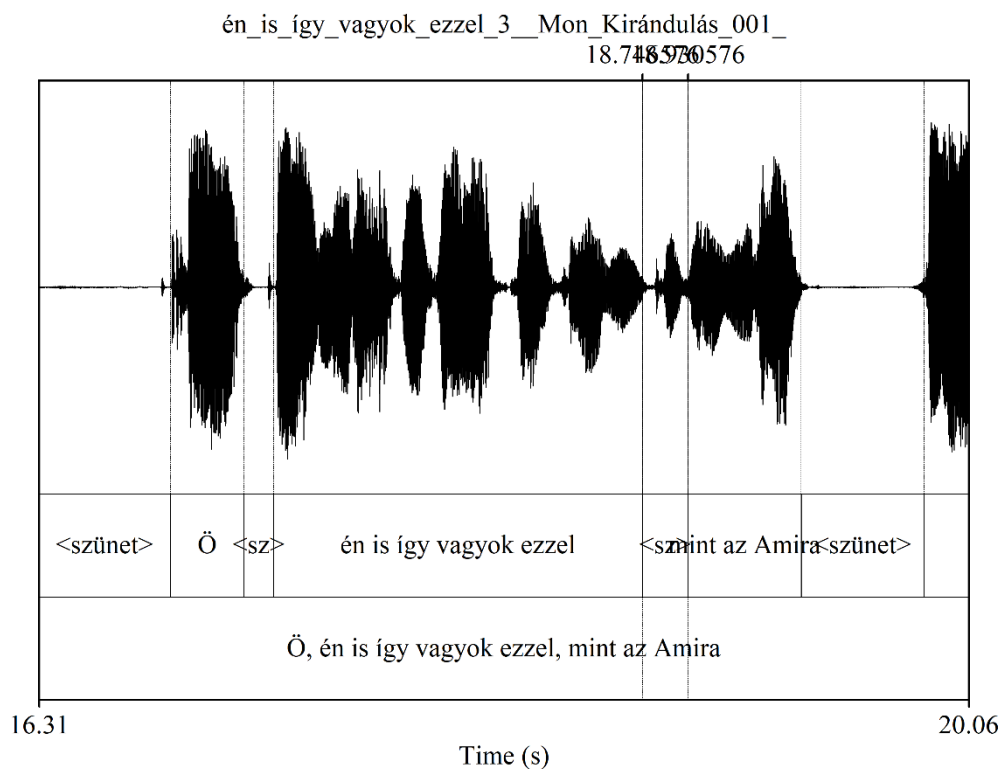
| Sz. | N-gram (szó)           | Gyakoriság | Gyakoriság / millió szó | Relatív DOKgy % |
|-----|------------------------|------------|-------------------------|-----------------|
| 1   | én is így vagyok vele  | 3          | 41,93                   | 3,85            |
| 7   | ezzel én is így vagyok | 2          | 27,95                   | 2,56            |
| 23  | én is így vagyok ezzel | 2          | 27,95                   | 1,28            |

A konkordanciasorokban történő vizsgálat alapján viszont megállapíthatjuk, hogy a Sketch Engine által három külön kategóriába osztályozott öt elemből álló lexikai egység jelentése és használata miatt inkább egyetlen kategóriába sorolandó. Úgy tűnik, hogy egy részben rögzült formulaszerű lexikai elemként értelmezhető. Azonban még felmerülhet a kérdés, hogy valóban erről van-e szó vagy inkább kisebb méretű teljesen rögzült elemek csoportjairól. Például egy négyelemű *én is így vagyok*, vagy két darab kételemű lexikai egységet (*én is* és *így vagyok*) találtunk. Az ilyen kérdéses esetekben a lexikai egységek kiválasztásához a találatokat a Praat programmal is megvizsgáltam, hogy megállapítsam, az adatközlő gyermekek egyben, szünet nélkül mondják-e ki mind az öt elemet a diskurzus közben. A következő ábrákon (72–74. ábra) a fentebb szöveggörnyezetben is bemutatott három találat eredményei láthatóak a Praat segítségével vizualizálva (Vermekei, 2022b, megjelenés alatt).

72. ábra Szünetek helye az *ezzel én is így vagyok* kimondása közben (doc#2: Mon\_Állatok\_003)

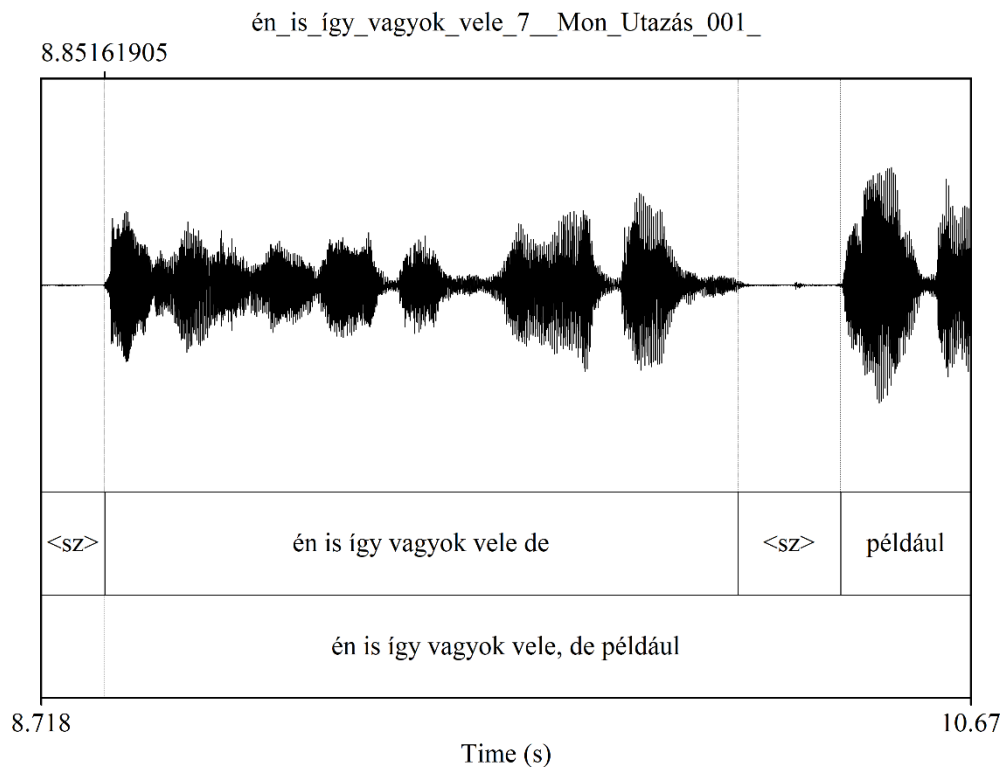


73. ábra Szünetek helye az *én is így vagyok ezzel* kimondása közben (doc#54: Mon\_Kirándulás\_001)





**74. ábra** Szünetek helye az *én is így vagyok veled* kimondása közben (doc#73: Mon\_Utazás\_001)



A Praat által annotált csendes [silent] és hangzó [sounding] szakaszokat átneveztem, az ábrákon. A megnyilatkozások mellett a szünetek helyét a <szünet> vagy <sz> rövidítéssel jelöltem az első sorban, a második sorban pedig az ábrán látható teljes megnyilatkozás olvasható. Az első két ábrán (72–73. ábra) látható eredmények alapján – bár a második ábrán látható, hogy enyhén zajos a felvétel – megállapíthatjuk, hogy az egységeket egyben, szünet nélkül mondták ki a gyermekek a beszédprodukciónk során. Az utolsó ábra (74. ábra) viszont azt mutatja meg, hogy a szünetet a megnyilatkozó egy lexémával később tartja, ezért érdemes megnézni az összes találat eredményeit. Az alábbi táblázatban (33. táblázat) ezt foglaltam össze, az eredmények vizualizációi pedig a mellékletek között találhatóak.

**33. táblázat** Az *én is így vagyok + veled/ezzel* közben tartott szünetek összefoglaló táblázata

| Dok. | N-gram (szó)           | Szünet helye<br>a lexikai egység<br>előtt / után van |     | Megjegyzés            |
|------|------------------------|--|-----|-----------------------|
|      |                        | igen   | nem |                       |
| #2   | ezzel én is így vagyok | x  |     | Jól mérhető a szünet. |
| #8   | én is így vagyok veled | x  |     | Jól mérhető a szünet. |

| Dok. | N-gram (szó)                   | Szünet helye<br>a lexikai egység<br>előtt / után van |     | Megjegyzés  |
|------|--------------------------------|--|-----|---|
|      |                                | igen   | nem |   |
| #54  | én is így vagyok ezzel         | x  |     | Enyhén zajos a felvétel, a szünetnél jelzi az ábra, de mérhető a szünet.  |
| #54  | én is így vagyok ezzel         |  | x   | A szünet mintája eltér:<br><sz> én is <sz> így vagyok ezzel a dologgal <sz> hogyha                                      |
| #65  | én is így vagyok vele          | x  |     | Többelemű egység, a szünet mintája:<br><sz> én is így vagyok vele mint az <sz>  |
| #66  | ezzel én is így vagyok         |  |     | Hallhatóan az ötelemű egység előtt és után van a szünet, de a gyermekek egyszerre beszélnek, így nem mérhető.           |
| #73  | én is így vagyok vele          | x  |     | Többelemű egység, a szünet mintája:<br><sz> én is így vagyok vele de <sz>   |
| #66  | ezzel én is nagyon így vagyok  | x  |     | Többelemű egység, a szünet mintája:<br><sz> huu, ezzel én is nagyon így vagyok és néha már kiborít, hogy nem bírok <sz> |
| #50  | én is pontosan így vagyok vele |  | x   | A szünet mintája eltér:<br><sz> én is <sz> pontosan így vagyok vele <sz>  |

A kilenc találat közül háromban jól mérhető a szünet az öt elemből álló egység előtt és után, háromban több mint öt elem után következik a szünet, egy felvételen nem mérhető a szünet, mert a gyermekek egyszerre beszélnek, kettőben pedig máshol tartja a beszélő a szünetet. Bár ebben az esetben nem hozott szignifikáns eredményt a szünetek helyének vizsgálata, hiszen a találatok 33%-ban valósult csak meg teljes mértékben a szünet helyével

kapcsolatos kritérium, mégis jól mutatja, hogy miért van szükség a többféle megközelítésre, illetve a több kutatási szekvenciára.

Az N-gram eszköz alkalmazásával talált többi elem esetében, amelyeknél teljes rögzítettségük miatt nem merültek fel hasonló kérdések, a már korábban említett kritériumok (34. táblázat) alapján választottam ki a végleges listán lévő elemeket. Majd ezeket kategorizáltam szövegben betöltött szerepük szerint (Biber, 2006). A következő alfejezetekben a KorSzak Gyermeeknyelvi Korpuszból ily módon kiválasztott két- vagy többelemből álló lexikai egységeknek a közlésére kerül sor.

**34. táblázat** A lexikai egységek kiválasztásának meghatározott kritériumai

| Lexikai egység típusa      | Minimum gyakorisági mutató     |
|----------------------------|--------------------------------|
| kételemű egység [2-gram]   | 100 előfordulás / 1 millió szó |
| háromelemű egység [3-gram] | 14 előfordulás / 1 millió szó  |
| négyelemű egység [4-gram]  | 4 előfordulás / 1 millió szó   |

| Lexikai egység típusa      | Minimum dokumentum gyakoriság | Minimum relatív dokumentum gyakoriság (%) |
|----------------------------|-------------------------------|---|
| egyelemű lexéma            | 39                            | 50  |
| kételemű egység [2-gram]   | 20                            | 25  |
| háromelemű egység [3-gram] | 5                             | 6,4                                       |
| négyelemű egység [4-gram]  | 3                             | 3,8                                       |

### 3.1 A leggyakoribb két- vagy többelemű lexikai egységek

#### 3.1.1 Kételemű lexikai egységek

Az első kategória a kételemű lexikai egységek kategóriája. A Sketch Engine összesen 4749 kételemű egységet talált, amelyből 600 egység maradt az első szűrés után, miután a feltételrendszer alapján kizártam a 100 előfordulás / egymillió szónál kisebb gyakorisági mutatóval rendelkező elemeket. Ezután a második szűréssel a következő kritériumnak megfelelően kizártam a 25%-nál kevesebb relatív dokumentum gyakorisági mutatóval

rendelkező egységeket, hogy a kiválasztott elemek ne csak néhány adatközlő nyelvhasználatát tükrözzék, hanem szélesebb eloszlásban jelenjenek meg a korpuszban. Így a végleges lista 60 itemet tartalmaz. A következő táblázatban (35. táblázat) az első 10 legnagyobb gyakorisági mutatóval rendelkező kételemű lexikai egység látható.

**35. táblázat** A tíz leggyakoribb kételemű lexikai egység a KorSzak GyermeKNyelvi Korpuszban

| Sz. | N-gram      | Gyak. | Gyakoriság /<br>millió | DOKgy | Relatív DOKgy<br>% |
|-----|-------------|-------|------------------------|-------|--------------------|
| 1.  | nem tudom   | 227   | 3172,78                | 38    | 48,72              |
| 2.  | és akkor    | 177   | 2473,93                | 36    | 46,15              |
| 3.  | hogy a / az | 147   | 2 054,62               | 39    | 50,00              |
| 4.  | én is       | 144   | 2012,69                | 32    | 41,03              |
| 5.  | nem is      | 126   | 1761,10                | 32    | 41,03              |
| 6.  | az a        | 107   | 1495,54                | 40    | 51,28              |
| 7.  | egy ilyen   | 106   | 1481,56                | 25    | 32,05              |
| 8.  | az ilyen    | 101   | 1411,68                | 22    | 28,21              |
| 9.  | meg a       | 96    | 1341,79                | 24    | 30,77              |
| 10. | és a        | 94    | 1313,84                | 41    | 52,56              |

### 3.1.2 Háromelemű lexikai egységek

A következő kategória a háromelemű lexikai egységek kategóriája. A Sketch Engine N-gram eszköze az első keresés során 2033 egységet talált, amelyet az előzőekben leírtakhoz hasonlóan és a háromelemű lexikai egységekkel kapcsolatosan korábban meghatározott kritériumoknak (14 előfordulás / millió szó, 6,4 % relatív dokumentum gyakoriság) megfelelően szűrtem meg. Így 120 lexikai egység maradt a listán, ezzel ez vált a legnagyobb kategóriává. A következő táblázatban (36. táblázat) az első 10 legnagyobb gyakorisági mutatóval rendelkező háromelemű lexikai egység látható.

**36. táblázat** A tíz leggyakoribb háromelemű lexikai egység a KorSzak Gyermekegyeségi Korpuszban

| Sz. | N-gram            | Gyak. | Gyakoriság /<br>millió | DOKgy | Relatív<br>DOKgy % |
|-----|-------------------|-------|------------------------|-------|--------------------|
| 1.  | nem is tudom      | 49    | 684,87                 | 16    | 20,51              |
| 2.  | mit tudom én      | 29    | 405,33                 | 13    | 16,67              |
| 3.  | vagy nem tudom    | 20    | 279,54                 | 10    | 12,82              |
| 4.  | vagy hogy mondjam | 18    | 251,59                 | 8     | 10,26              |
| 5.  | ott van a         | 12    | 167,72                 | 6     | 7,69               |
| 6.  | de nem tudom      | 12    | 167,72                 | 6     | 7,69               |
| 7.  | és akkor így      | 12    | 167,72                 | 10    | 12,82              |
| 8.  | mi a kedvenc      | 12    | 167,72                 | 11    | 14,10              |
| 9.  | hát nem tudom     | 12    | 167,72                 | 9     | 11,54              |
| 10. | van egy ilyen     | 11    | 153,75                 | 9     | 11,54              |

### 3.1.3 Négyelemű lexikai egységek

Az utolsó kategória, amely vizsgálatra került a jelen kutatásban a négyelemű lexikai egységek kategóriája, amelyből a Sketch Engine az első keresés után 313 egységet talált, amelyet 20 egységre szűkítettem le a kritériumok alapján (10 előfordulás / millió szó, 3,8% relatív dokumentum gyakoriság). A következő táblázatban (37. táblázat) az első 10 legnagyobb gyakorisági mutatóval rendelkező négyelemű lexikai egység látható.

**37. táblázat** A tíz leggyakoribb négyelemű lexikai egység a KorSzak Gyermekegyeségi Korpuszban

| Sz. | N-gram            | Gyak. | Gyakoriság /<br>millió | DOKgy | Relatív<br>DOKgy % |
|-----|-------------------|-------|------------------------|-------|--------------------|
| 1.  | vagy nem is tudom | 7     | 97,84                  | 5     | 6,41               |
| 2.  | én is így vagyok  | 7     | 97,84                  | 6     | 7,69               |
| 3.  | hogy mit tudom én | 5     | 69,89                  | 4     | 5,13               |

| Sz. | N-gram                         | Gyak. | Gyakoriság /<br>millió | DOKgy | Relatív<br>DOKgy % |
|-----|--------------------------------|-------|------------------------|-------|--------------------|
| 4.  | nagyon jól éreztük<br>magunkat | 4     | 55,91                  | 3     | 3,85               |
| 5.  | én is egyet tudok              | 4     | 55,91                  | 3     | 3,85               |
| 6.  | én úgy vagyok vele             | 4     | 55,91                  | 4     | 5,13               |
| 7.  | vagy mit tudom én              | 4     | 55,91                  | 4     | 5,13               |
| 8.  | az volt az első                | 4     | 55,91                  | 4     | 5,13               |
| 9.  | az nagyon jó volt              | 4     | 55,91                  | 3     | 3,85               |
| 10. | de ő is nagyon                 | 3     | 41,93                  | 3     | 3,85               |

A fenti eredményekből és az összefoglaló táblázatból (38. táblázat) láthatjuk, hogy a meghatározott feltételrendszer alapján kiválasztott lexikai egységek közül a legnagyobb kategória a háromelemű lexikai egységeké (120 item), majd ezt követik a két- (60 item) és végül a négyelemű egységek (20 item). Az összesen 200 darab kiválasztott lexikai egységet a mellékletek között található táblázatokban ismerhetjük meg. A legnagyobb gyakorisági mutatóval rendelkező többelemű lexikai egységek a kételeműek között találhatóak, ilyenek a NEM TUD (292), az ÉN IS (188), valamint az ÉS AKKOR (177). Ugyanakkor már itt is láthatjuk, hogy a kételemű és háromelemű egységek közül több nagy valószínűséggel beágyazódott hosszabb egységekbe. Ilyen lehet a *mit tudom én* > *vagy mit tudom én*, illetve az *így vagyok* > *én is így vagyok velem / ezzel*. Ezeknek a vizsgálatának bemutatása a későbbiekben kerül kifejtésre (IV.5. fejezet).

**38. táblázat** A leggyakoribb kételemű, a háromelemű és négyelemű lexikai egységek összehasonlító táblázata

| Sz.  | Kételemű l.e. | Gyak. | Háromelemű l.e.    | Gyak. | Négyelemű l.e.              | Gyak. |
|------|---------------|-------|--------------------|-------|-----------------------------|-------|
| 1.   | nem tudom     | 227   | nem is tudom       | 49    | vagy nem is tudom           | 7     |
| 2.   | és akkor      | 177   | mit tudom én       | 29    | én is így vagyok            | 7     |
| 3.   | én is         | 144   | vagy nem tudom     | 20    | hogya mit tudom én          | 5     |
| 4.   | hogya a / az  | 147   | vagy hogya mondjam | 18    | nagyon jól éreztük magunkat | 4     |
| 5.   | nem is        | 126   | ott van a          | 12    | én is egyet tudok           | 4     |
| 6.   | az a          | 107   | de nem tudom       | 12    | én úgy vagyok vele          | 4     |
| 7.   | egy ilyen     | 106   | és akkor így       | 12    | vagy mit tudom én           | 4     |
| 8.   | az ilyen      | 101   | mi a kedvenc       | 12    | az volt az első             | 4     |
| 9.   | meg a         | 96    | hát nem tudom      | 12    | az nagyon jó volt           | 4     |
| 10.  | és a          | 94    | van egy ilyen      | 11    | de ő is nagyon              | 3     |
| 11.  | hogya nem     | 92    | és akkor ott       | 11    | azt is nagyon szeretem      | 3     |
| 12.. | ez a          | 85    | egy idő után       | 10    | hogya mi a kedvenc          | 3     |
| 13.  | van egy       | 68    | vagy nem is        | 10    | nektek mi a kedvenc         | 3     |
| 14.  | és hogya      | 60    | én nem tudom       | 10    | az lenne a kérdésem         | 3     |
| 15.  | az egyik      | 57    | az nagyon jó       | 9     | nekem nem nagyon van        | 3     |

## 4. A KorSzak Gyermeknyelvi Korpuszban található gyakori lexikai egységek használati mintázatai

*4. kutatási kérdés: Hogyan használják a KorSzak Gyermeknyelvi Korpuszban szereplő gyermekek a kiválasztott lexikai egységeket? Milyen használati mintázatok figyelhetők meg a korpuszban?*

Az előzőekben bemutatott gyakorisági vizsgálatok során elkészült listákban – ahogy látható volt – mind az egyelemű lexémák, mind a két- vagy többelemű lexikai egységek esetében megfigyelésük lemmatizált formában (vö. Durrant, 2013: 32) történt. A lemmatizálásra azért volt szükség, hogy a keresés során a Sketch Engine az egységek minden alakját, így minden előfordulását is megtalálja. Miután elkészültek a gyakorisági listák, vagyis a feltételrendszer alapján kiválasztásra került a kételemű lexikai egységek esetében a 60, a háromeleműek esetében a 120, valamint a négyeleműek közül 20 item, a következő lépés a kiválasztott egységek használati mintázatainak megfigyelése volt. Ehhez a Sketch Engine Concordance és Word Sketch eszközeit alkalmaztam. Ebben az alfejezetben azoknak a formulaszerű lexikai elemek használati mintázatainak a bemutatására kerül sor, amelyek a gyermekeknek készülő tananyag készítése szempontjából érdekesek lehetnek.

### 4.1 Az *ugye* használati mintázatai

Az első bemutatásra kerülő, a korpuszban gyakori lexikai egység az *ugye*, amelyet azért emeltem ki, mert a használata erősen jellemző a beszélt nyelvre, ezenkívül nincs kialakult tanítási módja a magyar mint idegen nyelv órákon. Már az első gyakorisági vizsgálatok után, a konkordanciasorokban történő elemzés során látszott, hogy a gyermekek a korpuszban többféleképpen is használják ezt a szót. Érdemes ezeket az előfordulásokat bővebb kontextusban is megfigyelni, mert a tananyag készítéséhez segítséget jelenthet.

A Korszak Gyermeknyelvi Korpuszban az *ugye* szó összesen 231 alkalommal, a dokumentumok 52%-ában fordul elő. Kollokációit tekintve (39. táblázat), a határozott névelőkkel, az *is* és az *és* szavak társaságában jelenik meg a legtöbbször.



**39. táblázat** Az *ugye* kollokációi a KorSzak Gyermeknyelvi Korpuszban

| Sz. | Szó   | Együttes előfordulások (-5 /+5) | Összes együttes előfordulás a korpuszban | LogDice |
|-----|-------|---------------------------------|--|---------|
| 1.  | a     | 113                             | 2779                                     | 10,26   |
| 2.  | is    | 47                              | 1218                                     | 10,05   |
| 3.  | és    | 47                              | 1267                                     | 10,01   |
| 4.  | hogya | 44                              | 1766                                     | 9,50    |
| 5.  | az    | 41                              | 1459                                     | 9,63    |
| 6.  | van   | 30                              | 587                                      | 10,23   |
| 7.  | meg   | 30                              | 736                                      | 9,99    |
| 8.  | mert  | 28                              | 410                                      | 10,48   |
| 9.  | nem   | 22                              | 1231                                     | 8,95    |
| 10. | volt  | 18                              | 495                                      | 9,67    |

Az *ugye* a korpuszban négyféle funkcióval jelenik meg, amelyeket a következőkben néhány kontextusban bemutatott példán figyelhetünk meg. Az első példa (7. példa) azt mutatja meg, amikor a gyermekek a szót időkitöltő elemként használják gondolkodás közben. A következő funkció a megerősítés (8. példa), amikor a beszélő biztos abban, hogy a hallgató egyetért vele, valamilyen közös háttérismeretet vagy megértést feltételez. A harmadik funkciója a megerősítés kérése (9. példa), amely során a beszélő visszaigazolást kér arra vonatkozóan, hogy az általa említett kijelentés valóban megfelel a valóságnak. A megerősítés történhet egy válasszal, ahogy a példában is látható vagy egy gesztussal, mimikával, például fejbólintással is. Ennek okán az előző két funkció sokszor csak a videó megfigyelésével különíthető el egymástól. Végül az *ugye* utolsó, a korpuszban megtalálható funkciója, amikor a beszélő saját maga iránti bizalmát fejezi ki (10. példa).

(doc#1 Mon\_Állatok\_002: A gyermekek arról beszélgetnek, hogy kinek hány háziállata van.)

- (7) A: Nekünk állatkertünk van.  
B: Az igaz.  
A: Van **ugye** a Csomó, ő egy kutya.

(doc#0 Mon\_Állatok\_001: A gyermekek arról beszélgetnek, hogy mit tehetnének az állatok védelméért.)

- (8) A: Szerinted mit tehetünk az állatok védelméért így kollektíven?  
B: Hát interneten van egy csomó ilyen csoport, hogy tüntetnek értük, illetve az állatkerteket nem támogatják, mert oda **ugye** be vannak zárva egész életükben.

(doc#5 Mon\_Állatok\_Kutya\_002: A gyermekek a kirándulásokról beszélgetnek.)

- (9) A: [...] de erről te is tudnál mesélni, mert **ugye** ott laktok az erdő közelében.  
B: Igen, igen.

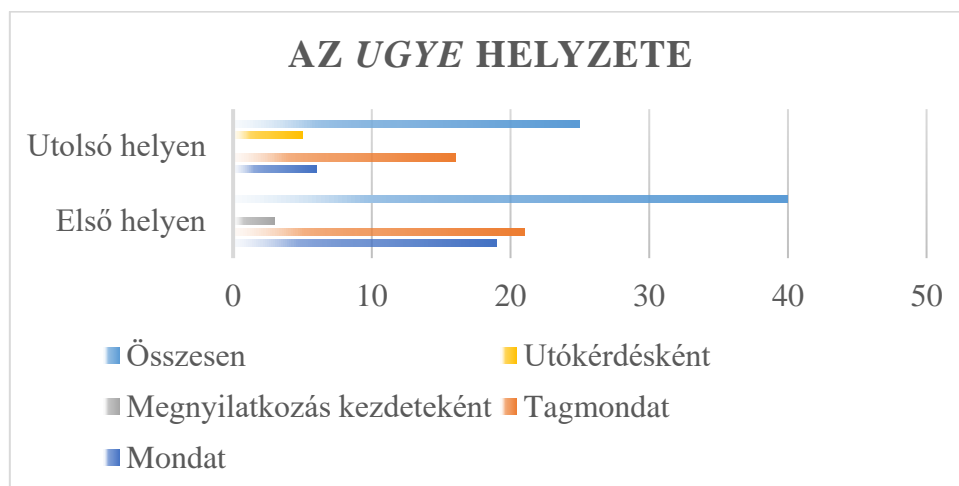
(doc#0 Mon\_Állatok\_001: A gyermekek arról beszélgetnek, hogy kinek mi a kedvenc állata.)

- (10) A: Én **ugye** minden állatot szeretek, de a rókát, meg az őzet különösen szeretem.

(KorSzak Gyermeeknyelvi Korpusz, 2020)

A fenti példákból (7–10. példa) az *ugye* használatának egy másik lényeges jellemzőjét, a megnyilatkozásokban elfoglalt pozícióit is megfigyelhetjük. A korpuszban számottevően magas százalékban (71%, 166 előfordulás) a mondat vagy tagmondat második vagy harmadik elemeként jelenik meg. A többi 65 előfordulásból (70. ábra) mondat- vagy tagmondatkezdő pozícióban 40-szer, mondat, illetve tagmondat végén összesen 25-ször szerepel. Ez utóbbi közül utókérdésként mindösszesen 5 alkalommal használják a gyermekek.

75. ábra Az UGYE pozíciója a megnyilatkozásokban



A kolligációját tekintve általában (95%) kijelentő mondatokban szerepel, kérdésekben a 231 előfordulásából 10 alkalommal jelenik meg. Ezek közül összesen 6 alkalommal utókérdésként, amelyekből ötször – ahogy már említésre került – a mondat utolsó eleme, egyszer pedig az utolsó előtti helyen fordul elő negatív utókérdésként (11. példa).

(doc#57 Mon\_Sajáttörténet\_001: A gyermekek egy kirándulásról beszélgetnek.)

- (11) A: Igen, és ott hagytam el a pulóvert, **ugye** nem?  
B: De!

(KorSzak Gyermekegyelvi Korpusz, 2020)

A többi kérdésben főleg időkitöltő szerepe van, ezenkívül egyszer jelenik meg megerősítés kérése funkcióban. Ez a 12. példában olvasható.

(doc#50 Mon\_Hp\_001: A beszélgetés témája a Harry Potter.)

- (12) A: **Ugye** milyen menőn hangzik?  
B: Nagyon.

(KorSzak Gyermekegyelvi Korpusz, 2020)

A tananyagkészítés szempontjából megtekintve a fenti eredményeket, arra következtethetünk, hogy az *ugye* használata – a korpuszban lévő nagyarányú előfordulása miatt – a mindennapi beszélgetések fontos részét képezik az adatközlő gyermekek körében. Ennek okán, ha a nyelvtanuló gyermekek is használják, akkor nagy valószínűséggel amellet, hogy hatékonyabbá és magabiztosabbá válhat a kommunikációjuk, a nyelvhasználatuk is természetesebbé válik.

Az *ugye* arányaiban kevesebb kérdőmondatban történő megjelenéséből feltételezhetjük, hogy a megerősítés kérése leggyakrabban nem kérdés formájában történik ebben a korpuszban, hanem inkább olyan módon, ahogy azt az 9. példában is olvashatjuk. Ezt az aspektust szintén érdemes előtérbe helyezni a tananyagkészítés során. Természetesen a pontosabb értelmezés érdekében érdemes lenne a beszéd más jellemzőit is megvizsgálni, például a hanglejtést, illetve a videókat is részletesen elemezni (gesztusok, mimika), de mivel a korpuszban találhatóak olyan felvételek is, amelyeken nem láthatók a gyermekek, vagy csak hanganyag, ezért jelenleg ez nehezen lenne megoldható. Ugyanakkor további felvételek gyűjtése után, egy másik kutatási projekt keretében célszerű lenne erre is fókuszálni.

A megerősítés kérése funkció használatának jellemzői mellett még azt is lényeges megjegyezni, hogy az *ugye* időkitöltő szerepe hangsúlyos a gyermekek nyelvhasználatában, ezért ezt a funkcióját ajánlatos lehet már a nyelvtanulás korai szakaszában megismertetni a nyelvtanulókkal.

## 4.2 A NEM TUD használati mintázatai

A leggyakoribb kételemű lexikai egység a NEM TUD lemmákból származik, amely összesen 292-szer jelenik meg a KorSzak GyermeKNyelvi Korpuszban. Ha részletesebben megtekintjük az előfordulásait (40. táblázat), akkor láthatjuk, hogy a leggyakrabban megjelenő alakja az egyes szám első személyben lévő határozott ragozású *nem tudom* forma, amely a találatok 77,74%-át teszi ki. A következő, a már jóval alacsonyabb találatú határozatlan ragozású változata, vagyis a *nem tudok*, amely összesen 17 alkalommal, a találatok 5,82%-ában jelenik meg. A többi ragozási formában megjelenő alakja a két első találathoz képest lényegesen alacsonyabb gyakorisági számokkal rendelkezik.

**40. táblázat** A NEM TUD kételemű lexikai egység előfordulása a KorSzak GyermeKNyelvi Korpuszban

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| <b>NEM TUD</b> | <b>292</b>      |                            |
| nem tudom      | 227             | 77,74                      |
| nem tudok      | 17              | 5,82                       |
| nem tudjuk     | 7               | 2,40                       |
| nem tudod      | 7               | 2,40                       |
| nem tudja      | 5               | 1,71                       |
| nem tud        | 5               | 1,71                       |
| nem tudtam     | 4               | 1,37                       |
| nem tudunk     | 4               | 1,37                       |
| nem tudnék     | 3               | 1,03                       |
| nem tudsz      | 2               | 0,68                       |
| nem tudta      | 2               | 0,68                       |
| nem tudtunk    | 2               | 0,68                       |
| nem tudnám     | 1               | 0,34                       |
| nem tudják     | 1               | 0,34                       |
| nem tudnának   | 1               | 0,34                       |
| nem tudhatjuk  | 1               | 0,34                       |

A leggyakoribb alak, a *nem tudom* a találatok 84%-ában (192 előfordulás) egy egységet alkot, illetve hosszabb lexikai egységek részét képezi. Az előfordulásait gyakran követik kérdőszók (Vermeki, 2023a). Az alábbi konkordanciasorban (76. ábra) található néhány példa erre. Ezenkívül 5 esetben megjelenik a *nem tudom* + főnévi igenév szerkezet is, amely a 77. ábrán látható.

## 76. ábra Találatok a *nem tudom* + kérdőszó mintázatra a KorSzak Gyermeknyelvi Korpuszban

| Left context   | KWIC                                    | Right context ↑   |
|--|---|---|
| és akkor hű, de menő, hogy ott állt, jött az állat a | <b>nem tudom</b><br>N.Nom V.Prs.Def.1Sg | <b>hány</b> négyzetméteres kennelebe, és akkor de menő, hogy levadászta, amikor n         |
| vesen gitározni, és akkor még ott van a másik,       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>hány</b> elfoglaltság, vagy éppen hegedűt kell gyakorolni, zongorát kell gyakoroln     |
| s>Ő biztos, hogy ott van, mert nem férne el az a     | <b>nem tudom</b><br>N.Nom V.Prs.Def.1Sg | <b>hány</b> tonna régiség, ami még a szüleié nagyszüleié volt, és azokat úgy őrizget      |
| m.</s><s>Így őszintén megmondom, hogy most           | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>hol</b> van, de biztos, hogy megvan még valahol. - És Bubó hogyan teljesített?</s>     |
| : YouTube-on, hogy bement valami kínai erdőbe,       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>hol</b> , és ott azt hiszem talált egy halottat és kiröhögte.</s><s>Pontosabban nei    |
| sen van megformálva.</s><s>Halálfaló.</s><s>         | <b>Nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>melyik</b> részben van először, de egy a Voldemort leghűségesebb szolgája, és e        |
| szeretem a meglepetéseket, mert idegesít, hogy       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>mi</b> az, de persze, amikor megkapom, örülök neki, és van, amikor jó érzés meg        |
| semelyik. - Hát én inkább kutya.</s><s>Igazából      | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> , de valahogy én a kutyákat talán jobban kedvelem.</s><s>Nem azt mor         |
| reg a macskát, pedig egyébként nem is.</s><s>        | <b>Nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> , de nem tudom.</s><s>Múltkor jöttem rá, hogy igazából annyira nem is        |
| jt.</s><s>Hát valamiért a szülei elleneztek, de      | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> . - Nem tudom, én nem úgy talán azért félek háziállatot tartani, mert ha n   |
| csótányok, illetve a poloskáktól, így undorodom,     | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> így, nem azoktól is, mint a csótányoktól. - Én a kigyóktól félek, mert ha v. |
| </s><s>Tessék. - Nekem a Grindelwald.</s><s>         | <b>Nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> , de a Grindelwald. - De miért? - Nekem azért tetszik a Grindelwald karal    |

## 77. ábra A *nem tudom* + Inf. mintázat megjelenése a KorSzak Gyermeknyelvi Korpuszban

CONCORDANCE

KorSzak Gyermeknyelvi Korpusz

simple **nem tudom** • 227  
3.172.78 per million tokens • 0.32%

filter ( [word="abbahagyni"] ) ( [word="kikefélni"] ) ( [word="el..."] ) • 5  
69.89 per million tokens • 0.007%

Sort word x

Details

|   | Left context  | KWIC                                  | Right context ↑   |
|---|---|---------------------------------------|---|
| 1 | doc#65 dom, én egy éve nézem körülbelül, és már hatszor végignéztém az egészet és       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>abbahagyni</b> .</s><s>Nekem ez a kedvenc sorozatom. - Nekem nem nagyon v            |
| 2 | doc#66 .</s><s>Hát szerintem azért, mert úgy hogyha elkezdünk egy sorozatot, akkor      | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>abbahagyni</b> , mivel izgalmas. - Hű ezzel én is így vagyok.</s><s>Én függővé       |
| 3 | doc#0 k a szarvashúst étteremben.</s><s>Ez így, hát így határeset.</s><s>Hogy így       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>eldönteni</b> , hogy ez most így jó vagy nem. - Meg ugye te is beszéltél arról, hog  |
| 4 | doc#5 .utya a legjobb, vagy legideálisabb háziállat, mert én egy macskáról egyáltalán   | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>elképzeln</b> , hogy ő akkor most szuperhős, meg és megment, hanem inkább €          |
| 5 | doc#6 jlje, hogy le kell nyírni a szűrét. - Igen.</s><s>A Rozival az a baj, hogy nagyon | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>kikefélni</b> , mert ugye ilyen nagyon. 5 percig nem bír nyugton lenni, de a Benit i |

A második leggyakoribb alak a *nem tudok*, amelynek összesen 17 előfordulása van a korpuszban. A *nem tudok* után a legtöbbször főnévi igenév áll (10 előfordulás), amely vagy közvetlenül utána (például *nem tudok beszélni / hegedülni / leállni / mondani / síelni / spórolni / választani*), vagy egy beékelődött *mit* után áll (13. példa).

(doc#38 Mon\_Hobby\_001: A gyermekek a kreatív hobbikról beszélgetnek.)

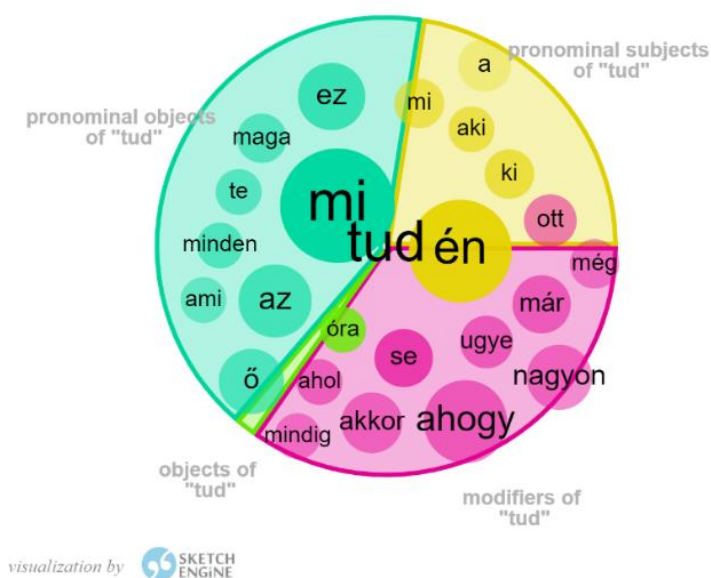
(13) A: Hát, én még megpróbáltam varrni is, bár azt még annyira nem sajátítottam el, de rajta vagyok.

B: Hát az a bajom leginkább, hogy **nem tudok mit varrni**, hogy mit varrjak, mert érted hímezni tudok állatokat, meg minden, de varrni, mit varrjak?

(KorSzak Gyermeknyelvi Korpusz, 2020)

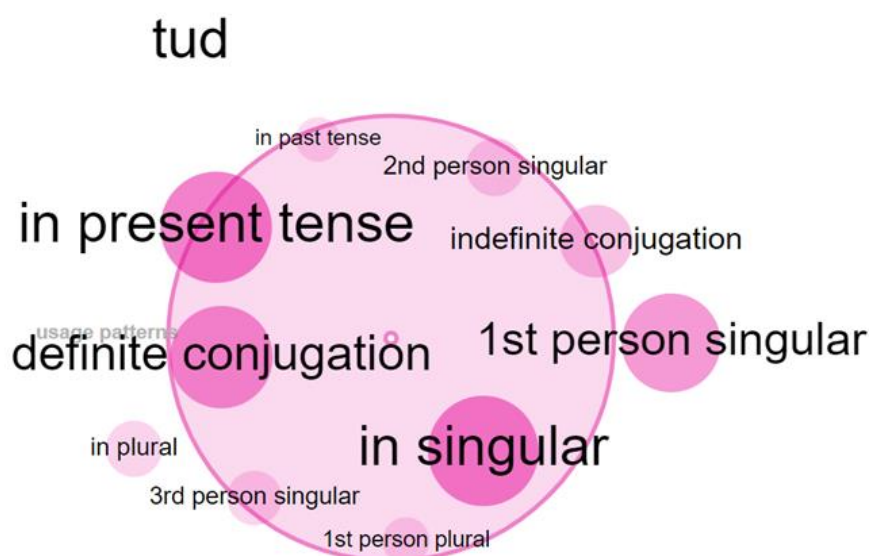
A NEM TUD találatait érdemes összevetni az egyelemű TUD találati eredményeivel is, hogy további következtetéseket vonhassunk le a használatával kapcsolatban. Ahogy a IV.2.1 fejezetben olvashattuk a TUD a 13. leggyakoribb lemma, amely 650-szer jelenik meg a korpuszban. A 650 találat közül 591-szer, tehát a találatok 90%-ában többemű lexikai egységek részét képezi, amelyek közül 346 alkalommal *tudom* alakban jelenik meg, és ahogy fentebb láthattuk (40. táblázat), ebből 227-szer, tehát a találatok 65%-ban *nem tudom* szerkezetben. Az alábbi ábra (78. ábra), amelyen a TUD kollokációit láthatjuk, szintén azt mutatja, hogy a TUD lemma legszorosabb kapcsolatban az alanyként megjelenő egyes szám első személlyel (*én*), valamint a tárgyként megjelenő MI lemmával áll (Vermeki, 2023a).

78. ábra A TUD lemma kollokációi (Vermeki, 2023a: 120)



A TUD lemma kolligációt tekintve (79. ábra) arra a megállapításra juthatunk, hogy a használata korlátozott grammatikai mintázatokkal társul. A leggyakrabban jelen idejű határozott ragozás egyes szám első személyében, vagyis *tudom* formában szerepel a korpuszban (gyakorisági mutatója: 346). Más grammatikai mintázatokban ennél jóval kevésbé gyakrabban jelenik meg. A gyermekek dialógusaiban előfordul még jelen idejű határozatlan ragozás első szám első személyben (*tudok*), ennél sokkal ritkábban jelen idejű határozott ragozás második (*tudsz*) és harmadik személyben (*tud*), valamint 27 alkalommal múlt időben (*tudtam*, *tudta*) is megjelenik. Láthatjuk tehát, hogy a TUD lemmát vizsgálva is hasonló eredményeket kapunk, mint a NEM TUD kételemű lexikai egység vizsgálatánál (Vermeki, 2023a).

79. ábra A TUD lemma kolligációi (120)



Visszatérve a kételemű lexikai egységre, a *nem tudom* kifejezés többféle jelentése és funkciója miatt fontos szerepet játszik a kommunikációban. A KorSzak Gyermeeknyelvi Korpuszban négy funkciója jelenik meg. Amellett, hogy a *nem tudom* használata egy módja lehet a tudásunk hiányának kifejezésének (14. példa), használhatjuk kérdések elkerülésére (15. példa), időkitöltő elemként (16. példa) vagy arra, hogy tudassuk a beszélgetőtársunkkal, hogy hallottuk és értjük, amiről beszél, ahogy ez a 17. példában második megnyilatkozásában történik, ahol az utolsó megnyilatkozásban elhangzó (*Igen, igen.*) mondathoz hasonló a jelentése és a funkciója. Ebből látható, hogy ez egy igen hasznos kifejezés, amelyet a nyelvtanulóknak érdemes megtanulni (Vermeki, 2022c, megjelenés alatt).

(doc#2 Mon\_Állatok\_Kutya\_002: A gyermekek egy túrázási lehetőségről beszélgetnek, amelyet korábban már kipróbált egyikük.)

(14) A: Tök jó, hogy ott voltatok amúgy a Rékával.

B: Kár, hogy **nem tudom**, hogy idén lesz-e, mert feloldják a karantént, akkor is lehetne együtt menni ilyen...

A: Igen, ez tök jó lenne.

(doc#52 Mon\_Hp\_003: A gyermekek a kedvenc Harry Potter karaktereiről beszélgetnek. C nem szeretne válaszolni arra, hogy a három főszereplő közül ki a kedvence.)



- (15) A: De ha így a három közül választanotok kéne, Ron, Hermione, Harry, akkor kit választanátok?  
B: Ron. Ron.  
C: Azt **nem tudom**.

(doc#1 Mon\_Állatok\_002: A gyermekek az interneten terjedő videókról beszélgetnek.)

- (16) A: Ez annyira gáz.  
B: Szerintem ezek a videók kiakasztóak, vagy **nem tudom**, hogy ezt minek teszik föl, hogy másokat bosszantsanak, vagy ennek valami oka van.

(doc#8 Mon\_Állatok\_Kutya\_005: A gyermekek egy közös kirándulásukról beszélgetnek.)

- (17) A: Emlékszel, amikor elmentünk kirándulni, **nem tudom** már, és hogy ott volt egy kutya, a falu kutyájának hívták, **nem tudom**, hogy emlékszel, ahol ilyen túra volt és pecsétet kellett gyűjteni?  
B: **Nem tudom**.  
A: Nem? Ott volt a Vilmos, és vettük azt a banánt.  
B: Igen, igen.

(KorSzak Gyermeeknyelvi Korpusz, 2020)

### 4.3 A MI és TUD használati mintázatai

A Korszak Gyermeeknyelvi Korpuszban található legtöbb magas gyakorisági mutatókkal rendelkező egyelemű lexéma, kételemű és háromelemű lexikai egység beágyazódik egy-egy hosszabb lexikai egységbe. Ha például visszatérünk a TUD lemmához (78. ábra) és a leggyakoribb kollokációjával a MI lemmával közös kollokációit részletesebben is megnézzük, akkor azt figyelhetjük meg, hogy a leggyakoribb lexikai egység, amelyben megjelenik, a *mit tudom én* háromelemű egység, amely a második leggyakoribb háromelemű lexikai egység a korpuszban.

**80. ábra** A *mit tudom én* háromelemű lexikai egység leggyakoribb megjelenéseinek konkordanciasora

| Left context   | KWIC ↑                       | Right context  |
|--|------------------------------|--|
| nvalóan a haszonállatokat azokat munkára tartják, <b>mit</b>     | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , igás lovat szállításra használnak, és. - Vagy h  |
| gy mennyi ideje van vele foglalkozni, mármint egy <b>mit</b>     | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , egy hörcsög nem feltétlenül igényli, hogy fogl   |
| z állnak a ketrecen, és sosem látják a napot, vagy <b>mit</b>    | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , Magdi néni udvarán kaparászik a csirke, és b     |
| most egy jobb gazdája is, ha nem hoztam el volna <b>mit</b>      | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , a menhelyről, vagy bármi, és igazából én talá    |
| it tartani.</s><s>Ilyen kis pici kutyát.</s><s>Ilyen, <b>mit</b> | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> pomerániai törpe spicc, az még úgy elmegy tal      |
| állatok ezrei, vagy direkt nem szeretik az állatokat, <b>mit</b> | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , most tudok egy ilyen példát, hogy nem mondt      |
| kor már elegend van belőle, akkor egyszerűen csak <b>mit</b>     | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , kiviszed az országútra, és otthagytott, de el se |
| az oroszlán levadássza a leggyengébb gnút, vagy <b>mit</b>       | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , zebrát akkor, nem arra törekszik, hogy az az :   |
| <s>Igen. - Meg ez is, hogy mindenki mondja, hogy <b>mit</b>      | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , az emberek mennyivel intelligensebbek az áll     |
| és tudod, ez a mennek a reptéren, és akkor jön a <b>mit</b>      | <b>tudom</b><br>V.Pr:Def:1Sg | <b>én</b> , a vakvezető kutya, éppen ő is megy föl a gép     |

A 46 találatból 38-szor, vagyis az előfordulások 82%-ában szerepel ebben az egységben (41. táblázat). Azt is láthatjuk, hogy a MI + TUD mind a 46 előfordulásában jelen időben ragozott, amelyből 38-szor határozott, és 8-szor határozatlan ragozásban jelenik meg. Három személyben találhatjuk meg a korpuszban: első sorban egyes szám elsőben (40 előfordulás), többes szám elsőben (4 előfordulás) és végül egyes szám második személyben (2 előfordulás).

**41. táblázat** A MI + TUD együttes megjelenéseinek használati mintázatai a KorSzak Gyermeknyelvi Korpuszban

| Lexikai egység                                | Találatok száma | Változatok gyakorisága (%) |
|---|-----------------|----------------------------|
| <b>MI + TUD</b>                               | <b>46</b>       |                            |
| <b>MI + TUD + Def: 1Sg</b>                    | <b>38</b>       | <b>82</b>                  |
| mit tudom én                                  | 29              | 63                         |
| hogy mit tudom én                             | 5               | 10,86                      |
| vagy mit tudom én                             | 4               | 8,69                       |
| <b>MI + TUD + NDef: 1Sg (+ Adv/Pro) + Inf</b> | <b>2</b>        | <b>4,34</b>                |

| Lexikai egység                             | Találatok száma | Változatok gyakorisága (%) |
|--|-----------------|----------------------------|
| hogy mit tudok belőle kihozni              | 1               | 2,17                       |
| mit tudok még erre mondani                 | 1               | 2,17                       |
| <b>MIT + TUD + NDef: 1Pl (+ Adv) + Inf</b> | <b>4</b>        | <b>8,69</b>                |
| nem tudunk mit csinálni                    | 1               | 2,17                       |
| nem igazán tudunk mit csinálni             | 1               | 2,17                       |
| hát nem tudunk mit csinálni                | 1               | 2,17                       |
| már nem tudunk mit szidni                  | 1               | 2,17                       |
| <b>MIT + TUD + NDef: 1Sg + Inf</b>         | <b>2</b>        | <b>4,34</b>                |
| nem tudsz mit kezdeni                      | 1               | 2,17                       |
| nem tudsz mit mondani                      | 1               | 2,17                       |

Az összefoglaló táblázatból (41. táblázat) láthatjuk, hogy a MI + TUD lemmák együttes előfordulásai közül a leggyakoribb alak a *mit tudom én*, amelynek három variációja is megjelenik a szövegekben. Az első 29 megjelenésnél biztosan egy lexikai egységet alkot, és pontatlanság, kétség vagy bizonytalanság kifejezése (18. példa), illetve időkitöltő funkcióban (19. példa) használják a korpuszban szereplő gyermekek.

(doc#49 Mon\_Hobbi\_Videójáték\_003: A gyermekek a szokásos ballagási ajándékokról beszélgetnek, illetve arról, hogy néha túl nagy ajándékokat kapnak a ballagók.)

(18) A: 250 ezer forintért egy laptopot, vagy 300-ért.

B: Mert 500 ezer, vagy kevesebért, de **mit tudom én**, olyan 450 ezer forintért egy új iPhone-t.

(doc#0 Mon\_Állatok\_001: A gyermekek a felelős állattartásról beszélgetnek.)

(19) A: Az nem az ő környezete, és nem biztos, hogy jól érzi magát.

B: Igen, meg az is szerintem fontos, fontos érv, hogy vagy fontos szempont, hogy legyen az embernek állata vagy nem, hogy mennyi ideje van vele foglalkozni, mármint egy **mit tudom én**, egy hörcsög nem feltétlenül igényli, hogy foglalkozz vele [...]

(KorSzak Gyermeknyelvi Korpusz, 2020)

A további találatokat, vagyis az ötször előforduló *hogymit tudom én*, illetve a négyszer előforduló *vagy mit tudom én* egységeket is ezekben a funkciókban használják a gyermekek. A találatok alacsony száma miatt még nem vonható le következtetés, ugyanakkor látható egy olyan különbség a két egység között, amellyel a későbbiekben, egy másik kutatási projektben érdemes lehetne esetleg foglalkozni. A *hogymit tudom én* egységet ugyanis inkább használják időkitöltő elemként a gyermekek, míg a *vagy mit tudom én* szerkezetet fele-fele arányban használják a két funkcióban.

Bizonytalanság, kétség, pontatlanság kifejtése:

- Mivel eltört a szárnya, **vagy mit tudom én**, milyen baja volt.
- Igen, én meg sietni akartam, hogy visszaérjünk Örsire, **vagy mit tudom én** mire.

Időkitöltő elemként:

- De hát az, **hogymit tudom én**, valaki most nem tudom, vissza tudom-e nézni a tegnapi TikTok használatomat.
- Meg ez is, hogy mindenki mondja, **hogymit tudom én**, az emberek mennyivel intelligensebbek az állatoknál.
- És jó volt így egy ilyen szakértővel beszélgetni arról, **hogymit tudom én**, mondtam, hogy most látok egy barnás madarat.
- És ezért meg is vásároljuk, **hogymit tudom én**, lehessen ott focizni, vagy mást csinálni.
- Nem tudom, én még úgy nem nagyon nézegettem a fiúknak az Instát, **hogymit tudom én**, kit követnek.
- És sosem látják a napot, **vagy mit tudom én**, Magdi néni udvarán kaparászik a csirke.
- Tehát, hogy amikor az oroszlán levadássza a leggyengébb gnút, **vagy mit tudom én**, zebrát.

#### 4.4 Az *annyira nem* használati mintázatai

Az *annyira nem* kételemű lexikai egység az előzőekben bemutatott NEM TUD egységnél jóval kisebb gyakorisági mutatóval rendelkezik, összesen 39-szer fordul elő a korpuszban, a dokumentumok 28%-ban szerepel (42. táblázat) egymás mellett ez a két szó. Ugyanakkor, ha az *annyira* kollokációit megtekintjük, láthatjuk, hogy a *nem* nem közvetlenül mellette, de a közelében összesen 81 alkalommal jelenik meg (43. táblázat) a korpuszban.

**42. táblázat** Az *annyira nem* kételemű lexikai egység előfordulása a KorSzak Gyermekegyelvi Korpuszban

| Lexikai egység     | Találatok száma | Változatok gyakorisága (%) |
|--------------------|-----------------|----------------------------|
| <b>ANNYIRA NEM</b> | <b>39</b>       |                            |
| <i>annyira nem</i> | 39              | 100                        |

**43. táblázat** Az ANNYIRA és a NEM együttes előfordulásai a KorSzak Gyermekegyelvi Korpuszban

| Sz. | Szó        | Együttes előfordulások (-5 /+5) | Összes együttes előfordulás a korpuszban | LogDice |
|-----|------------|---------------------------------|--|---------|
| 1.  | <i>nem</i> | 81                              | 1412                                     | 10,83   |

Az *annyira nem* korpuszban lévő előfordulásait tekintve a legtöbbször ige (*annyira nem szeretem, annyira nem támogatom*) vagy melléknév (*annyira nem egészséges, annyira nem ugrálós*) követi, illetve három különböző funkcióban jelenik meg. A kifejezést egyrészt a tagadás nyomatékosítására, erősítésére használják a gyermekek (21. példa) a KorSzak Gyermekegyelvi Korpuszban. Ezzel a funkciójával a korpuszban lévő megjelenések 41%-ban (16 előfordulás) fordul elő. A másik funkciója, amely még ennél is gyakoribb, az a vélemény kifejezése (21. példa), amely esetén pont az ellenkezője történik annak, ami az első funkciónál. A 21. példában jól látható, hogy ebben a funkcióban a véleményét kifejező gyermek az *annyira nem* lexikai egységet – a *hát így* és a *hát valahogy* kifejezések segítségével – azért használja,

hogy ne hangozzon túl elutasítónak vagy keménynek a véleménye. Ezzel a funkcióval a korpuszban lévő megjelenéseinek 59%-ában jelenik meg.

(doc#1 Mon\_Állatok\_002: A gyermekek a háziállatok tartásáról beszélgetnek.)

(20) A: A kutya nagyon büdös, tényleg mert hatalmas bundája van, és minél nagyobb a bundája, annál büdösebb szerintem, de nyáron fürdetjük rendszeresen. Nyáron jó illata van, de amikor hideg van, akkor nem.

B: A macskák pedig, hát ők **annyira nem** büdösek, csak behemperegnek a porba, és akkor lesznek még jobban büdösek.

(doc#0 Mon\_Állatok\_001: A gyermekek a vadásatról beszélgetnek.)

(21) A: Most egy kicsit más témára térve. Te mit gondolsz a vadásatról mint, a vadásatról?

B: Hát így hobbiból **annyira nem** támogatom, de hát valahogy meg kell akadályozni, hogy elszaporodjanak, mert hogyha abbahagyják, akkor be fognak jönni a falvakba, meg így borul az egész rendszer.

(KorSzak Gyermeeknyelvi Korpusz, 2020)

Az *annyira nem* harmadik funkciója az összehasonlítás (22. példa), amely összesen egy alkalommal jelenik meg a korpuszban.

(doc#6 Mon\_Állatok\_Kutya\_003: Az egyik gyermek a kutyáit mutatja be a többieknek.)

(22) A: Rozi már kicsit kezd lenyugodni, már **annyira nem** ugrálós [mint régen].

(KorSzak Gyermeeknyelvi Korpusz, 2020)

## 4.5 A ILYEN és OLYAN használati mintázatai

Az ILYEN és az OLYAN lemmák a KorSzak Gyermeknyelvi Korpuszban jelentős gyakorisági mutatókkal rendelkeznek, amely azt jelenti, hogy rendszeresen használják őket a gyermekek a beszélgetések közben. Ezek a lemmák is több többelemű lexikai egységek részeként is előfordulnak. A korpuszon végzett összehasonlító elemzésből kiderült (44. táblázat), hogy az ILYEN lemmát jóval sűrűbben használják a gyermek adatközlők, amíg az OLYAN lemmát szélesebb körben alkalmazzák. Az OLYAN lemmával kapcsolatos adatok szerint összesen 57 dokumentumban, vagyis 57 beszélgetésben fordul elő. A gyakori használatuk és sokoldalúságuk miatt ezek a szavak és azok a formulaszerű kifejezések, amelyekben szerepelnek fontos építőelemei a spontán beszédnek, ezért a nyelvtanítás szempontjából is nagy jelentőséggel bírnak.

44. táblázat Az ILYEN és az OLYAN lemma megjelenése a KorSzak Gyermeknyelvi Korpuszban

| Sz. | Lemma | Gyak. | Gyakoriság<br>/ millió | DOKgy | Relatív<br>DOKgy<br>% |
|-----|-------|-------|------------------------|-------|-----------------------|
| 1.  | ILYEN | 541   | 7561,57                | 49    | 62,82                 |
| 2.  | OLYAN | 298   | 4165,15                | 57    | 73,08                 |

Az összehasonlító elemzésből az is kiderült, hogy az ILYEN a találatok közel felében (266 előfordulás) egy olyan többelemű lexikai egység részét képezi, amelyet a korpuszból elemzésre kiválasztottunk (45. táblázat). Ezen kívül számos alkalommal jelenik meg melléknév és főnév előtt. Az ILYEN után álló melléknévek közül a leggyakoribbak a *kis* (16 előfordulás), a *nagy(obb)* (15 előfordulás), a *kicsi*, a *jó*, az *érdekes* (5 előfordulás), a *menő*, a *vicces*, a *kreatív* (4 előfordulás) és a *furcsa* (3 előfordulás). A főnevek közül pedig a *szempont* (8 előfordulás), a *dolog* (7 előfordulás), az *ember* (5 előfordulás) és a *hobby* (3 előfordulás).

**45. táblázat** Az ILYEN használati mintázatai a többemű lexikai egységekben

| Lexikai egység          | Találatok száma | Változatok gyakorisága (%) |
|-------------------------|-----------------|----------------------------|
| <b>ILYEN</b>            | <b>541</b>      |                            |
| <b>AZ ILYEN</b>         | <b>108</b>      |                            |
| az ilyen                | 101             | 93,52                      |
| azok ilyen              | 3               | 2,78                       |
| az ilyennel             | 1               | 0,93                       |
| az ilyenek              | 1               | 0,93                       |
| az ilyeneket            | 1               | 0,93                       |
| az ilyet                | 1               | 0,93                       |
| <b>EGY ILYEN</b>        | <b>108</b>      |                            |
| egy ilyen               | 106             | 98,15                      |
| egy ilyet               | 1               | 0,93                       |
| első ilyen              | 1               | 0,93                       |
| <b>VAN EGY ILYEN</b>    | <b>20</b>       |                            |
| van egy ilyen           | 11              | 55                         |
| volt egy ilyen          | 7               | 35                         |
| voltam egy ilyen        | 1               | 5                          |
| voltunk egy ilyen       | 1               | 5                          |
| <b>VAN AZ ILYEN</b>     | <b>10</b>       |                            |
| volt az ilyen           | 6               | 60                         |
| vannak az ilyen         | 4               | 40                         |
| <b>EZ EGY ILYEN</b>     | <b>6</b>        |                            |
| ez egy ilyen            | 5               | 83,33                      |
| ezek egy ilyen          | 1               | 16,67                      |
| <b>AZ ILYEN NAGY</b>    | <b>5</b>        |                            |
| az ilyen nagy           | 4               | 80                         |
| az ilyen nagyobb        | 1               | 20                         |
| <b>IS AZ ILYEN</b>      | <b>5</b>        |                            |
| is az ilyen             | 5               | 100                        |
| <b>IS VAN EGY ILYEN</b> | <b>4</b>        |                            |



| Lexikai egység   | Találatok száma | Változatok gyakorisága (%) |
|------------------|-----------------|----------------------------|
| is van egy ilyen | 4               | 100                        |

Ezzel szemben az OLYAN összesen 89 alkalommal jelenik meg a kiválasztott többelemű lexikai egységekben (46. táblázat), amely az előfordulásainak mindössze 29%-a.

**46. táblázat** Az OLYAN használati mintázatai a többelemű lexikai egységekben

| Lexikai egység       | Találatok száma | Változatok gyakorisága (%) |
|----------------------|-----------------|----------------------------|
| <b>OLYAN</b>         | <b>298</b>      |                            |
| <b>VAN OLYAN</b>     | <b>65</b>       |                            |
| van olyan            | 28              | 43,08                      |
| vannak olyan         | 12              | 18,46                      |
| volt olyan           | 12              | 18,46                      |
| vannak olyanok       | 3               | 4,62                       |
| van-e olyan          | 3               | 4,62                       |
| vagyunk olyan        | 2               | 3,08                       |
| voltak olyan         | 1               | 1,54                       |
| lenne olyan          | 1               | 1,54                       |
| lehet olyan          | 1               | 1,54                       |
| vagyok olyan         | 1               | 1,54                       |
| volna olyan          | 1               | 1,54                       |
| <b>VAN EGY OLYAN</b> | <b>11</b>       |                            |
| volt egy olyan       | 5               | 45,45                      |
| van egy olyan        | 5               | 45,45                      |
| lenne egy olyan      | 1               | 9,09                       |
| <b>DE VAN OLYAN</b>  | <b>7</b>        |                            |
| de van olyan         | 6               | 85,71                      |

| Lexikai egység       | Találatok száma | Változatok gyakorisága (%) |
|----------------------|-----------------|----------------------------|
| de vannak olyan      | 1               | 14,29                      |
| <b>NEM VAN OLYAN</b> | <b>6</b>        |                            |
| nem volt olyan       | 3               | 50                         |
| nem vagyunk olyan    | 2               | 33,33                      |
| nem vagyok olyan     | 1               | 16,67                      |

Az ILYEN lemmához hasonlóan az OLYAN is több alkalommal jelenik meg melléknevek és főnevek előtt. A melléknevek közül a *jó* (13 előfordulás) a *sok* (8 előfordulás), a *nagy* (5 előfordulás), a *vicces* (4 előfordulás) jelenik meg mellette a legtöbbször. Főnevek közül pedig az *ember* (12 előfordulás), a *dolog* (6 előfordulás), a *kutya* (5 előfordulás), a *tulajdonság* és a *sorozat* (4 előfordulás) található meg mellette. Ahogy a felsorolásból látható vannak olyan szavak, amelyek inkább az *olyannal* és vannak olyanok, amelyek inkább az *ilyen* mellett jelennek meg. Ezeknek az összefoglalását a 47. és 48. táblázatban láthatjuk. Az alábbi táblázatokban zöld színnel az ilyen mellett megjelenő, narancssárga színnel az olyan mellett megjelenő, végül pedig szürke színnel a mindkét szó mellett megjelenő főnevek, illetve melléknevek, határozószók, névmások, kötőszók láthatóak.

**47. táblázat** Az ILYEN és az OLYAN mellett megjelenő leggyakoribb főnevek

| Csak az ILYEN mellett jelennek meg a korpuszban |   | Az ILYEN és az OLYAN mellett is megjelennek a korpuszban |    |   | Csak az OLYAN mellett jelennek meg a korpuszban |   |
|---|---|--|----|---|---|---|
| szempont  | 8 | dolog  | 18 | 7 | nap   | 2 |
| mindenféle                                      | 4 | videó  | 10 | 5 | szereplő  | 2 |
| telefon   | 4 | sorozat  | 7  | 4 | fajta   | 2 |
| edzés   | 4 | kutya  | 12 | 8 | varázsvilág                                     | 2 |
| játék   | 4 | rész   | 4  | 4 | kreativitás                                     | 2 |
| hobbi   | 4 | ajándék  | 2  | 2 | terület   | 2 |
| csoda   | 3 | tulajdonság  | 4  | 4 | helyzet   | 3 |
| alap  | 3 | állat  | 3  | 4 | fej   | 3 |
| smink   | 3 | sport  | 2  | 3 | szint   | 4 |

|       |   |        |   |    |          |   |
|-------|---|--------|---|----|----------|---|
| kép   | 3 | kérdés | 2 | 3  | helyszín | 5 |
| tanár | 3 | ember  | 7 | 12 |          |   |

**48. táblázat** Az ILYEN és az OLYAN mellett megjelenő leggyakoribb melléknevek, határozószók, névmások, kötőszók

| <b>Csak az ILYEN mellett jelennek meg a korpuszban</b> |   | <b>Az ILYEN és az OLYAN mellett is megjelennek a korpuszban</b> |    |   | <b>Csak az OLYAN mellett jelennek meg a korpuszban</b> |   |
|--|---|---|----|---|--|---|
| inkább   | 7 | hogya   | 13 | 6 | igazából   | 2 |
| akár   | 2 | mindig  | 2  | 2 | még  | 4 |
| teljes   | 2 | már   | 2  | 6 | egyszer  | 2 |
| sok  | 2 |   |    |   | mégis  | 2 |
| csak   | 5 |   |    |   | konkrét  | 2 |
| tényleg  | 2 |   |    |   |  |   |
| ahol   | 2 |   |    |   |  |   |
| nagyon   | 7 |   |    |   |  |   |
| például  | 2 |   |    |   |  |   |
| ugye   | 2 |   |    |   |  |   |
| amikor   | 2 |   |    |   |  |   |
| ott  | 2 |   |    |   |  |   |

Az elemzés során kiemeltük a többemű lexikai egységek közül a VAN EGY ILYEN (20 előfordulás) és a VAN EGY OLYAN (11 előfordulás) egységeket, mivel mindkettő fontos szerepet tölt be a gyermekek dialógusaiban. A VAN EGY ILYEN kifejezés kétféle funkcióban jelenik meg a korpuszban. Ez az egység a spontán beszédben gyakran használt szókapcsolat, amelyet általában valamilyen emlék, történet vagy anekdota elmesélése során használunk. Ennek a többemű lexikai egységnek a fő funkciója a kiemelés, előkészítés (23. példa) (Vermekei, 2023b, megjelenés alatt).

(doc#07 Mon\_Állatok\_Kutya\_004: A gyermekek kutyás emlékeikről beszélgetnek.)

- (23) A: Ő eléggé hírhedt lett ott, és ő egy olyan kutyas, aki igazából ő is bichon, kicsi, de nagyon aktív és nagyon ilyen kis bolondos kutya. Ezért szerették őt nagyon a gyerekek.
- B: Hát nekünk pont a szomszédban **van egy ilyen** kutyas, úgy hívják, hogy Maci. Ő egy nagyobb testű kutya, de nem tudjuk, hogy milyen fajta, keverék. És tényleg nagyon lökött, de én nagyon sokat játszok vele.

(KorSzak Gyermeknyelvi Korpusz, 2020)

Ez az a funkciója, amelyben a legtöbbször megjelenik (18 előfordulás). Ugyanakkor a korpuszban szereplő gyermekek néhányszor pontatlanság indikálására is alkalmazzák a kifejezést (24. példa).

(doc#38 Mon\_Hobbi\_001: A gyermekek a kreatív hobbijaikról beszélgetnek.)

- (24) A: [...] meg **van egy ilyen**, nem tudom, hogy jó vagy rossz szokásnak lehet mondani, hogy úgy szoktam néha hímezni vagy varrni, hogy közben megy a tévé [...]

(KorSzak Gyermeknyelvi Korpusz, 2020)

A másik többemű lexikai egység, a VAN EGY OLYAN összesen három funkcióban fordul elő a korpuszban. A VAN EGY OLYAN kifejezés – a VAN EGY ILYEN egységhez hasonlóan – szintén az adatközlők narratív beszédére jellemző. Ugyanakkor az előző kifejezéssel ellentétben, ezt az egységet akkor használják a gyermekek, amikor egy konkrét dolgot vagy helyzetet akarnak megnevezni, de vagy nem emlékeznek pontosan a nevére, vagy nem akarják kimondani (25. példa) (Vermeki, 2023b, megjelenés alatt).

(doc#53 Mon\_Influenzerek\_001: A gyermekek influenzazerekről beszélgetnek. Az egyik beszélő elmesélt egy negatív történetet az egyikükről. Erre reagál a másik.)

(25) A: Ha **van egy olyan** influenzszer, akiről nem vagy jó véleménnyel, meg hülyeségeket posztol, akkor miért követed?

(KorSzak Gyermeknyelvi Korpusz, 2020)

Egy másik funkciója a VAN EGY OLYAN szerkezetnek a történetmesélés közben egy dolog, helyzet vagy személy azonosítása (26. példa). A használata által a hallgató számára könnyen követhetővé válik a történet.

(doc#1 Mon\_Állatok\_002: A gyermekek egy kirándulásról beszélgetnek.)

(25) A: És akkor ott sétáltunk az erdőben, és akkor ott **volt egy olyan** sódarab, amit a szarvasok nyalogattak.

(KorSzak Gyermeknyelvi Korpusz, 2020)

A VAN EGY OLYAN előfordulásainak többségében (9 előfordulás) a korábban említett két funkcióval használják az adatközlő gyermekek. A tananyagkészítés szempontjából azonban érdemes figyelmet szentelni egy kevésbé gyakori, de meghatározó diskurzusszervező funkciójának is, amely a témabevezetés. Ezt a funkciót két előfordulás mutatja, amelyekben a VAN EGY OLYAN szerkezetet arra használják a gyermekek, hogy egy kérdés feltevése során bevezessenek egy új témát a beszélgetésbe (Vermeki, 2023b, megjelenés alatt). Az ebben a funkcióban megjelenő két előfordulás a következő:

- Nekem még **van egy olyan** kérdésem, hogy
- Nekem még **lenne egy olyan** kérdésem, hogy

Ennek tanítása és gyakorlása fontos lehet a nyelvoktatásban, hiszen nagy segítséget jelenthet a nyelvtanulóknak abban, hogyan vezessenek be egy új témát egy már megkezdett beszélgetés során, hogyan szakítsanak félbe egy beszélőt, illetve egy már lecsengőben lévő beszélgetést hogyan indítsanak újra.

#### 4.6 Az *igazából* használati mintázatai

A következő kiemelésre kerülő lexikai egység a három elemből álló HÁT IGAZÁBÓL ÉN. Ez az egység nem rendelkezik nagy gyakorisági mutatókkal, ugyanakkor az *igazából* lemma

használati mintázatainak megismerése, mivel számos kommunikációs célt szolgál, hasznos lehet a tananyagkészítés szempontjából. Az *igazából* lemma összesen 170 alkalommal jelenik meg a korpuszban, amelyből 6 előfordulása a már említett háromelemű lexikai egység történik. A korpuszban három funkcióját használják az adatközlők. Az előfordulásainak nagy arányában (46 alkalommal) nyitószóként, vagy bevezető szóként használják a gyermekek, egy kijelentés bevezetésére vagy egy állítás megerősítésére (26. példa). Kötőszavak és a *hát* után, a mondat vagy a tagmondat második helyén is hasonló funkciókat lát el (82 előfordulás). Egy másik funkciója a pontatlanság, kétség, bizonytalanság kifejezése (27. példa) egy adott állítással vagy helyzettel kapcsolatban. A harmadik funkció pedig egy diskurzusszervező funkció, amely során az *igazából* a tulajdonképpen vagy valójában szinonimájaként használják a gyermekek (28. példa).

(doc#1 Mon\_Állatok\_002: A gyermekek a háziállataikról beszélgetnek.)

(26) A: **Igazából** rengeteg macskánk volt.

(doc#1 Mon\_Állatok\_002: A gyermekek a háziállatok felelős tartásáról beszélgetnek.)

(27) A: De nem tudom **igazából**. Ha valaki például bent lakik a városban, ott lehet, hogy nem a legelőnyösebb ilyen nagy kutyát tartani.

(doc#8 Mon\_Állatok\_Kutya\_005: A gyermekek arról beszélgetnek, hogy miért szeretik a kutyákat.)

(28) A: **Igazából...**

B: Mond?

A: Nem semmi.

B: Mondjad.

A: **Igazából** nekem, én azért szeretem nagyon a kutyákat, mert hogyha a kutyák azok úgy megértik, megértik az érzéseidet, amikor nekünk volt kutyánk, és mondjuk sírtam, vagy ilyesmi,

és mindig kimentem hozzá, és mindig odabújt hozzám, és ez nagyon-nagyon aranyos volt.

(KorSzak Gyermeknyelvi Korpusz, 2020)

A kolligációit tekintve az *igazából* több alkalommal fordul elő negatív jelentéstartalmú megnyilatkozásokban, mint pozitív jelentéstartalmúakban. Sokszor meg is jelenik mellette a *nem* vagy a *nincs* tagadószavunk (56 előfordulás).

### 81. ábra Példák az *igazából* + *nem* használatára

| Left context  | KWIC                                  | Right context  |
|---|---------------------------------------|--|
| ...tettő.</s><s>Mindkettőt imádom. - Én semelyik. - Hát én inkább kutya.</s><s>       | <b>igazából</b> <b>nem</b><br>Adv Adv | tudom miért, de valahogy én a kutyákat talán jobban kedvelem.</s><s>       |
| ...szűnkbe jut. - Igen. -És Zsófi néni is mondta, hogy ő is olyat írt fel, hogy utána | <b>igazából</b> <b>nem</b><br>Adv Adv | is annyira szereti, csak az jutott eszébe.</s><s>És én is, ha én megkér    |
| </s><s>Olyan, mintha látná a jövőt. - Hát már nem sokáig.</s><s>Nem tudom,            | <b>igazából</b> <b>nem</b><br>Adv Adv | szeretnék ilyen lenni, meg mondjuk egy szünyog is hiába él mondjuk va      |
| ...ondjuk egy ilyen tudományos, hogy mondjuk, tehát az ilyen rosszindulatú, ami       | <b>igazából</b> <b>nem</b><br>Adv Adv | jó semmire, az mondjuk nem, de ami így nem árt az állatnak sem.</s><s>     |
| ...jes mértékben, tehát volt esélye arra, hogy túléljen. - Ez tök jó.</s><s>Nekünk    | <b>igazából</b> <b>nem</b><br>Adv Adv | volt madarunk, csak egyszer, de anyuék azzal vigasztaltak, hogy elrepi     |
| ...g minden, már fű allig evett, meg allig ivott, az nagyon szörnyű volt.</s><s>Meg   | <b>igazából</b> <b>nem</b><br>Adv Adv | is tudtuk végül, hogy mi lett a baja.</s><s>Szóval ez a baj az ilyen állat |
| ...s><s>Úgy bent. - Hát nálam egy megint tapasztalatból mondott sztori.</s><s>        | <b>igazából</b> <b>nem</b><br>Adv Adv | teljesen szilveszterkor történt, de egyszer elkezdtek tűzijátékozni nálun  |

Amikor az IGAZÁBÓL a háromelemű lexikai egység, vagyis a HÁT IGAZÁBÓL ÉN részét képezi (82. ábra), akkor egyetlen funkciójával jelenik csak meg a korpuszban, amely az utoljára említett diskurzusszervezés, témabevezetés. Mivel ennek az egységnek csak 6 előfordulása található a korpuszban, ezért nem lehet messzemenő következtetéseket levonni, ugyanakkor az megállapítható, hogy ezekben a megjelenésekben a gyermekek minden alkalommal hasonló módon használják.

### 82. ábra A HÁT IGAZÁBÓL ÉN konkordanciasora

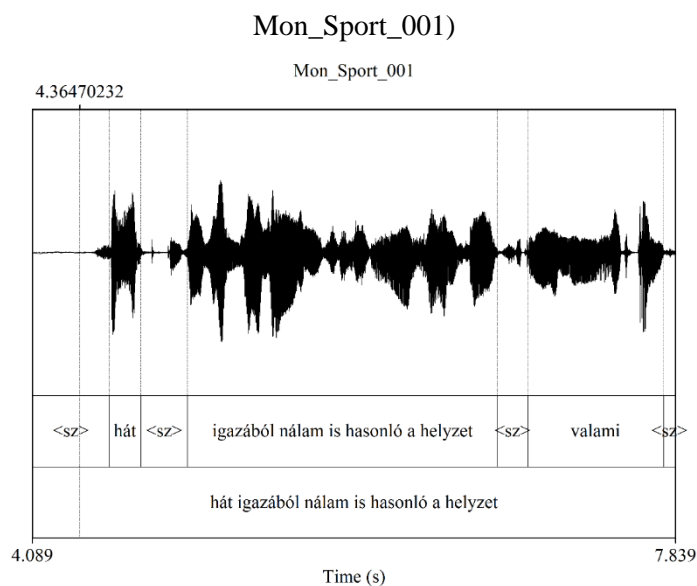
| Left context   | KWIC   | Right context                                      |
|--|--|--|
| ...; mert félek, hogy nem jól gondoskodik róla. -    | <b>Hát igazából én</b><br>Inj-Utt Adv N_Pro:1Sg:Nom    | pont attól félek, hogy nem tudok róla gondoskodni, |
| ...t, hogy szegényke még előbb meghalt volna. -      | <b>Hát igazából nekem</b><br>Inj-Utt Adv N_Pro:Dat:1Sg | a hörcsögömmel az volt, hogy én így az elején mo   |
| ...lis háziállat? - Szerintem teljesen ideális, mert | <b>hát igazából én</b><br>N:Nom Adv N_Pro:1Sg:Nom      | úgy tudom, hogy a kutyák érzik, hogyha bármi bajc  |
| ...jére.</s><s>Mesélj a kedvenc mozifilmedről. -     | <b>Hát igazából nekem</b><br>Inj-Utt Adv N_Pro:Dat:1Sg | a kedvencem az talán a Star Wars, amit még kiset   |
| ...edvem se.</s><s>Én ezt a témát elhagytam. -       | <b>Hát igazából én</b><br>Inj-Utt Adv N_Pro:1Sg:Nom    | is egyszer kezdtem bele a Riverdale-be, de aztán s |
| ...oardoznak, és én ezt szívesen kipróbálnám. -      | <b>Hát igazából nálam</b><br>Inj-Utt Adv N_Pro:Ade:1Sg | is hasonló a helyzet.</s><s>Valami extrémebb spc   |

A HÁT IGAZÁBÓL ÉN kifejezés a korpuszban egy megnyilatkozás, egy válasz elején jelenik meg, és abban támogatja a beszélőt, hogy a hallgatók számára világossá tegye, hogy őszintén, nyíltan és közvetlenül kommunikál. Tehát ebben a szerkezetben az *igazából* jelentése eltérő, inkább azt sugallja, hogy a beszélő most készül őszinte választ adni, vagy elmondani valamilyen számára lényeges dolgot. A fenti nyelvhasználati példák (82. ábra) között található egy, amely eltér ettől, és amely még a nyelvtanítás és a tananyagkészítés szempontjából érdekes lehet. Az egyetértés kifejezésére is használható *nálam is hasonló a helyzet* kifejezés, amely úgy tűnik, hogy beágyazódott a *hát igazából* kezdetű mondatba. Mivel a konkordanciasorban nem látható egyértelműen, hogy egy egységet alkot-e a két kifejezés, ezért megnéztem a Praat programban, hogy a megnyilatkozást tévő gyermek, mikor tart szünetet a beszéde közben. Ahogy a 83. ábrán láthatjuk a szünetet a gyermek a *hát* és az *igazából nálam is hasonló a helyzet* között, illetve a hosszabb egység után tartja. Ebből arra következtethetünk, hogy a *hát* ebben az esetben a beszélő gondolatainak az egység előkészítését szolgálja, vagy időkitöltő elemként működik. Maga az egység pedig az *igazából nálam is hasonló a helyzet*. Ebben a felvételben még két variációja jelenik meg ennek a szerkezetnek. Ezek a következők:

- *nekem igazából ilyen hasonló*
- *nekem is hasonló, mint az Csillánál*

Az egyetértés kifejezésével kapcsolatosan további ebben a funkcióban használt lexikai egységekről még lesz szó a későbbiekben, a IV.6. fejezetben.

**83. ábra** Szünetek helye a *hát igazából nálam is hasonló a helyzet* kimondása közben (doc#68:





## 5. A két- és többelemű lexikai egységek hosszabb egységekbe történő beágyazódása

5. kutatási kérdés: *Hogyan ágyazódnak be a két- vagy többelemű lexikai egységek hosszabb szekvenciákba? A beágyazódott lexikai egységek milyen használati mintázatot mutatnak?*

Ahogy a doktori értekezés elején, az elméleti keret kifejtésekor megjegyeztük (II.3.3 fejezet: A lexikai előfeszítés elmélete) a priming jelensége több hierarchikus szinten is jelen van, nem csupán a szótagoknál, szavaknál és szókapcsolatoknál, hanem hosszabb, több szókapcsolatból álló szekvenciáknál is megfigyelhető. Ezt jelenséget Hoey a Krishnamurty után nesting-nek vagyis beágyazódásnak nevezi (2005: 8). A beágyazódás akkor következik be, amikor a priming hatásának eredményeképpen létrejött szókapcsolat további priming folyamat részévé válik, amely során egy hosszabb szókapcsolatot vagy szekvenciát építünk fel az elsődleges szókapcsolatban található szavaktól függetlenül. A beágyazódás jelenségének a memória hatékonyabb működésében van jelentős szerepe, hiszen a nagyobb egységeket könnyebben és gyorsabban elő tudjuk hívni (Hoey, 2005: 8). Azt is megjegyeztük, hogy a gyermekek a nyelvelsajátítás folyamán ezeket a hosszabb egységeket memorizálják először és csak ezután következik a kisebb egységek megjegyzése (vö. Tomasello, 2003). Az előző fejezetekben (IV.1.–4. fejezet) látható volt, hogy a KorSzak Gyermeeknyelvi Korpuszban is találhatóak ilyen beágyazódott két- vagy többeleműből álló lexikai egységek. A korpuszban leggyakrabban előforduló beágyazódott lexikai egységek bemutatására a következőkben kerül sor.

### 5.1 Az *én is* beágyazódása

Az *én is* a második leggyakoribb kételemű lexikai egység a KorSzak Gyermeeknyelvi korpuszban, összesen 144 előfordulása található a gyermekek beszédében. Az *én is* önmagában egy válaszként alkalmazott rövid kifejezés, amelyet a beszélő arra használ, hogy jelezze, ő is egyetért azzal, amit a másik fél mondott vagy tett. A korpuszban is több ilyen jelentésben való előfordulása található (84. ábra). Hasonló funkcióban, tehát az egyetértés kifejezéseként hosszabb lexikai egységek részét is képezi. Ezeket az egységeket az alábbi táblázatban (49. táblázat) láthatjuk.

**84. ábra** Az *én is* használata rövid válaszként a KorSzak Gyermeeknyelvi Korpuszban

| Left context                                  | KWIC                              | Right context  |
|---|-----------------------------------|--|
| - Na, ezt támogatom. - Ezt <b>mondjuk</b>     | <b>én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Tudjátok, a luxusfeleségekkel kik? - Igen.      |
| ratén kívül? - Más sportra nem <b>járok</b>   | <b>én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Régebben kézilabdáztam. - Tényleg? - Igy        |
| írom, de talán leginkább Hermione. -          | <b>Én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Pontosan így vagyok vele. - Hát, nem tud        |
| íadom. - Nagyon-nagyon szeretem. -            | <b>Én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Szerintem az még talán jobb is, mint a Harry Po |
| emberek. - Elfogadnám férjemnek. -            | <b>Én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Nem, a Hagridot nem, túl sok furá dolga v       |
| elmennél? - Én az Abszol útra. - Ó,           | <b>én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Meg a hú, mi volt annak a falunak a neve'       |
| ogramokból még lesz valami. - Igen,           | <b>én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Nagyon remélem, hogy elmegyünk Horvá            |
| t néztem Bibi Blocksberget. - Juj, <b>azt</b> | <b>én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Én is.</s><s>Én azt is néztem, amikor nem a b   |
| n Bibi Blocksberget. - Juj, azt én is. -      | <b>Én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Én azt is néztem, amikor nem a boszorká         |
| port. - Ha sport, akkor én kiszállok. -       | <b>Én is</b><br>N_Pro:1Sg:Nom Adv | ./s><s>Nem néztek sportot? - Rendeset nem, autóspor    |

49. táblázat Az én is beágyazódásai hosszabb lexikai egységekbe a KorSzak Gyermekegyetemi Korpuszban

| Lexikai egység                            | Gyak. |
|---|-------|
| <b>én is</b> így vagyok ezzel/vele        | 8     |
| • ezzel <b>én is</b> így vagyok           | (2)   |
| • ezzel <b>én is</b> nagyon így vagyok    | (1)   |
| <b>én is</b> szoktam                      | 6     |
| ezzel <b>én is</b> egyetértek             | 3     |
| • ezzel <b>én is</b> teljesen egyetértek  | (1)   |
| <b>én is</b> nagyon szeretek + Inf.       | 2     |
| ezzel <b>én is</b> egyet tudok érteni     | 1     |
| <b>én is</b> nagyon szeretem              | 1     |
| meg <b>én is</b> ezen a véleményen vagyok | 1     |

A megnyilatkozásokban lévő pozícióját tekintve, a fenti táblázatban (49. táblázat) felsorolt lexikai egységekben látszik, hogy az *én is* ezekben az egységekben is általában megtartja a kezdő pozícióját vagy maximum egy szót enged maga elé. Beágyazódását tekintve

tehát az *én is* leggyakrabban az *én is így vagyok ezzel/vele* lexikai egység változataiban jelenik meg a leggyakrabban, amelynek használati mintázatai a következő alfejezetben térünk ki.

## 5.2 Az *így vagyok* és az *úgy vagyok* beágyazódása

Ahogy az előzőekben láthattuk az *én is* beágyazódása kapcsán, a korpuszban a második leggyakoribb kételemű lexikai egység az előfordulásainak számottevő részében az *én is így vagyok ezzel/vele* változataiban jelenik meg. Ebben a részben ennek az egységnek egy másik kételemű beágyazódását és egy ahhoz hasonló szerkezetű másik kételemű egység, vagyis az *így vagyok* és az *úgy vagyok* beágyazódásainak, valamint használati mintázatainak bemutatására kerül sor. Mindkét kételemű lexikai egység egy névmásból és a létige első szám első személyű ragozott alakjából tevődik össze. Az *így vagyok* 10 alkalommal, az *úgy vagyok* kifejezés pedig 9 előfordulással szerepel a KorSzak Gyermeknyelvi Korpuszban.

Kollokációikat tekintve (50.táblázat) láthatjuk, hogy vannak hasonló szavak (*én* és *vele*), amelyekkel mindkét kifejezés gyakran fordul elő, de ugyanakkor azt is észrevehetjük, hogy az *így vagyok* kollokációi a korpuszban változatosabbak, pedig a két lexikai egység gyakorisági mutatói között nincs nagy különbség. Az *így vagyok* leggyakrabban az *én* személyes névmással, amíg az *úgy vagyok* legtöbbször a *hogyan* kötőszóval jelenik meg a korpuszban.

50. táblázat Az *így vagyok* és az *úgy vagyok* kollokációi a KorSzak Gyermeknyelvi Korpuszban

| Sz. | így vagyok   |                                 |                          | úgy vagyok    |                                 |                          | LogDice |
|-----|--------------|---------------------------------|--------------------------|---------------|---------------------------------|--------------------------|---------|
|     | Szó          | Együttes előfordulások (-5 /+5) | Összes együttes előford. | Szó           | Együttes előfordulások (-5 /+5) | Összes együttes előford. |         |
| 1.  | <b>én</b>    | 10                              | 744                      | <b>hogyan</b> | 9                               | 1791                     | 8,76    |
| 2.  | <b>is</b>    | 9                               | 1218                     | <b>én</b>     | 6                               | 744                      | 7,91    |
| 3.  | <b>ezzel</b> | 6                               | 36                       | <b>vele</b>   | 4                               | 74                       | 12,06   |
| 4.  | <b>vele</b>  | 4                               | 74                       |               |                                 |                          | 10,61   |
| 5.  | <b>és</b>    | 3                               | 1687                     |               |                                 |                          | 5,86    |

Amíg az *így vagyok* beágyazódását tekintve általában az *én is így vagyok vele/ezzel*

változataiban jelenik meg, az *úgy vagyok az én úgy vagyok (vele)* ötelemű lexikai egységekben szerepel a legtöbbször. A két hosszabb lexikai egység jelentése között a különbség, hogy az előbbi azt jelenti, hogy a beszélő gyermek egyetért vagy azonosul azzal, amit a másik fél mond (29. példa), míg az utóbbi használatával a beszélő a saját álláspontját fejezi ki (30. példa).

(doc#54 Mon\_Kirándulás\_001: A gyermekek arról beszélnek, hogy kivel – barátok vagy család – szeretnek kirándulni menni.)

(29) A: [...] családdal, ismerősökkel, barátokkal, szóval úgymond bárkivel, akivel jól érzem magam, igen.

B: **Én is így vagyok ezzel**, mint az Amira [...].

(doc#54 Mon\_Kirándulás\_001: A gyermekek a kiránduláshoz megfelelő időjárási körülményekről beszélgetnek.)

(30) A: Nekem is, ha borús az idő, akkor annyira nem is szoktam kimenni, de ha jó idő van, akkor mindenképpen én nagyon szeretek kint lenni, úgyhogy nekem is mindenképpen boldogság.

B: **Én úgy vagyok vele**, hogy én kifejezetten szeretem, ha kicsit borúsabb az idő.

(KorSzak Gyermeknyelvi Korpusz, 2020)

A megnyilatkozások közben tartott szünetek helyét megvizsgálva láthatjuk, hogy míg az *én is így vagyok ezzel/vele* esetében korábban bemutatásra került (IV.3 fejezet), hogy a lexikai egységek Praat programmal történő vizsgálata során kiderült, hogy az előfordulások 33%-ában tartanak a korpuszban szereplő gyermekek pontosan az egység előtt és után szünetet, az *én úgy vagyok (vele)* lexikai egység előfordulásainak 55%-ában a négy elem előtt és után vagy pedig egy hosszabb szekvenciát követően kerül sor a szünet tartására (51. táblázat).

51. táblázat Szünetek helye az (én) úgy vagyok (vele) kimondása közben

| Dok. | N-gram (szó)        | Szünet helye<br>a lexikai egység<br>előtt / után van |     | Megjegyzés   |
|------|---------------------|--|-----|--|
|      |                     | Igen   | Nem |  |
| #2   | úgy vagyok          | x  |     | Jól mérhető a szünet.  |
| #8   | én úgy vagyok vele  | x  |     | Többelemű egység, a szünet mintája:<br><sz> én úgy vagyok vele, hogy rá<br><sz>                              |
| #40  | én úgy vagyok vele  | x  |     | Többelemű egység, a szünet mintája:<br><sz> mert én úgy vagyok vele, hogy<br>én<sz>                          |
| #49  | én úgy vagyok       |  | x   | A szünet mintája eltér:<br><sz> én úgy vagyok a játékoknál,<br>hogy most újabban elkezdtem<br>podcastet <sz> |
| #49  | úgy vagyok          | x  |     | Többelemű egység, a szünet mintája:<br><sz> aztán én úgy vagyok vele <sz>                                    |
| #51  | én úgy vagyok vele  | x  |     | Jól mérhető a szünet.  |
| #54  | én úgy vagyok vele  | x  |     | Jól mérhető a szünet.  |
| #66  | úgy vagyok          |  | x   | A szünet mintája eltér:<br><sz> és <sz> most úgy <sz> vagyok,<br>hogy így <sz>                               |
| #68  | én úgy vagyok ezzel | x  |     | Jól mérhető a szünet.  |

A hosszabb lexikai egységekbe beágyazódott *így vagyok* és *úgy vagyok* minden alkalommal egy mondatot vagy egy mondatrészt alkot, az előfordulásiak túlnyomó többségében mondatkezdő, illetve megnyilatkozást kezdő pozícióban vannak. A további használati mintázataikkal kapcsolatosan (52. táblázat) meg kell jegyeznünk, hogy bár az alacsony gyakorisági mutatóik miatt még nem vonhatunk le nagyobb következtetéseket, az látszik, hogy az *én is így vagyok ezzel/vele* egység változatait, amikor egy mondatrész szerepét töltik be, leggyakrabban a *hogya* és a *mint* kötőszók követik, amíg az *én úgy vagyok (vele)* kifejezés

minden megjelenése után a *hogy* kötőszó található meg a korpuszban.

**52. táblázat** Az *én is így vagyok ezzel / vele* és az *én úgy vagyok (vele)* lexikai egységeket követő kötőszók és előfordulásaiknak százalékos aránya

| én is így vagyok ezzel / vele |             |                         | én úgy vagyok (vele) |             |                         |
|-------------------------------|-------------|-------------------------|----------------------|-------------|-------------------------|
| Kötőszó                       | Előfordulás | Az összes előfordulás % | Kötőszó              | Előfordulás | Az összes előfordulás % |
| <b>hogyha</b>                 | 2           | 20                      | <b>hogy</b>          | 9           | 100                     |
| <b>mint</b>                   | 2           | 20                      |                      |             |                         |
| <b>de</b>                     | 1           | 10                      |                      |             |                         |
| <b>és</b>                     | 1           | 10                      |                      |             |                         |
| <b>hogy</b>                   | 1           | 10                      |                      |             |                         |

Az említett jelenség azért fontos a nyelvtanítás és a tananyagkészítés szempontjából, mert a korpuszban található adatok alapján megállapítható, hogy az *én is így vagyok ezzel/vele* kifejezés után általában vagy egy szituációt (*hogyha*), vagy egy másik személyt (*mint az Amira*) említene a beszélők, míg az *én úgy vagyok (vele)* kifejezés után minden esetben egy *hogy* kötőszóval bevezetett álláspont következik. Ennek megfelelően, ha tananyagunkba olyan minta mondatokat szeretnénk bevinni, amelyek a gyermekek természetes nyelvhasználatát tükrözik, érdemes figyelembe venni ezeket az adatokat.

### 5.3 A *mit tudom én* beágyazódása

Utolsó példaként a háromelemű egységek közül a *mit tudom én* hosszabb lexikai egységekbe történő beágyazódása kerül bemutatásra. Ahogy már korábban említésre került (IV.4.3 fejezet) a *mit tudom én* háromelemű lexikai egység összesen 29 előfordulással szerepel a KorSzak Gyermeknyelvi Korpuszban. Kollokációit tekintve a leggyakrabban a *vagy* kötőszóval szerepel együtt, de ezenkívül számos alkalommal megjelenik a közelében a *most az egy* és az *ilyen is*.

| Sz. | Szó          | Együttes előfordulások (-5 /+5) | Összes együttes előfordulás a korpuszban | LogDice |
|-----|--------------|---------------------------------|--|---------|
| 1.  | <b>vagy</b>  | 8                               | 516                                      | 8,91    |
| 2.  | <b>most</b>  | 4                               | 295                                      | 8,66    |
| 3.  | <b>egy</b>   | 6                               | 764                                      | 7,95    |
| 4.  | <b>ilyen</b> | 3                               | 541                                      | 7,43    |
| 5.  | <b>de</b>    | 4                               | 769                                      | 7,36    |
| 6.  | <b>akkor</b> | 3                               | 591                                      | 7,31    |
| 7.  | <b>hogy</b>  | 8                               | 1791                                     | 7,17    |
| 8.  | <b>az</b>    | 7                               | 1566                                     | 7,17    |
| 9.  | <b>nem</b>   | 5                               | 1412                                     | 6,83    |
| 10. | <b>meg</b>   | 3                               | 863                                      | 6,78    |

A mit tudom én kifejezés önmagában való megjelenése mellett általában egy-egy négyelemű lexikai egység részét képezi a korpuszban, amelyek a következők:

- *hogy mit tudom én* (5 előfordulás),
- *vagy mit tudom én* (4 előfordulás),
- *meg mit tudom én* (1 előfordulás),
- *és mit tudom én* (1 előfordulás).

A szünetek helyét a Praat programmal megvizsgálva láthatjuk, hogy a *hogy mit tudom én* előfordulásainak 80%-ában (53. táblázat), a *vagy mit tudom én* előfordulásainak 75%-ában (54. táblázat), illetve mind a *meg mit tudom én*, mind az *és mit tudom* esetében (55. táblázat) a szünetek tartására vagy közvetlenül a lexikai egység előtt és után kerül sor, vagy pedig egy hosszabb egység után.

**53. táblázat** Szünetek helye a *hogy mit tudom én* kimondása közben

| Dok. | N-gram            | Szünet helye<br>a lexikai egység<br>előtt / után van |     | Megjegyzés   |
|------|-------------------|--|-----|--|
|      |                   | Igen   | Nem |  |
| #2   | hogy mit tudom én |  | x   | A szünet mintája eltér:<br><sz> hogy <sz> mit tudom én |
| #2   | hogy mit tudom én | x  |     | Jól mérhető a szünet.                                  |
| #39  | hogy mit tudom én | x  |     | Jól mérhető a szünet.                                  |
| #42  | hogy mit tudom én | x  |     | Jól mérhető a szünet.                                  |
| #53  | hogy mit tudom én | x  |     | Jól mérhető a szünet.                                  |

**54. táblázat** Szünetek helye a *vagy mit tudom én* kimondása közben

| Dok. | N-gram            | Szünet helye<br>a lexikai egység<br>előtt / után van |     | Megjegyzés  |
|------|-------------------|--|-----|---|
|      |                   | Igen   | Nem |   |
| #0   | vagy mit tudom én | x  |     | Jól mérhető a szünet.   |
| #2   | vagy mit tudom én |  | x   | A szünet mintája eltér:<br><sz> vagy <sz> mit tudom én                    |
| #9   | vagy mit tudom én | x  |     | Többelemű egység, a szünet mintája:<br><sz> vagy mit tudom én milyen <sz> |
| #59  | vagy mit tudom én | x  |     | Többelemű egység, a szünet mintája:<br><sz> vagy mit tudom én mire <sz>   |

**55. táblázat** Szünetek helye a *mit tudom én* kifejezést tartalmazó egyéb lexikai egységekben  
kimondása közben

| Dok. | N-gram           | Szünet helye<br>a lexikai egység<br>előtt / után van |     | Megjegyzés            |
|------|------------------|--|-----|-----------------------|
|      |                  | Igen   | Nem |                       |
| #2   | és mit tudom én  | x  |     | Jól mérhető a szünet. |
| #58  | meg mit tudom én | x  |     | Jól mérhető a szünet. |



A *mit tudom én* kifejezés funkcióiról (pontatlanság, kétség vagy bizonytalanság, illetve időkitöltés) szintén már szó volt egy korábbi fejezetben (ld. IV.4.3 fejezet). A hosszabb lexikai egységekben történő megjelenéseivel kapcsolatban meg kell jegyeznünk, hogy amikor ezekben szerepel, akkor kizárólag időkitöltő funkcióban jelenik meg a gyermekek beszélgetéseiben. A nyelvtanítással, tananyagkészítéssel kapcsolatban azért lényeges, hogy az órákon megismerkedjünk az ilyen időkitöltő kifejezésekkel, mert a használatuk által a nyelvtanulók beszéde folyamatosabb és természetesebb lehet, hiszen gondolkodás közben, nem alakul ki csend. Ezáltal ahhoz is hozzájárul ezeknek a lexikai egységeknek az ismerete, hogy a tanulók magabiztossága növekedjen. Ezenkívül az időkitöltő funkcióban használható kifejezések nem csak azt mutatják, hogy a nyelvtanuló érti, amit mondanak neki, hanem azt is jelzik, hogy a beszélgetésben valóban részt vesznek, és aktívan hallgatnak a másik félre. Így alkalmazásuk javíthatja a kommunikációt és elősegítheti az interakciókat az anyanyelvi beszélővel.

## 6. A két- és többelemű lexikai egységek szövegben betöltött szerepük szerint

*6. kutatási kérdés: Mi a szövegben betöltött szerepe a KorSzak Gyermekegyelvi Korpuszban található két- és többelemű lexikai egységeknek?*

A tananyagkészítés során fontos szempont, hogy az általunk használt mintamondatok és mintaszövegek természetes nyelvhasználatot tükrözzenek. Ehhez elengedhetetlen, hogy megvizsgáljuk a KorSzak Gyermekegyelvi Korpuszban található két- és többelemű lexikai egységek szerepét a szövegben. Ez az elemzés lehetővé teszi számunkra, hogy kiválasszuk és megfelelő funkcióban használjuk fel azokat az egységeket, amelyeket a tananyagba szeretnénk beépíteni. A következőkben bemutatásra kerülő, kiemelkedően hasznos lexikai egységek Biber (2006: 139 – 146) kategóriái alapján kerültek besorolásra.

### 6.1 Álláspont kifejezése

A Biber (2006: 139–146) által definiált kategóriák közé tartozik az álláspont kifejezése, amelynek során a beszélő keretet ad a mondanivalójának. Ebben a kategóriában szerepelnek azok a kifejezések, amelyekkel a beszélő kifejezheti, hogy valami biztos, bizonytalan vagy lehetséges, illetve a cselekményhez való viszonyát, kívánságait, kötelezettségeit, szándékait, jóslatait és képességeit. Ebbe a kategóriába tartoznak azok a kifejezések is, amelyekkel a gyermekek a beszélgetések közben kifejezik a véleményüket, egyetértenek vagy nem értenek egyet a másikkal. Az ilyen kifejezések a természetes nyelvhasználatban gyakran előfordulnak, és lényegesek a kommunikáció hatékonysága szempontjából.

Az álláspont kifejezésének első alkategóriája az episztemikus kifejezések, olyan nyelvi elemek, amelyek kifejezik a beszélő tudását, bizonytalanságát, illetve az információ forrásának megbízhatóságát. A KorSzak Gyermekegyelvi Korpuszban a leggyakrabban előforduló ebbe az alkategóriába tartozó legnagyobb gyakorisági mutatóval rendelkező lexikai egységek a következők:

- nem tudom
- de nem tudom
- hogy nem tudom
- és nem tudom

- hát nem tudom
- én nem tudom
- nem tudom miért
- nem tudok mit + Inf.
- mert nem tudom
- így nem tudom
- de hát nem tudom

Ahogy már korábban megjegyeztük az első elem a *nem tudom* 227 alkalommal jelenik meg a korpuszban, amíg az utolsó, a *de hát nem tudom* mindössze háromszor. Ezek a kifejezések összesen 323 előfordulással szerepelnek a korpuszban, a kiválasztott két- vagy többemű lexikai egységek 2,53%-át alkotják.

A második kategória a szemléletet, modalitást kifejező lexikai egységek, amelybe beletartoznak a vágyak, óhaj, utasítás, szándék, előfeszítés, képesség és készség kifejezései, illetve a véleménynyilvánítás, egyetértést, egyet nem értést kifejező lexikai egységek. A KorSzak Gyermekeknél Korpuszban a leggyakrabban előforduló ebbe az alkategóriába tartozó legnagyobb gyakorisági mutatóval rendelkező lexikai egységek a következők:

- nagyon szeretem / nagyon szeretek / nagyon szerettem
- jó volt
- jó lenne
- a kedvenc / a kedvencem
- nagyon jó
- és szerintem
- (én) úgy vagyok vele / úgy voltam vele
- annyira nem
- jól érzem magam(at)
- nagyon jó volt / nagyon jó lenne
- ezzel én is egyetérték
- hát én is / hát nekem is
- az nagyon jó
- én is szeretnék
- én is szeretek / én is szeretem
- nem azt mondom

- nem lenne jó
- nem volt jó
- ki a kedvenc(ed)
- hát én nem / hát nekem nem
- de szerintem ez / ettől
- az jó volt
- az jó lenne / az jó lehet
- nagyon jól éreztük / érezzük magunkat
- tök jó volt
- tök jó lenne
- és szerintem nagyon + Adj.
- (hogya) mit gondolsz a
- és szerintem ez
- én is így vagyok vele /ezzel
- én is nagyon szeretek / szeretem / szerettem
- az nagyon jó volt
- én is egyet tudok (ezzel / veletek) érteni
- nekem az a kedvenc(em)
- nektek mi a kedvenc
- hogy mi a kedvenc
- nekem is hasonló a helyzet / véleményem
- nekem nem nagyon van

A szemléletet, modalitást kifejező lexikai egységek közül a leggyakoribb a *nagyon jó* kifejezés, amely összesen 44 alkalommal jelenik meg a korpuszban. Ehhez az alkategóriához tartozó lexikai egységek a korpuszból kiválasztott két- és többemű lexikai egységek mintegy 3,47%-át alkotják, összesen 443-szor szerepelnek a gyermekek beszélgetéseiben.

A tananyagkészítés szempontjából lényeges volt összegyűjteni azokat a lexikai egységeket, amelyeket a gyermekek a véleményük kifejezésére, illetve arra használnak, hogy másoktól megtudják az álláspontjukat. A korpuszban szereplő gyermekek a véleményüket a következő táblázatban (56. táblázat) lévő módokon fejezik ki (Vermekei, 2022c, megjelenés alatt).

**56. táblázat** Vélemény kérésére és vélemény kifejezésére szolgáló lexikai egységek a Korszak Gyermeknyelvi Korpuszban

| Vélemény kérése   | Vélemény kifejezése  |
|---|--|
| <ul style="list-style-type: none"> <li>• nektek mi a kedvenc</li> <li>• hogy mi a kedvenc</li> <li>• (hogy) mit gondolsz a</li> <li>• ki a kedvenc(ed)</li> </ul>   | <ul style="list-style-type: none"> <li>• nagyon szeretem / nagyon szeretek / nagyon szerettem</li> <li>• a kedvenc / a kedvencem</li> <li>• és szerintem</li> <li>• és szerintem nagyon + Adj.</li> <li>• és szerintem ez</li> <li>• nekem az a kedvenc(em)</li> <li>• (én) úgy vagyok vele / úgy voltam vele</li> </ul> |
| Egyetértés  | Egyet nem értés  |
| <ul style="list-style-type: none"> <li>• ezzel én is egyetértek</li> <li>• hát én is / hát nekem is</li> <li>• én is szeretnék</li> <li>• én is szeretek / én is szeretem</li> <li>• nagyon jó lenne</li> <li>• az jó lenne / az jó lehet</li> <li>• tök jó lenne</li> <li>• én is így vagyok vele /ezzel</li> <li>• én is nagyon szeretek / szeretem / szerettem</li> <li>• az nagyon jó volt</li> <li>• én is egyet tudok (ezzel / veletek) érteni<br/>nekem is hasonló a helyzet / véleményem</li> </ul> | <ul style="list-style-type: none"> <li>• annyira nem</li> <li>• nem azt mondom</li> <li>• nem lenne jó</li> <li>• hát én nem / hát nekem nem</li> <li>• de szerintem ez / ettől</li> </ul>   |

## 6.2 Diskurzusszervező kifejezések

A második kategóriát a diskurzusszervező kifejezések alkotják (vö. Biber, 2006: 139–146). Ezek a kifejezések segítik a beszélőt abban, hogy bevezesse a témát, pontosítsa azt, illetve egy bizonyos dologra koncentrálja a hallgató figyelmét. Ezek a kifejezések szintén fontos szerepet töltenek be a spontán beszédben.

A diskurzusszervező kifejezések első alkategóriája tehát a témabevezetés. A KorSzak Gyermeknyelvi Korpuszban lévő felvételeken a gyermekek a következő kifejezéseket használják erre a funkcióra a leggyakrabban:

- van egy / volt egy
- volt az /van az / vannak az
- van egy ilyen / volt egy ilyen
- volt az ilyen / vannak az ilyen
- volt egy olyan / van egy olyan
- (nekem) az lenne a /az első kérdésem

Ezek a témabevezető kifejezések a korpuszban összesen 203 alkalommal jelennek meg, amely a kiválasztott lexikai egységek 1,59%-át teszi ki.

A második alkategória a téma pontosítására használt lexikai egységek kategóriája, amelyek közül a leggyakoribbak a korpuszban a következők:

- EZ A / AZ A / EZ AZ / AZ AZ
- AZ IS
- ÉS EZ
- AZ VAN
- van ez a / volt ez a
- volt az az / vannak azok az

A fent felsorolt kifejezések közül az első négy helyen szereplő lemmatizált egységek minden változata ebbe a kategóriába tartozik, tehát például az EZ A változatai az *ez a*, *ezt a*, *ebben a*, *ezeket a* és a többi változata. Ezek a kifejezések összesen 637-szer jelennek meg a korpuszban, amely a kiválasztott két- vagy többeléből álló lexikai egységek 4, 99%-át teszik ki.

A harmadik kategória az azonosításra, a hallgató témára való fókuszálására szolgáló kifejezések, amelyek a gyermekek felvételeiben a következők a leggyakoribbak:

- AZ EGYIK
- A MÁSIK
- az volt a / az lenne a
- azzal van a baj
- az a baj / az a bajom
- az a lényeg(e)

Ennek a kategóriának a lexikai egységei közül – az előzőhöz hasonlóan – az első kettőnek (AZ EGYIK, A MÁSIK) szintén az összes változata ide sorolható. Ezek a kifejezések összesen 146 alkalommal jelennek meg a korpuszban, a kiválasztott lexikai egységek, 1,14%-át alkotva.

### 6.3 Hivatkozás kifejezése

Az utolsó kategória, a hivatkozás kifejezése szintén több alkategóriát foglal magában. Ezt a funkciót a beszélő arra használja, hogy beazonosítson valamit, vagy valaminek a szerinte fontos tulajdonságát.

A pontatlanság indikálása az első alkategória, amelyet meg kell említenünk. Ezt a funkciót a következő lexikai egységek segítségével használják a gyermekek a leggyakrabban:

- nem is tudom
- mit tudom én
- vagy nem tudom
- vagy hogy mondjam
- vagy nem is tudom
- hogy is mondjam
- hogy mit tudom én
- vagy nem is tudom
- vagy mit tudom én
- hát igazából én /nekem / nálam

Ezek a kifejezések összesen 126 alkalommal jelennek meg a korpuszban, ami azt jelenti, hogy a kiválasztott két- vagy többelemből álló lexikai egységek 0,98%-át teszik ki.

A következő alkategória a tulajdonság megadása, amely az egyik legkisebb kategória, összesen 12 alkalommal jelennek meg az eben található kifejezések a korpuszban, ami azt jelenti, hogy a kiválasztott lexikai egységeknek csak a 0,09%-át teszik ki. Ebben a funkcióban a KorSzak Gyermeeknyelvi Korpuszban lévő felvételeken a gyermekek a következő kifejezéseket használják a leggyakrabban:

- az ilyen nagy(obb)
- ő is nagyon + Adj

- de ő is nagyon + Adj

A harmadik alkategória az idő és a hely megadása, amelyben szereplő lexikai egységeket a gyermekek történetmesélés közben használják. Az itt található kifejezések szinte mindegyike rögzült formulaszerű egység. A gyermekek a korpuszban a következő lexikai egységeket használják a leggyakrabban ebben a funkcióban:

- és akkor
- és akkor ott
- ott van / volt / voltam / voltunk / voltak
- ott van / volt / voltam / voltunk / voltak a/az
- ott van / volt egy
- egy idő után
- és egyszer csak
- is ott volt / van / voltam
- akkor ott van
- amikor kicsi voltam / amikor kisebb voltam / mikor kisebb voltam

A gyakoriságukat tekintve ezek a lexikai egységek 284 alkalommal szerepelnek a korpuszban, amely a kiválasztott két- vagy többelemű lexikai egységek 2,22%-át teszi ki. A tananyagkészítés szempontjából a leghasznosabb kifejezések az *és akkor*, az *egy idő után*, az *és egyszer csak*, illetve az *amikor kicsi/kisebb voltam*, amelyek mind nagy segítségére lehetnek a nyelvtanulóknak egy-egy történet mesélése során.

Az utolsó alkategória ebben a csoportban a multifunkcionális hivatkozás, amelyek közül a korpuszban szereplő gyermekek a következőket használták a leggyakrabban:

- és akkor így
- és akkor úgy
- az volt az első
- még akkor is

Ezek a kifejezések összesen 30 alkalommal jelennek meg a korpuszban, vagyis a kiválasztott lexikai egységek 0,23%-át teszik ki.



## 6.4 A lexikai egységek szövegben betöltött szerepeinek összefoglalása

A következő táblázatban összefoglaltam (57. táblázat), hogy a kiválasztott lexikai egységek mely Biber-féle funkciókban jelennek meg a korpuszban. A második, harmadik és negyedik oszlop, amelyeket A, B és C betűvel jelöltem mutatják a kategóriákat, illetve ezeken belül láthatóak az alkategóriák is. Az 1–3. sorokban láthatóak a lexikai egységek típusai (két-, három- és négyelemű). A 4. sorban láthatóak a lexikai egységeknek azok a változatainak száma, amelyek az alkategóriákhoz sorolhatóak, az 5. sorban pedig ezek összesítése, vagyis az jelenik meg, hogy a fő kategóriákban hányféle változata jelenik meg a lexikai egységeknek. Az utolsó sorban, vagyis a hatodikban a lexikai egységek abszolút gyakorisági mutatója látható alkategóriánkénti eloszlásban. Az E oszlop pedig azt mutatja meg, hogy a kutatás során kiválasztott lexikai egységeknek hány változata van összesen, illetve a 6. sorban megjelenő szám az összes kiválasztott lexikai egység összesített abszolút gyakorisági mutatóra utal.

Ami először ez alapján az összefoglaló táblázat alapján feltűnhet nekünk, az az, hogy a korpuszból a feltételrendszer alapján kiválasztott lexikai egységek összes változata 816 darab, ugyanakkor a Biber-féle kategóriákba és alkategóriákba besorolt lexikai egységeké összesen 237. A két szám azért mutat eltérést, mert a két- és a háromelemű lexikai egységek számos alkalommal beágyazódnak a náluk hosszabb egységekbe. A táblázatba kizárólag azok a két- és háromelemű lexikai egységek kerültek bele, amelyek önmagukban is az adott kategóriába tartoznak. Abban az esetben, ha beágyazódtak egy hosszabb lexikai egységbe, akkor a hosszabb egységben betöltött szerepük szerint kerültek bele a táblázatba.

Az összefoglaló táblázat adatait részletesebben megtekintve láthatjuk, hogy a legnagyobb kategória a diskurzusszervező kifejezések, amelyet az álláspont kifejezése és végül a hivatkozás kifejezése követ. A diskurzusszervező kifejezések közül a legtöbbféle lexikai egységet a téma pontosítására használják a gyermek adatközlők, majd ezután a fókusz kialakítására, a téma bevezetésére használt kifejezések következnek. A második legnagyobb kategóriában, vagyis az álláspont kifejezésére használt lexikai egységeknél a szemlélet, modalitás alkategória az első, amely után az episztemikus kifejezések következnek. Végül a harmadik kategóriában a legtöbbféle lexikai egységet a gyermekek a hely és idő megadása funkcióban, ezt követően a pontatlanság indikálása, a multifunkcionális hivatkozás, végül pedig a tulajdonság megadása funkcióban használják. Azt is láthatjuk az adatok részletesebb megfigyelésekor, hogy a kételemű lexikai egységeknek legtöbb esetben a téma pontosítása a funkciója. A háromeleműeknek a szemlélet, modalitás, illetve a hely és idő megadása, a négyelemű egységeknek pedig a szemlélet, modalitás funkcióban található a legtöbb változata.

57. táblázat A lexikai egységek szövegben betöltött szerepük szerinti kategóriáinak összefoglaló táblázata

|    | A                                      | B                    |                      | C                             |                    |                    | D                       |                      |                      |                              | E  |
|----|--|----------------------|----------------------|-------------------------------|--------------------|--------------------|-------------------------|----------------------|----------------------|------------------------------|--|
|    | A leggyakoribb lexikai egységek        | Álláspont kifejezése |                      | Diskurzusszervező kifejezések |                    |                    | Hivatkozás kifejezése   |                      |                      |                              | A kiválasztott lexikai egységek összes változata |
|    |  | episztemikus         | szemlélet, modalitás | témabevezetés                 | a téma pontosítása | azonosítás, fókusz | pontatlanság indikálása | tulajdonság megadása | hely és idő megadása | multifunkcionális hivatkozás |  |
| 1. | Kételemű lexikai egységek változatai   | 1                    | 10                   | 5                             | 87                 | 12                 | 0                       | 0                    | 8                    | 0                            | 448  |
| 2. | Háromelemű lexikai egységek változatai | 9                    | 27                   | 6                             | 4                  | 6                  | 8                       | 2                    | 20                   | 3                            | 340  |
| 3. | Négyelemű lexikai egységek változatai  | 1                    | 19                   | 2                             | 0                  | 1                  | 4                       | 1                    | 0                    | 1                            | 28   |
| 4. | Változatok összesen                    | 11                   | 56                   | 13                            | 91                 | 19                 | 12                      | 3                    | 28                   | 4                            | 816  |
| 5. |  | 67                   |                      | 123                           |                    |                    | 47                      |                      |                      |                              |  |
| 6. | Gyakoriság                             | 323                  | 443                  | 203                           | 637                | 146                | 126                     | 12                   | 284                  | 30                           | 12.748   |

## V. A kutatási eredmények felhasználása

### 1. A Magyar felfedezőről

A jelen doktori értekezésben bemutatott kutatás eredményei – és maga a KorSzak Gyermeknyelvi Korpusz – nem csupán elméleti értelemben lehetnek lényegesek, hanem számos gyakorlati alkalmazási lehetőséget is rejtenek magukban a nyelvtanítás és tananyagkészítés terén. Ebben a fejezetben ez utóbbi, vagyis a kutatás eredményeinek gyakorlati hasznosításának bemutatására kerül sor. Ahogy már többször is megemlítettem a korábbi fejezetekben, a korpuszépítés, valamint a kutatás fő kimeneti célja is egy korpuszinformált tananyag elkészítése magyarul tanuló gyermekek számára, amely a természetes nyelvhasználatot tükrözve és megfelelő módszertannal felépítve hatékony eszköze lehet a magyar mint idegen nyelvi óráknak. A következőkben tehát ennek a készülőben lévő tananyagnak, vagyis a *Magyar felfedezőnek* (Kasikci – Vermeki, 2021) az ismertetésére kerül sor olyan kérdések mentén haladva mint például, hogyan kerülnek felhasználásra a korpuszalapú kutatásból származó adatok, hogyan épül fel a tananyag, valamint, hogy milyen módszertant alkalmaznak a szerzők a tananyag készítése során.

A *Magyar felfedező* egy projektalapú, modellalapú (vö. Szita – Pelcz, 2017) és korpuszinformált komplex tananyag, amely a KER szerinti öt nyelvi szinten (A1–C1) készül, elsősorban a 12–13 éves gyermekek számára, valamint kitágítva az ideális időkeretet, a 11–14 éves korosztályban használható hatékonyan. A szerzők<sup>21</sup> a tananyag célcsoportjának elsősorban a származásnyelvi órákon tanuló gyermekeket tekintik, ugyanakkor úgy gondolják, hogy a tananyag több nyelvi szinten készülő fejezetei miatt magyar mint idegen nyelvi csoportokban is használhatóvá válhat. Jelenleg egy mintafejezet és még egy fejezet készült el mind az öt nyelvi szintre, tehát összesen tíz fejezet<sup>22</sup>. A további fejezetek előkészületei pedig már zajlanak. Ennek a bemutatásnak a fókuszában főként a mintafejezet, azaz a 4. fejezet áll, és a példák kiválasztása is nagyobb mértékben erre a fejezetre koncentrálódik (Vermeki, 2021b).

---

<sup>21</sup> A *Magyar felfedező* című tananyag szerzői Kasikci Katalin és Vermeki Boglárka

<sup>22</sup> [magyarfelfedezo.hu](http://magyarfelfedezo.hu)

## **2. Hogyan használható a KorSzak Gyermeeknyelvi Korpuszon végzett korpuszalapú kutatás a tananyag elkészítése során?**

Ahogy már korábban kifejtettük (II.6.5.1.1 fejezet) a tananyagok szerzői számára a korpuszok rendkívül hasznos eszközöket jelentenek a munkájuk során. A korpuszokat több területen is alkalmazhatják, például a hasznos lexika kiválasztásakor, a lexika és szövegtípus kapcsolatának vizsgálatokor, valamint a tanított nyelvi formák arányának és sorrendjének kialakításakor. Emellett a korpuszok segítségével létrehozhatnak egy szintezett lexikogrammatikai tanmenetet, és elkerülhetik az olyan, nyelvhasználatot rosszul reprezentáló szövegek írását, amelyek csak a saját intuíciójukra támaszkodva készülnek. A korpuszok továbbá segíthetnek a valós szituációk kiválasztásában, és olyan döntések meghozatalában, amelyek befolyásolják, hogy a tananyag valódi, természetes nyelvhasználatot tükrözzön. A *Magyar felfedező* készítése során a szerzők is hasonló módokon alkalmazzák a korpuszokat. Ebben a részben a II.6.5.1.1 fejezetben összegyűjtött, a korpuszok tananyagkészítés során történő felhasználásával kapcsolatos pontok alapján végig haladva kerül ismertetésre, hogy ez pontosan hogyan történik.

### **2.1 A hasznos lexika kiválasztása a tananyag számára**

A tananyagban megjelenő lexika kiválasztásakor a *MagyarOK* szerzőihez hasonlóan, a *Magyar felfedező* szerzői is a témakörökből indulnak ki (Szita – Pelcz, 2020). A III.1.2.1 fejezetben már szóba került, hogy a témakörök kiválasztása gyermekek segítségével történt. Egy felső tagozatos gyermekeket megcélzó felmérés során, amelyben 138 általános iskolás diák vett részt kerültek összegyűjtésre a lehetséges témakörök, amelyeket előbb életkor szerint csoportokba soroltunk, majd gyakorisági alapon választottuk ki közülük a legfontosabbakat. A tananyag témakörei ezek a témakörök kiegészítve azokkal az elemekkel, amelyeket a nyelvtanári tapasztalatunkból szükségesnek és hasznosnak tartottunk. A tananyag lexikáját a témákat alapul véve gyakorisági vizsgálatok által választottuk ki, illetve természetesen itt is hagyatkoztunk a tanítási tapasztalatainkra és figyelembe vettük azokat a céljainkat is, amelyeket az egyes fejezetekhez, illetve a teljes tananyaghoz rendeltünk. Ilyen cél például a 4. fejezetben lévő projekt, egy állatkertről szóló reklámvideó elkészítése, amelynek megvalósulása céljából a

fejezetben minden nyelvi szinten meg kellett jelennie azoknak az állatoknak az elnevezéseinek is, amelyek tipikusan állatkertben élnek (például 85. ábra).

85. ábra Milyen állatok élnek az állatkertben? (A1)

4

MILYEN ÁLLATOK ÉLNEK AZ ÁLLATKERTBEN?

1. Kösd össze a szavakat a képekkel!

|  |   |  |   |
|--|---|--|---|
|     |  |  |    |
|     | pingvin   | medve  | oroszlán  |
|   | hal   | tigris   | elefánt   |
|   | madár   | zsiráf   | zebra   |
|   | ló  | majom  | kenguru   |
|  |   |  |  |
|  |   |  |  |

Ezeknek nagyrészt a gyermekek beszélgetéseiből gyűjtöttük össze. Így jelent meg például gyakorisági alapon kiválasztva A1-es nyelvi szinten készült fejezetben a kutya, macska, és a hörcsög mellett, a medve, a róka, az őz, vagy A2-es nyelvi szinten a borz. Ezenkívül azt is fontosnak tartottuk, hogy emellett még az olyan lexikai egységek is megjelenjenek, amelyekkel egy-egy állatot be tudunk mutatni. A1-es nyelvi szinten (86. ábra) az állatkertben élő állatok bemutatása során öt kérdésre keressük a választ (*Hol él? Milyen színű? Hány kiló? Hány méter? Mit eszik?*), ugyanakkor megpróbáljuk megtartani a rövid szövegek természetes hangzását. Tehát nem leegyszerűsített mondatokat alkotunk, mint például:

- *Az elefánt Afrikában él. Az elefánt szürke. Az elefánt 3000 kiló. Az elefánt 3 méter. Az elefánt növényt eszik.*

Inkább arra törekszünk, hogy a szöveg természetességét adó lexikai egységek többször is feltűnjenek a fejezetben. Például:

- *Az elefánt Afrikában él. A szürke színű állat, akár 3 méter is lehet. A súlya elérheti a 6000 kilogrammot. Növényeket eszik.*
- *A kenguru Ausztráliában él. 1 méter 70 centiméter magas és a súlya akár 90 kilogramm is lehet. A barna színű állat növényeket eszik.*
- *A pingvin a Déli-sarkon él. Ez a fekete-fehér színű állat halat eszik. A pingvin 50 centiméter magas és 2 kilogramm.*
- *A tigris Ázsiában él. Ez a sárga-fekete színű állat húst eszik. A tigris akár 100 kilogramm is lehet és 120 centiméter a magassága.*

### 86. ábra Az állatok bemutatása (A1)

4. Melyik szóval tudsz válaszolni a kérdésekre? Írd be!

húst, 70, zöld, kék, 200, szürke, növényt, 50, halat, fekete, 3000, gyümölcsöt

|                                      |               |                           |
|--------------------------------------|---------------|---------------------------|
| Hány kilogramm / méter / centiméter? | Milyen színű? | Mit eszik?<br><i>húst</i> |
|--------------------------------------|---------------|---------------------------|

5. Beszélgetetek párokban a minta alapján!

6. Hallgasd meg a felvételt, és egészítsd ki a névjegyeket!

- Szerinted hány kiló egy elefánt?  
- Szerintem 100 kiló.

| elefánt                     | kenguru             | pingvin             | tigris              |
|-----------------------------|---------------------|---------------------|---------------------|
| Hol él?<br><i>Afrikában</i> | Hol él?             | Hol él?             | Hol él?             |
| Milyen színű?               | Milyen színű?       | Milyen színű?       | Milyen színű?       |
| Hány kiló? _____            | Hány kiló? _____    | Hány kiló? _____    | Hány kiló? _____    |
| Hány méter? _____           | Hány méter? _____   | Hány méter? _____   | Hány méter? _____   |
| Mit eszik?<br>_____         | Mit eszik?<br>_____ | Mit eszik?<br>_____ | Mit eszik?<br>_____ |

7. Egészítsd ki a mondatokat!

Az *elefánt* \_\_\_\_\_ Afrikában él. A \_\_\_\_\_ színű állat, akár 3 méter is lehet.

A súlya elérheti a \_\_\_\_\_ kilogrammot. Növényeket eszik.

A kenguru \_\_\_\_\_ él. 1 méter 70 centiméter magas, és a súlya akár \_\_\_\_\_ kg is lehet. A barna színű állat \_\_\_\_\_ eszik.

A \_\_\_\_\_ a Déli-sarkon él. Ez a \_\_\_\_\_-fehér színű állat halat eszik.

A pingvin 50 centiméter magas és \_\_\_\_\_ kilogramm.

A tigris \_\_\_\_\_ él. Ez a sárga-fekete színű állat \_\_\_\_\_ eszik. A tigris akár 100 kilogramm is lehet, és \_\_\_\_\_ centiméter a magassága.

A2-es nyelvi szinten újra visszatérnek ezek a leírást segítő lexikai egységek és ki is egészülnek újabb elemekkel (87. és 88. ábra). Például olyan kifejezésekkel, mint a *növényevő*, *ragadozó*, *testsúly*, *élelem*. Ezen a szinten azt is megtanuljuk, hogyan tudunk két állatot összehasonlítani, illetve hogyan tudunk az állati legekről beszélni. Így ezeket is be tudjuk építeni egy-egy állat bemutatásába.

### 87. ábra Az állatok bemutatása (A2) 1.

#### 4. Találd ki, melyik állatról szól a szöveg!

Ez az állat Afrikában él. Három méter magas, és szürke színű. Nagy füle van. Növényevő: fűvet, gyümölcsöt és leveleket eszik. Tud futni, úszni és ugrani. Akár 70 évig is él.

Ez az állat a Déli-sarkon él. Körülbelül egy méter magas, és fekete-fehér színű. Ragadozó, halat eszik. Madár, de nem tud repülni.

Ez az állat Ausztráliában él. Körülbelül egy és fél méter magas, és barna színű. Növényevő: fűvet és leveleket eszik. Tud ugrálni és bokszozni.

#### 5. Készíts hasonló leírást az 1. feladat állatairól!

#### 6. Összekeveredtek a mondatok. Találd ki, melyik két állatról írták őket!



1. állat: \_\_\_\_\_  
számok: \_\_\_\_\_

1. Nagyon gyors állat.
2. Az erdőben él.
3. Szürke színű bundája van.
4. Gazellákat és zebrákat eszik.
5. Sárga-fekete színű bundája van.
6. Nyulakat és más erdei állatokat eszik.
7. A kutya rokona.
8. A macska rokona.
9. A szavannán él.
10. Csapatban él.



1. állat: \_\_\_\_\_  
számok: \_\_\_\_\_

### 88. ábra Az állatok bemutatása (A2) 2.

Melyik állat

leg- + -bb    -'bb    -ebb    -abb

a leghangosabb?

Melyik a legbűdösebb?

#### 3. Hallgasd meg az állatokról szóló szöveget, és egészítsd ki a táblázatokat!

| Elefánt   |                                 |
|-----------|---------------------------------|
| Élőhely:  | Afrika                          |
| Szín:     | szürke                          |
| Méret:    | 3 m                             |
| Testsúly: | 3000–6000 kg                    |
| Élelem:   | növények: fű, levél             |
| Rekord:   | a legnehezebb szárazföldi állat |

| Kolibri   |                  |
|-----------|------------------|
| Élőhely:  | Közép-Amerika    |
| Szín:     |                  |
| Méret:    |                  |
| Testsúly: | 2 g              |
| Élelem:   | a virág nektárja |
| Rekord:   |                  |

| Bógómajom |                |
|-----------|----------------|
| Élőhely:  | Dél-Amerika    |
| Szín:     |                |
| Méret:    | 1 m            |
| Testsúly: |                |
| Élelem:   |                |
| Rekord:   | a leghangosabb |

Néhány további példát is szeretnék megemlíteni. Az A2-es szinten lévő állati rekordokról szóló részhez (89. ábra), valamint a A2-es és a B1-es szinten lévő *Milyen lehet egy kutya?* (90. ábra) és *Milyen lehet egy háziállat?* (91. ábra) című részekhez tartozó szókincset szintén főleg a KorSzak Gyermekenyelvi Korpuszban szereplő gyermekek beszélgetéseiből választottuk ki, a IV.2.2 fejezetben már felsorolt legnagyobb gyakorisági mutatóval rendelkező mellékneveket alapul véve, valamint a gyermekek által említett kutyafajtákat és állatfajtákat felhasználva.

## 89. ábra Állati rekordok (A2)

4

### MELYIK A LEGLASSABB ÁLLAT A VILÁGON?

Biztos mindenki hallott már biológiaórán vagy a tévében az állatvilág LEG-jeiről. Már mindenki tudja, hogy a leggyorsabb állat a gepárd, a lelassabb pedig a lajhár...

#### 1. Kösd össze a szavakat a jelentésükkel!

- |                    |                               |
|--------------------|-------------------------------|
| 1. elpusztult      | A) nem lát                    |
| 2. magányos        | B) nincs biztonságban         |
| 3. mindnyájan      | C) kar, láb                   |
| 4. ritka           | D) kevés van belőle           |
| 5. vak             | E) mindenki                   |
| 6. veszélyeztetett | F) nincs barátja, egyedül van |
| 7. végtag          | G) nem él                     |



#### A gepárd a leggyorsabb állat?

Nem. Igaz, hogy földön ő a leggyorsabb, de a levegőben van egy nála is gyorsabb állat. Amíg a gepárd 120 km/órával fut a földön, a vándorsólyom közel 380 km/órával repül az égen. A világ leggyorsabb embere csak 44,72 km/órával tud futni.



#### Melyik a lelassabb állat a világon?

A lajhár a lelassabb szárazföldi állat, de a tengerben élnek állatok, akik még a lajhárnál is lassabbak. Például a tengeri rózsák és a korallok.

#### Melyik a legcsúnyább állat a világon?



A csupasz földikutyanak nincs bundája, csak néhány szőrszál van a testén. Nincs füle, és a szeme sem jó. A véznaújju maki csúnyább, mint a földikutya. Hosszú orra és csúnya szemei vannak. A gyerek makik is ijesztőek. Sok ember szerint a világ legcsúnyább állata a pacahal (angolul blobfish). A pacahal az óceán mélyén él.

#### Melyik a legritkább állat a földön?

A Magányos George nevű óriásteknős teljesen egyedül élt a Santa Cruz-szigeten, de 2012 júniusában sajnos elpusztult. George volt az egyik utolsó galapagos óriásteknős. A tudósok szerint a teknős, a hiúz, a rinocérosz és a tigris a legveszélyeztetettebb állat a Földön.



90. ábra Milyen lehet egy kutya? (A2)

NÉZD EZT A SOK ÁLLATOT!

ÉN IS SZERETNÉK EGY KUTYÁT!

1. Írd be a szavakat a megfelelő helyre! Milyen lehet egy kutya?

*hosszú szőrű, nagyobb testű, ugrálás, barna szőrű, bolondos, juhászkutya, hiperaktív, lusta, ragaszkodó, yorkshire terrier, kisebb testű, fekete szőrű, rövid szőrű, kíváncsi, vadászkutya, golden retriever, falánk (sokat eszik), hűséges, okos*

| Külső tulajdonság    | Belső tulajdonság | Fajta               |
|----------------------|-------------------|---------------------|
| <i>hosszú szőrű,</i> | <i>bolondos,</i>  | <i>juhászkutya,</i> |
|                      |                   |                     |

91. ábra Milyen lehet egy háziállat? (B1)

4

ÉN IS SZERETNÉK EGY HÁZIÁLLATOT!

1. Írd be a szavakat a megfelelő helyre! Milyen lehet egy háziállat?

*csendes, hűséges, tollas, feltűnő, kameleon, szelíd, tengerimalac, barátságos, tiszta, hörcsög, magányos, szőrös, társas, teknős, játékos, kígyó, pikkelyes, mozgékony, görény, intelligens*



| Külső tulajdonság | Belső tulajdonság | Faj          |
|-------------------|-------------------|--------------|
| <i>tiszta</i>     | <i>játékos</i>    | <i>kígyó</i> |
|                   |                   |              |

## 2.2 A tanított nyelvi formák arányának és sorrendjének kialakítása, egy szintezett lexikogrammatikai tanmenet elkészítése

A *Magyar felfedező* lexikogrammatikai tanmenetének elkészítésekor szintén a *MagyarOK* szerzőihez hasonlóan jártunk el (vö. Szita – Pelcz, 2020; Szita, 2014), egyrészt a gyakoriság mellett a nyelvi formák hasznosságát és funkcióját is figyelembe véve, másrészt pedig a témához és az ahhoz tartozó lexikához hozzárendelve készítettük el a tanmenet alapjául szolgáló mátrixot. A teljes lexikogrammatikai tanmenet a tananyag elkészülése után elérhető lesz, ebben a részben csak a már elkészült 4. fejezethez készült részét ismertetem röviden a következő táblázatban összefoglalva (58. táblázat).

58. táblázat A Magyar felfedező 4. fejezetének lexikogrammatikai tanmenetének részlete

| Nyelvi szint | Téma  | Nyelvi forma   |
|--------------|---|--|
| A1           | Kedvenc állatok                                 | Főnevek többes száma;<br>léte: van/ vannak, nincs/ nincsenek;<br>főnevek többes száma és a tárgyrag  |
| A2           | Különleges állatok                              | Melléknév fokozás (közép és felsőfok)  |
| B1           | Ideális háziállat; Milyen állatot nem tartanál? | Feltételes mód (határozott és határozatlan ragozás)  |
| B2           | Állatok veszélyben                              | Mennyiségjelzős alárendelő összetett mondatok;<br>fok-mérték határozós alárendelő összetett mondatok   |
| C1           | Az állatok jogai                                | A felszólító mód funkciói, a felszólító mód használata az az összetett mondatokban (ismétlés) (beszédszándék: tiltás, megengedés, lehetőség adása) |

## 2.3 A témákhoz kapcsolódó lexika, válaszvariációk és szövegtípus kapcsolatának beépítése a tananyagba

A modellalapú nyelvoktatás módszere, amelyről már röviden olvashattunk a II.6.5.2 fejezetben, jól kapcsolódik a korpuszinformált tananyagoknak a felhasználásához, illetve a *Magyar felfedező* által is képviselt nyelvoktatási szemlélethez. A módszer arra a megfigyelésre épül, amelyről már többször is szó esett a jelen doktori értekezésben (ld. például II.4 fejezet), és amely szerint a kompetens nyelvhasználók megnyilatkozásaikban nagy arányban használnak előre felépített egységeket, amelyekre modellként támaszkodnak (Szita – Pelcz, 2017: 263; vö. O’Keeffe et al., 2007, Altenberg, 1998; Biber et al., 1999).

A modellalapú tananyagokra jellemző a nagy nyelvi input használata és a spirális szerkezet. A nyelvoktatás során a nyelvi produkciót mindig egy a nyelvi mintákat megfigyelő szakasz előzi meg (Szita – Pelcz, 2017: 264), amelyhez a korpuszinformált tananyagok jó lehetőséget biztosítanak. A szövegekben található többelemű lexikai egységek a modellalapú tananyagokban különböző variációkban spirális szerkezetben, újra és újra előkerülnek, így az ismétlés által sokkal könnyebben tudják megtanulni őket a nyelvtanulók. A módszer számos előnnyel jár a nyelvtanulók, így a gyermekek számára is. A természetes nyelvhasználatot tükröző modell segítségével a gyermekek is könnyebben tudnak megnyilatkozásokat alkotni, beszédük pedig folyékonyabbnak és pontosabbnak is tűnik a modellek használatával.

A *MagyarOK* tankönyvcsaládhoz hasonlóan a *Magyar felfedezőben* is megjelenik a spirális szerkesztés, hiszen az öt nyelvi szintre készülő fejezetek tematikája azonos, ugyanakkor egyre bővülő szókincset nyújt a nyelvtanulóknak, és más-más nyelvi szerkezetek tanítását tűzte ki célul. A teljes tananyag elkészülése után látható lesz, hogy ez a spirális szerkezet az egy nyelvi szintre készült fejezetekben is megtalálható. A tananyag szövegvariációkkal dolgozik, a szövegekben pedig a korpuszból gyakoriság alapján kiválasztott többelemű nyelvi egységek, szókapcsolatok jelennek meg, segítve ezzel, hogy a tanulók könnyebben megjegyezzék őket (vö. Szita – Pelcz, 2017). Jó példa lehet erre a következő, amely az A2-es nyelvi szintre készült mintafejezetben található:

- Az elefánt Afrikában él. A sötét színű állat, **akár 3 méter is lehet**. (67. oldal, 8. feladat)
- Még az elefántnál is nehezebb az óceánban élő bálna. Súlya elérheti a 140 000 kilogrammot, a hossza pedig **akár 30 méter is lehet**. (69. oldal, 3. feladat)

- Egy degu kisebb, mint egy nyúl, de hosszabb a farka. A farka **akár 15 cm hosszú is lehet.** (75.oldal, 5. feladat)
- A következő állat a fóká. Fókák élnek az Északi és a Déli-sarkon is. Kb. 50 centiméter magas és a súlya **akár 20 kilogramm is lehet.** (A2-es videó)

Az állatokról szóló leírásokból származó *akár + (méret) + is lehet* egy részben rögzült lexikai egység, amelynek a felfedezése nemcsak a projekt elkészítése kapcsán járhat nagy haszonnal a nyelvtanuló gyermekek számára, hanem az ilyen lexikai egységek használata által a tanulók nyelvhasználatát természetesebbnek, beszédprodukciónak pedig fluensebbnek érzékeljük (vö. II. 4–5 fejezet). A kiválasztott lexikai egységek több variációban történő megjelenése, bővített szerkezetekben való vissza-visszatérése támogatja a nyelvtanulókat abban, hogy könnyebben megjegyezzék ezeket, észrevegyék a használati mintázataikat, valamint különböző kontextusokban lévő jelentésüket.

A modellként szolgáló mondatokat és kifejezéseket, amelyek nagy részét a KorSzak GyermeKNyelvi Korpuszból választottuk ki gyakoriságuk, hasznosságuk és funkciójuk alapján, a beszédkészség fejlesztését megcélzó feladatokban is hasonlóan alkalmazzuk ahhoz, ahogy azt már korábban láthattuk a *MagyarOK* kapcsán (II.6.5.2 fejezet). Ebben az esetben szintén szinteken átívelve is megjelennek újra és újra, bővített szerkezetek formájában a kiválasztott lexikai egységek, valamint a magasabb nyelvi szinteken ismétlésként is szerepelnek. Az alábbi két példa (92. és 93. ábra; 94. és 95. ábra) ezt jól mutatja.

### 92. ábra Vélemény kifejtése (B2)

**8.** **Képzeljétek el,** hogy egy állatkert vezetőségének tagjai vagytok. Ideje lenne több ketrecet, kifutót is felújítani. Az állatkertnek rengeteg bevételt hozhatna egy fókashow, lenne pénz felújításra. De vajon a fókáknak is örömet okoz a szereplés? Megéri így pénzhez jutni? **Vitassátok meg,** hogy bevezényeltek-e az állatkert programjába a bemutatót!

- a) Alkossatok két csapatot! Az egyik csapat a fókashow bevezetése mellett, a másik a bevezetés ellen gyűjtsön érveket!
- b) Vitassátok meg a kérdést! Segítségképpen használjátok a táblázatban szereplő fordulatokat!

| Vélemény kifejtése                         | Ellenvélemény                           | Egyetértés kifejezése                  |
|--|---|--|
| <i>Én úgy gondolom ...</i>                 | <i>Szerintem ebben nincs igazad.</i>    | <i>Szerintem is fontos a/az ...</i>    |
| <i>(Nekem) az a véleményem, hogy ...</i>   | <i>Az igaz, hogy ..., de ...</i>        | <i>Ebben igazad van.</i>               |
| <i>Szerintem ...</i>                       | <i>Igen, de nem gondolod, hogy ...?</i> | <i>Teljesen egyetértek.</i>            |
| <i>Meg vagyok győződve arról, hogy ...</i> | <i>Én ezt másképp látom.</i>            | <i>Én is úgy gondolom, hogy ...</i>    |
| <i>Biztos vagyok benne, hogy ...</i>       | <i>Szerintem inkább ...</i>             | <i>Én is ezen a véleményen vagyok.</i> |

### 93. ábra Vélemény kifejtése, információ kérése, válaszadás késleltetése (C1)

6. Képzeljétek el, hogy titeket kért fel az iskolaújság egy állatokról szóló interjúorozat elkészítésére. Párokban, kis csoportokban **írjatok össze** olyan kérdéseket, amiket szívesen feltennétek az iskolatársaitoknak, barátaitoknak, rokonaitoknak.

| Információ kérése   | A válaszadás késleltetése  | Vélemény kifejtése  |
|---|--|---|
| <ul style="list-style-type: none"> <li>Válaszolnál/Válaszolna néhány kérdésre?</li> <li>Mit gondolsz/gondol ...ról/ről?</li> <li>Szeretném megtudni, hogy ...</li> <li>Arról érdeklődnék, hogy ...</li> <li>Meg tudnád/tudná mondani, hogy ...?</li> <li>Mi a véleményed/véleménye a/az ...ról/ről?</li> <li>Azon gondolkodtam, hogy ...</li> </ul> | <ul style="list-style-type: none"> <li>Hát ...</li> <li>Hadd gondolkozzak ...</li> <li>Ez nehéz kérdés ...</li> <li>Ez jó kérdés ...</li> <li>Hát, attól függ, hogy ...</li> <li>Hogy érted/érti, hogy ...</li> <li>Még sosem gondolkodtam ezen, de ...</li> </ul> | <ul style="list-style-type: none"> <li>Szerintem ...</li> <li>Én úgy gondolom, hogy ...</li> <li>Biztos vagyok benne, hogy ...</li> <li>Azt gondolom, hogy ...</li> <li>(Nekem) az a véleményem, hogy ...</li> <li>Azt hiszem, hogy ...</li> <li>Meg vagyok győződve arról, hogy ...</li> </ul> |

### 94. ábra Érvelés és meggyőzés (A2)

3. Ákos sokat mesél az új kutyájáról. Noel is nagyon vágyik egy kutyára, de az apukája azt gondolja, hogy néhány hal jobb választás lenne.

a) Gyűjtsetek ötleteket, amivel Noel meggyőzheti az apukáját!

Milyen egy...

kutya?

barátságos,



hal?



Egy kutya barátságosabb, mint egy hal.



b) Olvassátok el a táblázatban szereplő kifejezéseket! Válasszatok néhányat, ami szerintetek hasznos, ha valakit meg akarunk győzni valamiről.

| Kérés                       | Érv               | Ígélet                        |
|-----------------------------|-------------------|-------------------------------|
| Nagyon szeretném, (hogy)... | Biztos, hogy...   | Megígérem, hogy...            |
| Kérlek szépen!              | Hidd el, hogy ... | Ha ..., akkor...              |
| Légy szíves!                | Raadásul...       | Biztos lehetsz benne, hogy... |
| A kedvemért!                | Ezenkívül...      |                               |

## 95. ábra Érvelés és meggyőzés (B1)

5. Ákos sokat mesél az új kutyájáról. Noel nagyon vágyik egy háziállatra. Az apukája szerint néhány hal lenne az ideális választás, de Noel valami különlegesebbre vágyik.

- a) Válasszatok egy különleges háziállatot, és gyűjtsetek ötleteket, amivel Noel meggyőzhetné az apukáját!

Az állat, amit választottunk: \_\_\_\_\_

Érvek:

*(Pl. Ha lenne egy kecskénk, nem kellene füvet nyírni.)*



- b) Olvassátok el a táblázatban szereplő kifejezéseket! Válasszatok néhányat, ami szerintetek hasznos, ha valakit meg akarunk győzni valamiről.

| Kérés                          | Érv                                      | Ígéret                     |
|--------------------------------|--|----------------------------|
| <i>Kérlek szépen!</i>          | <i>Biztos, hogy ...</i>                  | <i>Megígérem, hogy ...</i> |
| <i>Légy szíves!</i>            | <i>Hidd el, hogy ...</i>                 | <i>Ha ..., akkor ...</i>   |
| <i>Úgy örülnék neki.</i>       | <i>Hát, az fontos szempont, hogy ...</i> |                            |
| <i>Annyira/Olyan jó lenne.</i> | <i>Te is biztos szívesen ...</i>         |                            |
| <i>A kedvemért!</i>            | <i>Neked is tetszene / jót tenne.</i>    |                            |
| <i>Gondold át!</i>             | <i>Ráadásul ...</i>                      |                            |
|                                | <i>Ezenkívül ...</i>                     |                            |

## 2.4 Valós és a tanulók számára releváns szituációk beépítése a tananyagba

A *Magyar felfedezőbe* kerülő kommunikációs helyzetek kiválasztásánál is törekszünk olyanokat meghatározni, amelyek relevánsak lehetnek a magyarul tanuló gyermekek számára. Azokat a témákat és kommunikációs helyzeteket választottuk ki, amelyek az anyanyelvi beszélők, jelen esetben a gyermekek mindennapi életében gyakran előfordulnak, így a nyelvtanulók számára is fontosak lehetnek (vö. II.6.5.2.4). A szituációk kiválasztásánál is elsősorban a KorSzak Gyermeznyelvi Korpuszban lévő gyermekek beszélgetéseire alapozunk, illetve a saját tanítási tapasztalatunkra. A 4. fejezetben ilyen szituáció például a kedvenc állatokról való beszélgetés (A1) vagy a saját háziállat bemutatása (A2), valamint idetartozik az elveszett állatról való beszélgetés és a plakátkészítés is (96. ábra).

## 96. ábra Beszélgetés egy elveszett állatról és plakát készítése (A1)

3. Alex eltűnt. **Olvasd el** a plakátot, amit Dani készített!

**KERESSÜK ALEXET!**



Név: Alex  
Kor: 1 éves  
Fajta: rottweiler  
Nem: szuka  
Szín: fekete, barna

Utoljára látták:  
2020. április 12.

Itt látták:  
Budapest, 5. kerület

**50 000 forint jutalom a megtalálónak!**

+361235745 ✕  
+361235745 ✕  
+361235745 ✕  
+361235745 ✕  
+361235745 ✕  
+361235745 ✕  
+361235745 ✕

4. Kérdezd a párod!

- Mi a kutya neve?  
- Alex.

Mikor...?

Mennyi...?

Mi...?

Hol...?

Hány éves...?

Milyen színű...?

Milyen...?

5. Hibakereső. A következő szövegben 7 helyesírási hiba van. **Találd meg** a hibákat, és **másold le** a mondatokat helyesen!

Ő Alex, a kutjám. Barna színű, és nagyon aranos. Sokat játszunk a kertben. A kutja a kedvenc álatom.

(Dani, 13 éves)

*Ő Alex, a kutjám.*

---

---

---

---

---

6. **Készíts** a 3.feladatban lévőhöz hasonló plakátot!

## 2.5 Saját intuíciók helyett a korpuszadatokra támaszkodva készített, a természetes nyelvhasználatot tükröző szövegek beépítése a tananyagba

A tananyag beszéd-készség-fejlesztésre fókuszáló részeiben a KorSzak Gyermekeknyelvi Korpuszban elhangzó szövegek jelennek meg, az adott nyelvi szinthez igazított változatban. A nyelvi szinthez való igazodás során a cél az volt, hogy a szövegek továbbra is tartalmazzák a gyakori többemlékű lexikai egységeket, amelyek a korpuszalapú kutatás során kiválasztásra kerültek, illetve ezeknek a szövegekben megjelenő használati mintázatuk is jelenjen meg, hogy ezáltal továbbra is a természetes nyelvhasználatot tükrözzék. Ennek eredménye igazán csak akkor lesz látható, amikor már a teljes tananyag elkészül, de a kutatás eredményeit bemutató fejezetben (IV. fejezet) kiválasztott lexikai egységek közül néhány használata már az elkészült fejezetekben is megfigyelhető.

Az *ugye* használatának például két funkciója jelenik meg az elkészült fejezetekben. Időkitöltő elemként jelenik meg például a 4. fejezetben:

- *Nekünk állatkertünk van. Van **ugye** Csomó, ő egy kutya, van két macska és 14 hal, azt hiszem* (4. fejezet, B1).

Valamint megerősítés kérése funkcióban az 1. fejezetben:

**Kitti:** Vízilabda, **ugye**?

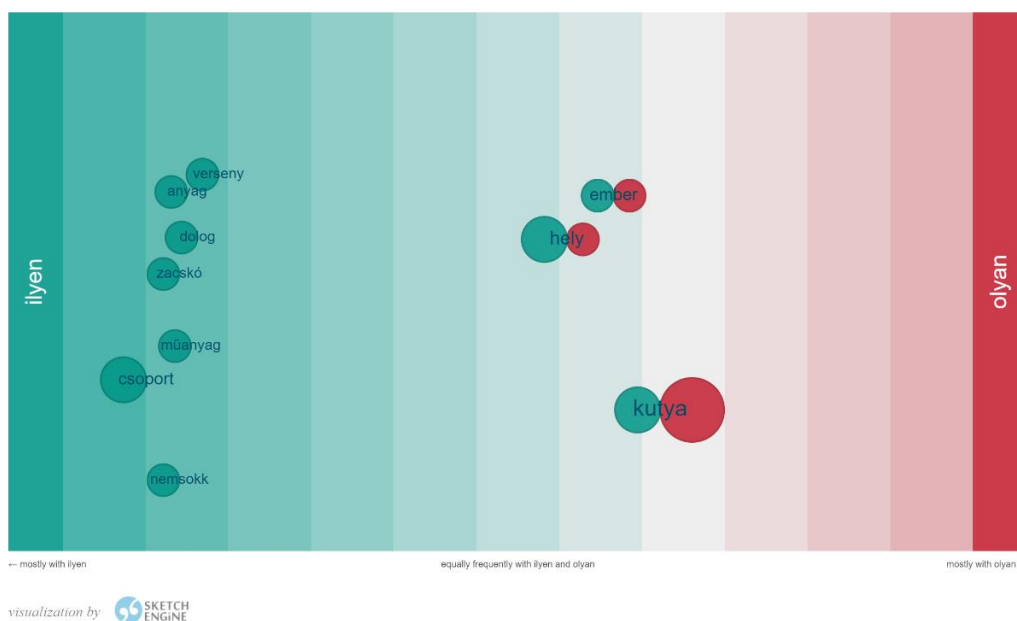
**Benji:** Igen.

A *nem tudom* kételemű lexikai egység szintén többször szerepel a *Magyar felfedező* szövegeiben, de egyelőre csak egy funkciójában, a tudás hiányának kifejezésésként. Például:

- Hát **nem tudom**, hogy a néptánc mennyire sport, de nekem a néptánc a hobbim (1.fejezet, A1).

Az *ilyen* és az *olyan* használati mintázatai még szintén nem látszanak teljes mértékben, de már megjelennek a tananyag elkészült fejezeteiben azok a mintázatok, amelyeket a korpuszalapú kutatás során is láthattunk. Ilyen például az *ember* főnév, amely mindkettő társaságában megjelenik, ugyanakkor a *dolog* egyelőre még csak az *ilyen* mellett szerepel (97. ábra).

**97. ábra** Az ILYEN és az OLYAN mellett megjelenő leggyakoribb főnevek a *Magyar felfedező* szövegeiben





A szövegekre visszatérve fent említésre került, hogy a tananyag beszédképesség-fejlesztő részeibe a KorSzak GyermeKNyelvi Korpuszból származó szövegeket válogattunk be. Az alacsonyabb nyelvi szintek esetében ezeket az autentikus szövegeket átalakítottuk, illetve rövidítettük, de törekedtünk a természetes nyelvhasználat megőrzésére. Ez látható, ha összevetjük a 31. példát a 98. ábrán látható szöveggel.

(doc#0 Mon\_Állatok\_001: A gyermekek a kedvenc állatokról beszélgetnek)

- (31) A: **Te szereted az állatokat?**  
B: **Igen. Nagyon szeretem őket.**  
A: **És minden állatot, vagy csak a vadon élő, vagy a háziállatokat inkább?**  
B: **Mindegyiket, mindegyik ugyanolyan fontos, és mindegyik ugyanolyan aranyos.**  
A: **Nekem is hasonló a véleményem.**  
[...] **És neked mi a kedvenc állatod?**  
B: **Hát én nagyon szeretem a medvéket, és te?**  
A: **Én ugye minden állatot szeretek, de a rókát, meg az őzet különösen szeretem, nem sűtve, de szeretem.**

98. ábra Szereted az állatokat? (A1)

**SZERETED AZ ÁLLATOKAT?**

**1. Olvasd el és hallgasd meg a párbeszédet!**

**NIKI:** Te szereted az állatokat?  
**ANDI:** Igen. Nagyon szeretem őket.  
**NIKI:** És minden állatot, vagy csak a vadon élőket, vagy a háziállatokat inkább?  
**ANDI:** Mindegyiket. Mindegyik ugyanolyan fontos, és mindegyik ugyanolyan aranyos.  
**NIKI:** Nekem is hasonló a véleményem. És neked mi a kedvenc állatod?  
**ANDI:** Hát, én nagyon szeretem a medvéket. És te?  
**NIKI:** Én minden állatot szeretek, de a rókát, meg az őzet különösen szeretem.

**NÉZD EZT A SOK ÁLLATOT!**



Visszaulva az *ugye* használatára a fent bemutatott A1-es szövegben látható, hogy ezen a szinten még kitöröltük a didaktizált szövegből. Ez azért történt, mert A1-es szinten még túl bonyolultnak ítéltük meg az eredeti szövegben szereplő funkcióját, vagyis amikor a beszélő saját maga iránti bizalmát fejezi ki.

Egy másik példa a szövegek nyelvi szinthez való igazítására a következő 32. példa, amelyben ismét az autentikus beszélgetés leírata látható és a 99. ábrán megjelenő szöveg, amely a tananyagban megtalálható változatát mutatja.

(doc#7 Mon\_Állatok\_Kutya\_004: A gyermekek a kedvenc kutyafajtáikról beszélgetnek)

(32) A: Szóval hogyha kezdhetném én, akkor nekem az első kérdésem az lenne, hogy **mi a kedvenc kutyafajtátok.**

B: **Én szeretem a kisebb testű kutyákat is, meg a nagyobb testűeket is. Igazából amit a legjobban szeretek, azok a kisebb kutyák,** mivel valahogy én **a nagyobb kutyáktól** egy kicsit **jobban tartok,** és én **szeretem a yorkikat,** amik így kisebbek. **Nagyobb testűből,** meg hogyha kéne abból is egyet választani, akkor abból **inkább a labradort** választanám, viszont **a kedvencem a yorki.**

C: **Én meg mindenféle kutyát szeretek, de mégis a kedvencem az inkább a border collie.**

A: **Nekem a kedvenc kutyafajtám, az egyik a golden retriever, a másik pedig a bichon, mert egyrészt olyan kutyám is van, és az egyik ugye nagy testű, a másik pedig kicsi kutya, de én kettőt nagyon szeretem.**

## 99. ábra Nektek mi a kedvenc kutyafajtátok? (A2)

### 2. Olvasd el és hallgasd meg a beszélgetést!



**BRIGI:** Szia, Ákos! Mi van?

**ÁKOS:** A szüleim végre megengedik, hogy legyen kutyám.

**NOEL:** De jó neked!

**ÁKOS:** De nem tudom, milyen legyen. Nektek mi a kedvenc kutyafajtátok?

**HANNA:** Én szeretem a kisebb testű kutyákat is, meg a nagyobb testűeket is. Igazából amit a legjobban szeretek, azok a kisebb kutyák, mert a nagyobb kutyáktól jobban tartok. Szeretem a yorkikat. Nagyobb testűekből pedig inkább a labradort, de a kedvencem a yorki.

**NOEL:** Én mindenféle kutyát szeretek, de a kedvencem a border collie.

**BRIGI:** Nekem az egyik kedvenc kutyafajtám a golden retriever, a másik pedig a bishon, mert olyan kutyám van. Az egyik nagyobb testű, a másik meg kicsi.

A magasabb nyelvi szinteken vagy nem hajtottunk végre semmiféle módosítást és az eredeti szöveget illesztettük be a tananyagba, vagy csak minimális változtatásra került sor. Például a szöveg elejét egy-egy bevezető kérdéssel igazítottuk a szöveg többi részéhez, hogy érthető legyen a beszélgetés kontextusa. Ez látható a 100. ábrán.

## 100. ábra És milyen volt az állatkert? (B2)

### 2. Olvasd el és hallgasd meg a beszélgetést!

**NIKI:** És milyen volt az állatkert?

**ANDI:** Jó volt, de tudod, én annyira nem támogatom az ilyen helyeket, hogy legszívesebben nem is mennék. Csak a családom miatt mentem.

**NIKI:** Mert be vannak zárva az állatok?

**ANDI:** Hát igen. Van egy csomó ilyen csoport az interneten, akik nem támogatják az állatkerteket, sőt tüntetnek is ellenük. Az állatok be vannak zárva egész életükben, és csak néznek ki a rácsokon. Annyira kicsi helyük van, hogy alig tudnak mozogni, és nem lehetnek kint a természetben. Lehet, hogy gyerek-szemmel az állatkert jó móka, de szerintem így annyira nem.

**NIKI:** Én is ezen a véleményen vagyok. Meg az sem mindegy, hogy például ott vannak a fókák, akiken látszik, hogy örömet okoz nekik, ha munkát kapnak, ha van valami feladatuk. De például az szerintem nem helyes, hogyha a jegesmedve, aki a természetes környezetében naponta tíz... vagyis több száz kilométert tesz meg úszva vagy esetleg a jégábránál mászkálva, be van zárva egy picike helyre.



### 3. Válaszolj a kérdésekre!

1. Miért nem szereti Andi az állatkerteket?
2. Nikinek mi a véleménye az állatkertekről?
3. Milyen állatokat említ Niki?
4. Milyen állatnak lehet jó az állatkert?

A tananyag készítése során természetesen nemcsak a KorSzak Gyermeknyelvi Korpusz beszélt nyelvi szövegeire támaszkodunk, hiszen egy ilyen jellegű tananyag célja az is, hogy a magyarul tanuló gyermekeknek a tanulmányaik során használható tudományos szókincse, szóbeli és írásbeli kifejezésmódja is fejlődjön. Ennek a célnak az elérése érdekében használt forrásokról talán egy tanulmányban érdemes lenne majd beszámolni, most azonban erre nem térnék ki részletesebben.

### 3. Néhány gondolat a projektalapú nyelvoktatásról

A projektalapú nyelvoktatás egy olyan tanulóközpontú módszer (Fried-Booth, 1997; Bilsborough, é.n.), amely a 11–14 éves korosztállyal hatékonyan alkalmazható. Nagy előnye, hogy vegyes nyelvi szinten lévő csoportokban is jól használható. Mivel a magyar nyelvet tanuló gyermekcsoportok is – a kis létszám miatt – sokszor nyelvi szintjük szerint heterogének, valamint a projektalapú nyelvoktatás ismert hatékonysága miatt (vö. Bilsborough, é.n.), ezt a nyelvoktatási módszert találták a szerzők a legmegfelelőbbnek. A vegyes nyelvi szintű csoportok támogatása az egyik fő célja a projektnek, ezért erről még a későbbiekben, a tananyag szerkezetére kapcsán is szó esik (Vermeki, 2021b).

Ennek az oktatási módszernek, a projektmódszernek a megjelenése egészen az 1800-as évek végéig nyúlik vissza és John Dewey nevéhez köthető, aki a *Pedagógiai hitvallásom* című művében beszélt a „cselekvés általi tanulás” [learning by doing] fontosságáról. A projektalapú oktatásnak számos meghatározása létezik (ld. Chard – Katz – Kogan, 2014; Capraro, 2013; Krauss – Boss, 2013), amelyben azonban mindegyik megegyezik az az, hogy a projekteknek bizonyos szempontoknak meg kell felelniük. Ezek a következők a *Magyar felfedező* mintafejezetében található gyakorlati példákkal kiegészítve:

1. A projekt egy kihívás, érdekes probléma vagy figyelemfelkeltő kérdés felvetésével kezdődik. A *Magyar felfedező* 4. fejezetében egy valós problémára keresik a tanulók a megoldást. A városi állatkert tulajdonosa szeretné, hogy ne csak kisgyerekek és felnőttek látogassák az állatkertet, hanem a fiatalok is. Megkéri ezért a tanulókat arra, hogy készítsenek egy olyan reklámvideót az állatkert számára, amely a kortársaik számára vonzó lehet, amelynek a segítségével az állatkertbe több fiatal tudnának becsalogatni.
2. A végeredménye egy termék, amelyet nyilvánosságra hozunk, kiviszünk az osztályból vagy vendégeket hívünk, hogy megtekintsék. A *Magyar felfedező* tanulói gyermekek az elkészült állatkert reklámozó videót feltölthetik a magyarfelfedezo.hu című weboldalra, de be is mutathatják a kortársaiknak, barátaiknak, vagy a családjuknak.
3. A projekt elkészülése közben a gyermekek nemcsak a tananyagot sajátítják el, hanem különböző készségeket, képességeket is. Például kommunikáció, kollaboráció, kritikus gondolkodás, problémamegoldás, kooperáció, időbeosztás és a többi. A 4. fejezetben is fontos a kollaboráció és a kooperáció, hiszen a csapatokban dolgozó gyermekeknek

különböző feladataik vannak: ki kell találniuk a forgatókönyvet, beszélniük kell az állatkertről, valakinek fel kell vennie a videót, valakinek pedig meg kell vágnia azt.

4. Fontos az érdeklődés fenntartása. A projektet több tanórán keresztül készítik a gyermekek, közben érdekes feladatokat, önálló kutatást is végeznek. Az érdeklődés fenntartását a *Magyar felfedező* szerzői a tananyagban található úgynevezett *kis projektekkel* próbálják meg elérni. Ezek a *kis projektek* szerves részeit képezik a fő projektnek, például egy felmérést kell a gyermekeknek készíteniük vagy éppen az állatkert térképét kell lerajzolniuk, vagy megtervezniük.
5. Autentikusság:
  - a. Egy valódi probléma, amely hatással van a gyermekek életére vagy számukra releváns téma. Egy olyan termék létrehozása, amely segíti a tanulók későbbi életét és munkáját. Például megtanulnak csapatban dolgozni, vagy megtanulják egy számítógépes alkalmazás használatát. A mintafejezetben szereplő állatkert tulajdonosának problémája, miszerint kevés fiatal jár hozzájuk, egy valós probléma, de a 4. fejezet projektje nem csak ez által válik autentikussá, hanem azáltal is, hogy elkészítése közben a gyermekeknek különböző infokommunikációs eszközöket és számítógépes programokat is meg kell tanulniuk használni, amely a későbbi életük során sok haszonnal járhat. A fejezetekhez kiválasztott témák azért mondhatóak ennek a korosztály számára relevánsnak, mert kiválasztásuk során azokat a témákat vették figyelembe a szerzők, amelyeket a KorSzak Gyermeeknyelvi Korpusz építése előtt a felmérésben résztvevő, az általános iskola felső tagozatába járó gyermekek meghatároztak (ld. III.1.2.1 fejezet).
  - b. Autentikus, valódi, jelentéssel teli nyelvhasználat, a nyelvet eszközként használják a projekt elkészítéséhez. A projekthez, valamint az elkészítése során történő kommunikációhoz a gyerekek használhatják a tananyagban található mintamondatokat, a kifejezésbankot, amely a kiegészítő anyagok között található, valamint a weboldalon lévő mintavideókat is. A videóban velük egykorú gyermekek és egy zoopedagógus beszél a Pécsi Állatkertről.
6. A tanulónak van választási lehetősége, önállósága a projekt elkészítése során. Ez nem azt jelenti, hogy a tanár passzív, hanem pont az ellenkezőjét: aktívan támogatja a gyerekek munkáját, partnerré válik, együtt tanul velük. A 4. fejezet projektjéhez a gyermekek maguk választhatnak állatkertet, ez akár egy kitalált állatkert is lehet,

valamint arról is ők döntenek, hogy milyen állatokat mutatnak be, illetve hogyan szerkesztik meg a videót.

7. A tanulók kaphatnak időt az önreflexióra.
8. Értékelhetik egymás munkáját. Nem csak a tanár ad visszajelzést, hanem a gyerekek is egymásnak. Ez azért is fontos, mert ennek a korosztálynak már nem a tanár véleménye a legfontosabb, hanem a társaiké. Ezáltal egymást motiválják a jobb teljesítményre.
9. Lehetőséget kapnak a visszajelzések alapján történő módosításra. A *Magyar felfedező* egy fejezetére körülbelül tíz tanórát szánnak a szerzők, de javasolják a feldolgozásra szánt idő csoporthoz való igazítását. Így a pedagógusoknak van lehetőségük a visszajelzésekre és a javításra szánt időt is a tanmenetükbe illeszteni (Vermekei, 2021b).

A *Magyar felfedező* szerzőinek meghatározása szerint a projekt egy hiányosan definiált feladat, amelynek a kimeneti követelményei jól meghatározottak (vö. Capraro, 2013: 4), releváns a gyermekek számára, a termék elkészítése során pedig valódi nyelvhasználaton keresztül 21. századi képességeket és készségeket is elsajátítanak (vö. Pieratt, 2019).

A *Magyar felfedező* tehát egy projektalapú tananyag, amelyet kétféleképpen is használhatunk. Egyrészt valódi projektalapú [project-based] tananyagként a végső termékre koncentrálva, amelyet a tankönyvi feladatokkal előkészíthetünk és kiegészítünk, másrészt pedig projekttel támogatott [project supported] tananyagként is, amely során a tananyag kerül a középpontba és a projektet, mint a megtanulandó nyelvi mintázatok, kifejezések gyakorlására szolgáló eszközt használjuk. A szerzők természetesen a *Magyar felfedező* projektalapú tananyagként való feldolgozását ajánlják a megfelelő hatékonyság elérése céljából, de fontosnak tartják, hogy a készülő tananyaggal a más módszerrel dolgozni kívánó pedagógusok munkáját is támogassák.

Végül lényeges megjegyeznünk, hogy a tananyag egyik különlegessége a projektalapúság, mert nem sok olyan nyelvkönyv létezik más, szélesebb körökben tanult nyelvekből sem (például angol, német), amely projektalapúnak mondhatja magát.

## 4. A tananyag szerkezete

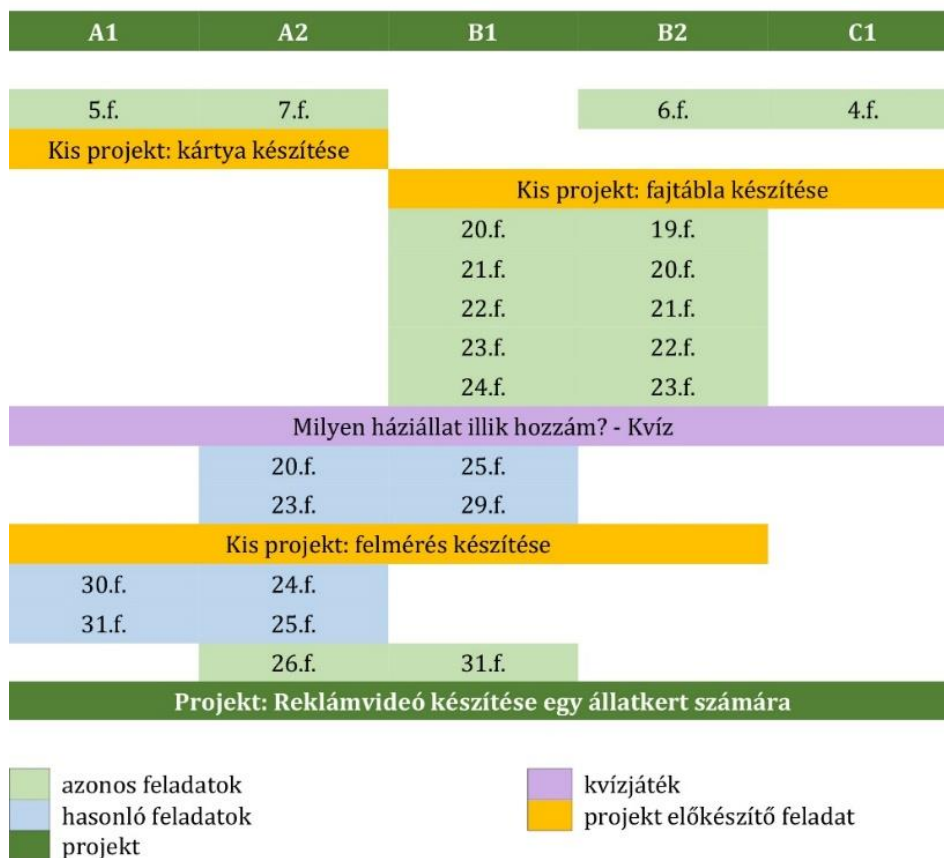
A *Magyar felfedező* egy fejezetét összesen körülbelül tíz negyvenötperces órán lehet feldolgozni, amelyből a projektet előkészítő tananyag hat, a kiegészítő magyar kultúrával kapcsolatos anyag egy, a projekt elkészítése és bemutatása három tanórát vesz igénybe. Természetesen a csoportok képességeihez és lehetőségeihez alkalmazkodva kevesebb vagy több időt is adhatunk a gyermekeknek. Minden fejezet öt nyelvi szinten készül (a KER szerinti A1–C1) és hasonló szerkezettel rendelkezik, így könnyebbé válik a differenciálás a vegyes nyelvi szintű csoportokban (Vermeki, 2021a).

101. ábra A *Magyar felfedező* 4. fejezetében lévő kapcsolódó feladatok összefoglaló ábrája



### Kapcsolódó feladatok

#### A 4.fejezet feladatai:





A fejezetek tartalmazzák szókincs-, hallott szövegértést, olvasott szövegértést, beszédképesség- és írásképesség-fejlesztést megcélzó feladatokat. A tananyag tanórákra bontott, minden tanórán mind a négy készség (beszédképesség, beszédértés, írásképesség, olvasás) fejlesztése zajlik, de egy-egy óra egy vagy két készségre fokozottan koncentrálnak. Például az első órán általában a beszédértés kerül a középpontba, míg a másodikon az olvasott szövegértés. Minden fejezethez tartozik egy a témával kapcsolatos kiegészítő anyag a magyar kultúrához kapcsolódóan.

A tananyag fejezeteinek témái minden nyelvi szinten azonosak, de ezen kívül vannak a nyelvi szinteken átívelő, kapcsolódó feladatok (87. ábra) is bennük, amelyek lehetőséget biztosítanak arra, hogy a vegyes nyelvi szinten lévő gyerekekből álló csoportok időnként együtt is dolgozhassanak.

A *Magyar felfedező*ben lévő kapcsolódó feladatokon kívül, a projekt is azonos az összes nyelvi szintre készült fejezetben, a mintafejezetben például egy állatkertről szóló promóciós videó elkészítése a feladatunk. A tananyagban lévő szókincs és a tanult nyelvi formák biztos alapot nyújtanak a projekt megvalósításához, de ezeket rugalmasan kell kezelniük, mert a projekt elkészítése során felmerülhetnek más igények is. Például olyan állatok nevei, amelyek nincsenek a szókincsfejlesztő részben. Ilyen esetben a pedagógusok nyugodtan segíthetnek a gyerekeknek, kiegészíthetik az anyagot (Vermeki, 2021ab).

**59. táblázat** A projekt elkészítésének lépései

| 1. lépés  | 2. lépés                                 | 3. lépés  | 4. lépés  | 5. lépés  |
|---|--|---|---|---|
| A projekt bemutatása, céljainak ismertetése, beszélgetés a termék bemutatásainak lehetőségéről. | Csoportok alkotása, ötletelés, tervezés. | Adatgyűjtés, újra tervezés, amennyiben szükséges. | A termék elkészítése, a bemutató szervezése és a bemutató gyakorlása. | Bemutató, egymás termékének megtekintése, külső érdeklődők bevonása, értékelések és visszajelzések. |

A projektfeladat elkészítését több kisebb lépésre bontottuk, az úgynevezett kis projektek fontos részét képezik a fejezeteknek, hiszen egyrészt időnként újra ráirányítják a figyelmünket a fő projektünkre, másrészt elő is készítik azt. Ilyen kis projektfeladat például A1-es nyelvi szinten egy felmérés készítése a csoportban, ahol a csoport tagjai által kedvelt állatok neveit

gyűjtjük össze, vagy C1-es nyelvi szinten az álcázó állatok jobb megismerése. A projektfeladatot mindig az adott nyelvi szinthez igazodva támogatja is a tananyag: szókincset, hasznos kifejezéseket, illetve mintaszövegeket és mintavideókat is tartalmaz.

A *Magyar felfedezőhöz* tartozó weboldalon a mintafejezethez készült, több mint kétszáz interaktív feladat található, amely az órán vagy az órák után, otthon is használható. Ezenkívül ott található meg a tananyag hanganyaga, a hozzá készült videók, valamint a pedagógusoknak és szülőknek szóló kiegészítő anyagok. A tanárok felkészülését támogató anyagok például a Tanári kézikönyv, amely óravezetés-javaslatokat és megoldókulcsot is tartalmaz, kiegészítő képek, kártyák és feladatlapok, valamint az órákon is használható szószedet és kifejezésbank. Ezekre példákat a mellékletek között találhatunk.

## VI. Összegzés és lezárás

A jelen doktori értekezésben bemutatott korpuszalapú kutatás célja a KorSzak Gyermeeknyelvi Korpuszban lévő magyar anyanyelvű adatközlő gyermekek természetes nyelvhasználatának megfigyelése és elemzése volt, az eredmények felhasználásával pedig egy korpuszinformált tananyag készítése. A kutatás során a gyermekek spontán beszédében megjelenő a nyelvtanulás szempontjából is hasznos lexikai egységek feltérképezésére és használati mintázatainak megismerésére került sor. Ehhez az elméleti keretet öt alappillérre építettük.

Az első alappillér a használattal alapú nyelvelméleti modellek (vö. Tomasello, 2003, 2013, 2015; Bybee, 2008; Ortega et al., 2016; Ellis et al., 2016; Pelcz, 2023) voltak, amelyek főként a gyermekek nyelvelsajátítási folyamatával foglalkoznak, de lényeges megállapításait a nyelvtanulás során is érdemes figyelembe vennünk. A használattal alapú nyelvelméletek szerint a nyelvi fejlődés azoknak az asszociációnak a következménye, amelyek a nyelv használata közben alakulnak ki. Az ilyen asszociációk az elérhető nyelvhasználati példákból állnak, amelyeket az ismételt használat révén gyorsabban és hatékonyabban érhetünk el a memóriánkból. Ehhez a gyermekeknek – és a nyelvtanulóknak is – biztosítani kell a gyakori találkozások lehetőségét és a komplex nyelvi környezetet (Tomasello, 2013).

A második alappillér a lexikai előfeszítés elmélete (vö. Hoey, 2005; Pace-Sigge, 2013). A Hoey (2005) által kidolgozott elmélet fő gondolata azon a megfigyelésen alapul, miszerint bizonyos szavak és azok formái elősegítik vagy megakadályozzák egymás előfordulását. Az elmélet szerint minden, amit hallunk, olvasunk, leírunk vagy kimondunk megerősíti bennünk, hogyan használjuk a szavakat, illetve, hogyan ne. Amikor találkozunk egy adott szókapcsolattal, az elménk automatikusan asszociációkat kezd kialakítani vele. Az ismételt találkozások során az előzőleg kialakult asszociációk megerősödhetnek vagy gyengülhetnek (Hoey, 2005: 9).

A kutatás harmadik elméleti alappillére a formulaszerű nyelvhasználat (vö. Wray, 2002, 2013) volt. Korpusznyelvészeti kutatások bebizonyították, hogy a kompetens nyelvhasználók beszédprodukciójuk során nagymértékben támaszkodnak a formulaszerű nyelvhasználatra (Wray, 2002; Wray – Fitzpatrick, 2008; Wray, 2019), és nem élnek a grammatika által elméletileg biztosított korlátlan variációs lehetőségekkel (Pawley – Syder, 1983). Az ilyen formulaszerű egységek segítik a könnyebb megértést és a gyors kommunikációt, és ezáltal hozzájárulnak mind a produktív, mind a receptív fluensséghez.

A negyedik alappillér a beszédflowékonysághoz kapcsolódó elméletek voltak. Például

ismertetésre került a Segalowitz-féle hármas modell, a legfontosabb beszédprodukcións modellek (Levelt, 1999; Kormos, 2006), Tavakoli és Hunter piramis modelleje (2018), valamint az automatizálás fogalmának értelmezése. Ezenkívül kitértünk a beszédfolyékonyság nyelvoktatással kapcsolatos szerepére is.

Végül a kutatás ötödik alappillére a korpusznyelvészeti megközelítés (vö. Szirmai, 2005; Stefanowitsch, 2020; Hunston, 2022) volt, amely azon az elgondoláson alapul, hogy a természetes nyelvhasználatban megfigyelhető nyelvi jelenségek elemzéséhez és már magához a megfigyeléshez is elengedhetetlen a nagy mennyiségű nyelvi adatok felhasználásával történő empirikus kutatás.

A kutatást a KorSzak Gyermeknyelvi Korpuszra alapoztuk, amely kifejezetten olyan korpusznyelvészeti kutatások céljából jött létre, amelyek eredményeinek felhasználásával gyermekek számára szóló tananyag készülhet, ezért a korpuszépítés folyamatának bemutatása is fontos részét képezte az értekezésnek. Ennek kapcsán azonban fontos megemlíteni, hogy a korpuszépítés csupán eszközként szolgált a gyermekeknek szóló tananyag elkészítése során.

Lényeges azt is kiemelni, hogy nemcsak a magyar mint idegen nyelv, hanem más nyelvek oktatásában sem jellemző még a gyermekek nyelvhasználatának vizsgálatára alapozott korpuszinformált tananyagok készítése. A KorSzak Gyermeknyelvi Korpuszhoz hasonló gyermekek számára készült pedagógiai korpuszok, amelyeket az értekezésben nyelvoktatásra készült autentikus szövegek halmazaként értelmezzünk, nem állnak még rendelkezésünkre. Bár léteznek korpuszok, amelyeket kifejezetten a gyermekek számára készítettek (mint például a SACODEYL, CorpusMate), a pedagógusok ezekhez vagy egyáltalán nem, vagy csak kevés mennyiségű feladatlapot érhetnek el. A szövegek didaktizálásával létrejött, komplex tananyag nem készült hozzájuk.

A kutatás egy magyarázó szekvenciális tipológia alapján történt (vö. Dörnyei, 2007), amely során a kvantitatív vizsgálatokat kvalitatív elemzések követték. A kutatás több szekvenciából tevődött össze, amelyek alapvetően minden alkalommal vagy az egyelemű lexémákból, vagy a két- és többelemű lexikai egységekből indultak ki.

A kutatás előre meghatározott feltételrendszere alapján az egyelemű lexémák közül az első száz legnagyobb gyakorisági mutatóval rendelkezőt vetettük össze a huTenTen12 korpuszsal. Az eredményekből az látható, hogy a gyermekek nyelvhasználatában megtalálható legnagyobb gyakorisági mutatóval rendelkező egyelemű lexémák és a huTenTen12 korpusz leggyakrabban elemei között van különbség. Az első 20 leggyakoribb elem összevetésekor egyértelmű volt,

hogy a négy legnagyobb eltérést mutató lemma közül az ILYEN 88%-kal, a NAGYON 80%-kal, az AKKOR 75%-kal és a TUD 71%-kal magasabb értéket mutatott, valamint az N-gram korpusszal – amely a KorSzak Gyermeeknyelvi Korpuszban található, a kritériumok alapján kiválasztott lexikai egységekből épült fel – összevetve azt is észrevehettük, hogy ezek mindegyike a fókuszkorpusz többemű lexikai egységeiben beágyazódva is több százszor megjelennek.

A vizsgált egyelemű lexémák közül a 20–100. helyen több olyan elem is található, amely a referenciakorpusz első száz lemmája között egyáltalán nem jelenik meg. Ilyen például a dokumentumok 50%-ban szereplő IGAZÁBÓL, amelynek használati mintázatával bővebben is foglalkoztunk. A gyermekek beszélgetéseiben természetesen megjelentek a beszélt nyelvi fordulatok, mint az UGYE, TÖK, ÚGYHOGY, TÉNYLEG, NA, SZÓVAL, valamint az AMÚGY. A fókuszkorpuszból kiválasztott többemű lexikai egységekben nem szerepel az UGYE, INKÁBB, ÚGYHOGY, TÉNYLEG, NA, HISZ, SZÓVAL, AMÚGY, AZTÁN, valamint a BESZÉL lemma. Ezt azt jelenti, hogy a korpuszban bár megjelenik a többemű lexikai egységek részeként, de nem szignifikáns számban, illetve többségük önmagában alkot egy-egy formulaszerű egységet, amely főként diskurzusjelölőként funkcionál. A kutatás feltételrendszere alapján a felsoroltak közül az ugye felel meg annak a kritériumnak, hogy egyelemű lexémaként a dokumentumok legalább 50%-ában szerepel, ennek nyelvhasználati mintázataival részletesebben is foglalkoztunk.

A vizsgált szófajokat (főnév, melléknév, ige) tekintve elmondható, hogy a főnevek és az igék esetében a fókuszkorpuszban magasabb arányban szerepeltek a mindkét korpuszban megjelenő elemek, a melléknéveknél viszont megállapítható volt, hogy bizonyos elemek jóval alacsonyabb arányban (az új 79%-kal, az utolsó 60%-kal) voltak jelen a fókuszkorpuszban, mint a referenciakorpuszban, valamint a fókuszkorpuszban nagy arányban jelentek meg olyan melléknévek, amelyek a referenciakorpusz első száz eleme között egyáltalán nem szerepeltek (például ARANYOS, BOLDOG, BÜDÖS, CUKI, CSODÁLATOS). Az így kapott, a tananyagkészítés szempontjából leghasznosabb eredményeket beépítettük a tananyag elkészült fejezeteibe. Ilyen például a *Milyen lehet egy kutya?* (A2) vagy a *Milyen lehet egy háziállat?* (B1) című feladatok, amelyekben az előbb említett melléknévek szerepelnek.

Az előre meghatározott feltételrendszer alapján kiválasztott lexikai egységek közül a legnagyobb kategória a háromelemű lexikai egységeké (120 item), majd ezt követően a két- (60 item) és végül a négyelemű lexikai egységeké (20 item) lett. Az összesen 200 darab kiválasztott lexikai egységet közül a tíz legnagyobb gyakorisági mutatóval rendelkező az alábbi felsorolásban látható.

- Kételemű egységek: nem tudom, és akkor, hogy a / az, én is, nem is, az a, egy

ilyen, az ilyen, meg a, és a

- Háromelemű egység: nem is tudom, mit tudom én, vagy nem tudom, vagy hogy mondjam, ott van a, de nem tudom, és akkor így, mi a kedvenc, hát nem tudom, van egy ilyen
- Négyelemű egység: vagy nem is tudom, én is így vagyok, hogy mit tudom én, nagyon jól éreztük magunkat, én is egyet tudok, én úgy vagyok vele, vagy mit tudom én, az volt az első, az nagyon jó volt, de ő is nagyon

A felsoroltak közül néhány lexikai egység használati mintázatával bővebben is foglalkoztunk (például *nem is tudom, mit tudom én, én is így vagyok ezzel / vele* is), valamint ezek a szövegben betöltött szerepük szerint csoportosításra is kerültek (vö. Biber 2006: 139–146). Az így kapott eredményeket szintén felhasználtuk a tananyag elkészült fejezeteiben, bár ez majd csak a teljes tananyag elkészülése után lesz megfelelően látható.

Az értekezés utolsó nagy egységében, az V. fejezetében került sor a Magyar felfedező című tananyag felépítésének és módszertani jellemzőinek példákkal illusztrál bemutatására, amelyben arra is kitértünk, hogy a KorSzak Gyermeknyelvi Korpusz pontosan hogyan járult hozzá a tananyagkészítés során felmerülő kérdésekkel kapcsolatos hozott döntések meghozatalához.

A jelen doktori értekezésben bemutatott korpusznyelvészeti kutatás jelentőségeként még meg kell említenünk, amellet, hogy a magyar nyelv, illetve a magyar mint idegen nyelv oktatásának kevésbé kutatott területével foglalkozik, hozzájárul a gyermeknyelvi és pedagógiai korpuszok területén, valamint az agglutináló nyelvek vizsgálatával történő korpusznyelvészeti kutatások bővítéséhez is. Felhívja a figyelmet a gyermekek, ezen belül a 11–15 évesek számára szóló korpuszinformált tananyagok jelentőségére, valamint bemutatja, hogyan alkalmazhatják a szerzők a korpuszokat a munkájuk során.

Fontosságaként kiemelendő még, hogy a kutatás hasznos, látható eredményt, egy tananyag elkészülését produkálta, amely segítheti a magyarul tanuló gyermekek és tanáraik munkáját, valamint két olyan módszertani megközelítés is beépítésre került ebbe a tananyagba, amely a gyermekek magyar mint idegen nyelv oktatásában, illetve a nyelvoktatásukban hozhat megújulást. Ezek közül az első a modellalapú megközelítés (vö. Szita – Pelcz, 2017), amely a magyar mint idegen nyelv oktatása területén már széles körben alkalmazott, de a gyermekek nyelvoktatásában újdonságot jelent. A második pedig a tananyag projektalapú megközelítése (vö. Capraro, 2013; Pieratt, 2019), amely bár azt gondolnánk, hogy a nagyobb nyelvek

tananyagaiban már elterjedt, nemzetközi viszonylatban is rendkívül korlátozott számúak az ilyen módszertant alkalmazó gyermekeknek szóló nyelvkönyvek.

Végül kitekintésként szeretném elmondani, hogy a KorSzak Gyermeknyelvi Korpusz dinamikus mivoltából adódóan újabb és újabb felvételekkel folyamatosan bővül, valamint több területen is fejlődik. A korpuszépítés megkezdése előtt ezeket a terveket már elkészítettük, így a korpusz úgy épült, illetve olyan metaadatokkal van ellátva, amely ezt lehetővé teszi. A terveink közül már elkezdődött a korpuszban szereplő gyermekek életkorának kiterjesztése (16–17 évesek) is. A következő korosztállyal már megkezdődtek a próbafelvételek készítése. Ezenkívül szeretnénk a felvételek gyűjtését Magyarország más régióira is kiterjeszteni. Továbbá több alkorpusz létrehozását is tervezzük a jövőben, amelyek közül már megindultak a kétnyelvű gyermekekkel készülő felvételek (vö. Schmidt, 2022), amely új távlatokat nyithat a gyermekek nyelvhasználatának vizsgálatában.

## Referenciák

- Adolphs, S., Knight, D. (2010). Building a spoken corpus: what are the basics? In: O'Keeffe A., McCarthy M. (Eds.). *The routledge handbook of corpus linguistics*. Routledge. 38–52.
- Aksan, M., Aksan, Y. (2014). Patterns and frequency: Evidence from the Turkish National Corpus (TNC). *Proceedings of ICTL 2014*. Berlin: Harrasowitz Verlag. Web: [https://www.tnc.org.tr/wp-content/uploads/2018/05/aksan\\_aksan\\_ictl2014-copy.pdf](https://www.tnc.org.tr/wp-content/uploads/2018/05/aksan_aksan_ictl2014-copy.pdf) (Utolsó megtekintés: 2023. 01. 26.).
- Aksan, M., Aksan, Y. (2015). Multi-word expressions in genre specification. *Mersin University Journal of Linguistics and Literature*, 12, 1–42.
- Allington, R. L. (1983). Fluency: The Neglected Reading Goal. *The Reading Teacher*, 36, no. 6, 1983, 556–561.
- Alrite (2020). Áttekintés - Információ a beszéd felismerő programról. web: <https://alrite.io/ai/hu/> (Utolsó megtekintés: 2022. október 25.).
- Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations, In: A.O. Cowie (Ed.), *Phraseology: theory, analysis and application*, Oxford: Oxford University Press, 101–122.
- American Psychological Association. (2020). Publication manual of the american psychological association: the official guide to apa style (Seventh). American Psychological Association.
- Anderson, J. R. (1989). A rational analysis of human memory. In: Roediger, H. L. I., Craik, F. I. M. (Eds.) *Varieties of memory and consciousness: Essays in honour of Endel Tulving*. Hillsdale, NJ: Erlbaum, 195–210.
- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbau.
- Anderson, J. R., Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, 96, 703–719.
- Atkins, S., Clear, J., Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7 (1), 1–16.
- BAAL. (2021). *Recommendations on Good Practice in Applied Linguistics*. 4th Edition. Web: <https://www.baal.org.uk/wp-content/uploads/2021/03/BAAL-Good-Practice-Guidelines-2021.pdf> (Utolsó megtekintés: 2022. október 26.).
- Baayen, R. H. Schreuder, R. (2003). *Morphological structure in language processing*. Berlin: Mouton de Gruyter.
- Babarczy A., Lukács Á., Pléh Cs. (2014). A nyelvelsajátítás elméleti modelljei. In: Pléh Cs., Lukács Á. (Szerk.). *Pszicholingvisztika 1–2*. Budapest: Akadémiai Kiadó, 445–481.
- Baddeley, A. (1986). *Working Memory*. Oxford: Clarendon.
- Baddeley, A., Gathercole, S., Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* 105, 158–73.
- Baddeley, A., Logie, R. H. (1999). Working memory. The multiple-component model. In: Miyake, A., Shah, P. (Eds). *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control*. Cambridge, MA: Cambridge University Press, 28–61.



- Badecker, W., Miozzo, M., Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, 57(2), 193–216.
- Bannard, C., Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241–248.
- Barlow, M., Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford, CA: The Center for the Study of Language and Information Publications.
- Baumann T., Majoros J., Pelcz K., Schmidt I., Szita Sz., Vermeki B. (2020). Bemutatkozik a Korpusznyelvészeti és Szakmódszertani Munkacsoport. *Hungarológiai Évkönyv*, 21 (1-2), 32–41.
- Biber D. (2006). *University language: a corpus-based study of spoken and written registers*. J. Benjamins.
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press.
- Biber, D. (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.
- Biber, D., Conrad, S., Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, D., Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24, 199–208.
- Bilsborough, K. (é.n.). *TBL and PBL: Two learner-centred approaches*. British Council. Web: <https://www.teachingenglish.org.uk/article/tbl-pbl-two-learner-centred-approaches/> (Utoljára megtekintve:2021.10.09).
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Boersma, P., Weenink, D. (2020). *Praat: Doing phonetics by computer*. Web: <http://www.praat.org/> (Utoljára megtekintve: 2023.02.09.).
- Bóna J. (2017). GABI–Gyermeknyelvi beszédadatbázis a kutatásban. In: Bóna J. (Szerk). *Új utak a gyermeknyelvikutatásokban*, Eötvös Kiadó, 35–50. Web: [http://www.eltereader.hu/media/2017/11/Bona\\_Gyermeknyelv\\_READER.pdf](http://www.eltereader.hu/media/2017/11/Bona_Gyermeknyelv_READER.pdf) (Utoljára megtekintve: 2020.11.29.).
- Branigan, H. P., Pickering, M. J., McLean, J. F., Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition*, 104, 163–197.
- Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review*, 101, 327–341.
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47–64.
- Braun, S. (2006). ELISA—a pedagogically enriched corpus for language learning purposes. In: Braun, S., Kohn, K., Mukherjee, J. (Eds.). *Corpus technology and language pedagogy: New resources, new tools, new methods*, Frankfurt: Peter Lang, 25–47.
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19 (3), 307–328.

- Brazil, D. (1995). *A grammar of speech*. Oxford, England: Oxford University Press.
- Brumfit, C. (1992). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.
- Buerki, A. (2016). Formulaic sequences: A drop in the ocean of constructions or something more significant? *European Journal of English Studies*, 20 (1), 15–34.
- Buntinx, N., Van Goethem, K. (2018). Cross-linguistic perspectives on intensification in speech: A comparison of L1 French and L2 English and Dutch. Poster presented at Louvain-la-Neuve, Université catholique de Louvain, Using Corpora in Contrastive and Translation Studies. Web: [https://www.researchgate.net/publication/327691004\\_Cross-linguistic\\_perspectives\\_on\\_intensification\\_in\\_speech\\_A\\_comparison\\_of\\_L1\\_French\\_and\\_L2\\_English\\_and\\_Dutch](https://www.researchgate.net/publication/327691004_Cross-linguistic_perspectives_on_intensification_in_speech_A_comparison_of_L1_French_and_L2_English_and_Dutch) (Utoljára megtekintve: 2022.11.29).
- Busch, B. (2017). Expanding the notion of the linguistic repertoire: On the concept of Spracherleben –The lived experience of language? *Applied Linguistics*, 38(3), 340–358.
- Butterworth, B. (1983). Lexical representation. In: Butterworth, B. (Ed.). *Language Production*, 2, Academic Press, London, 257–294.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10 (5), 425–455.
- Bybee, J. (2008). Usage-Based Grammar and Second Language Acquisition. In: Robinson P. J., Ellis, N. C. (Eds.). *Handbook of cognitive linguistics and second language acquisition*. Routledge.
- Byrne, D. (1977). *Teaching Oral English*. London: Longman.
- Capraro, R.M., Capraro, M.M., Morgan, J.R. (Eds.) (2013). *STEM Project-Based Learning*. Sense Publishers, The Netherlands: Rotterdam.
- Carroll, G., Conklin, K. (2020). Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences. *Language and Speech*, 63 (1), 95–122.
- Cheng, W., Greaves, C., Sinclair, J. M., Warren, M. (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, 30(2), 236–252.
- Cheng, W., Greaves, Ch., Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics* 11(4). 411–433.
- Cheng, W., Warren, M. (2007). Checking understandings: Comparing textbooks and a corpus of spoken English in Hong Kong. *Language Awareness*, 163, 275–291.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34, 548–560.
- Corpus for Schools. (é.n.). Web: <https://wp.lancs.ac.uk/corpusforschools/> (Utoljára megtekintve: 2023.04.10.).
- CorpusMate (2023). *Information*. Web: <https://corpusmate.baisa.cz/info> (Utoljára megtekintve: 2023.04.11).
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory In Typological Perspective*. Oxford: Oxford University Press.
- Croft, W. (2009). *Explaining language change: an evolutionary approach*. Pearson Longman.

- Crosthwaite, P. (Ed.). (2019). *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*, Routledge.
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., Hulstijn, J. H., (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In: Housen, A., Kuiken F., Vedder, I. (Eds.). *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins, 121–142.
- de Mornay Davies, P. (1998). Automatic semantic priming: The contribution of lexical- and semantic-level processes. *European Journal of Cognitive Psychology*, 10 (4), 389–412.
- Dér Cs. I. (2019). *Grammatikalizáció*. Budapest: Akadémiai Kiadó.
- Deuchar, M. (2022) Multilingualism in Early Childhood: The Role of the Input. In: Stavans, A. Jessner U. (Eds.). *The Cambridge Handbook of Childhood Multilingualism*. 58–81.
- Dewey, J. (1976). Pedagógiai hitvallásom. *A nevelés jellege és folyamata*. Budapest: Tankönyvkiadó.
- Di Vito, N. O. (1991). Incorporating native speaker norms in second language materials. *Applied Linguistics*, 12, 383–396.
- Dóla M. (2006a). Formulaszerű nyelvhasználat–az angol formulaic language terminus beillesztése a magyar nyelvtudományi terminológiába. In: Fóris, Á. – Pusztay, J. (szerk.): *Utak a terminológiához*. Terminologia et Corpora–Supplementum. Tomus I, Szombathely, 94–116.
- Dóla M. (2006b). Formulaszerű elemsorok a meghívás beszédaktusában. *THL2*, 2 (1–2), 41–55.
- Dóla M. (2014). Lexikon és grammatika kapcsolatáról – különös tekintettel az idegennyelvtanulásra. *Hungarológiai Évkönyv*, 15 (1), 8–29.
- Dóla M. (2014). Lexikon és grammatika kapcsolatáról – különös tekintettel az idegennyelvtanulásra. *Hungarológiai Évkönyv*, 15 (1), 8–29.
- Dóla M. (2016). *Tettek, szavak, szabályok. Formulaszerűség és interakciós rutinok küszöbszintű magyar mint idegen nyelvi tanulók beszélt nyelvi szövegeiben*. Doktori disszertáció. PTE BTK Nyelvtudományi Doktori Iskola, Pécs. Web: <https://pea.lib.pte.hu/bitstream/handle/pea/14541/dola-monika-phd-2016.pdf?sequence=1&isAllowed=y> (Utoljára megtekintve: 2023.01.23.).
- Dóla M. (2018). Formulaszerű kifejezések a felnőttkori idegennyelvtanulásban–elvárások és tények. *Hungarológiai Évkönyv*, 19 (1), 17–44.
- Dörnyei Z. (2007). *Research methods in applied linguistics : quantitative qualitative and mixed methodologies*. Oxford University Press.
- Dörnyei, Z. (2009). *The Psychology of Second Language Acquisition - Oxford Applied Linguistics*. Oxford University Press.
- Durrant, P. (2013). Formulaicity in an agglutinating language: the case of Turkish. *Corpus Linguistics and Linguistic Theory*, 9(1). 1– 38.
- ECL szóbeli nyelvvizsga értékelési kritériumai (é.n.). Web: <https://ecl.hu/a-beszedkeszseg-ertekelesi-kriteriumai/> (Utoljára megtekintés: 2022. 11. 17.).
- Ellis, N. 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.

- Ellis, N. C., O'Donnell, M. B., Romer, U. (2016). Constructions and Usage-based Approaches to Language Acquisition. *Language Learning*, 66, 23–44.
- Ellis, N., Simpson-Vlach, R., and Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396.
- Ellis, N., Wulff, S. (2019). Cognitive Approaches to Second Language Acquisition. In: Schwieter, J. Benati, A. (Eds.). *The Cambridge Handbook of Language Learning*. Cambridge: Cambridge University Press. 41–61.
- Ellis, R., Barkhuizen, G. (2005). *Analyzing Learner Language*. Oxford: Oxford University Press.
- Eskey, D. E. (1983). Meanwhile, Back in the Real World...: Accuracy and Fluency in Second Language Teaching. *TESOL Quarterly*, 17 (2), 1983, 315–323.
- Fillmore, C. J. (1979). On fluency. In: Fillmore, C. J., Kempler, D., Wang, W.S. (Eds.). *Individual differences in language ability and language behaviour*. New York: Academic Press, 85–101.
- Fillmore, C. J., Kay, P., O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language*, 64(3), 501–538.
- Firth, J. R. (1957a). *Papers in linguistics 1934–1951*. London: Oxford University Press.
- Firth, J. R. (1957b). *A synopsis of linguistic theory 1930–1955*. Studies in Linguistic Analysis. Oxford: Blackwell.
- Fóris Á. (2013). Írott és íratlan szabályok a kutatásban—a kutatói etika. In: Sepsi E. (Szerk.) *Tudomány és etika. Studia Caroliensia*, KRE–L'Harmattan Kiadó, Budapest. 203–218. Web: [http://real.mtak.hu/27020/1/Foris\\_2013\\_A%20kutat%C3%B3i%20etika....pdf](http://real.mtak.hu/27020/1/Foris_2013_A%20kutat%C3%B3i%20etika....pdf) (Utoljára megnézve: 2022. 10. 26.).
- Fried-Booth, D. L. (1997). *Project work*. Oxford: Oxford University Press.
- Garrod, S., Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1, 292–304.
- GDPR (2018). *Az Európai Unió általános adatvédelmi rendelete*. Web: [https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index\\_hu.htm](https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_hu.htm) (Utoljára megnézve: 2022.10. 26.).
- Ghaleb, R., Sadighi, F. (2015). The Usage-based Theory of Language Acquisition: A review of Major Issues. *Journal of Applied Linguistics and Language Research*, 2 (6), 190–195.
- Gibbs, R.W., Nayak, N.P., Cutting, P. (1989). How to kick the bucket and not decompose: analyzability and idiom processing. *Journal of memory and language*, 28, 576–593.
- Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal*, 58, 363–374.
- Goldberg A. E. (2007). *Constructions: a construction grammar approach to argument structure*. University of Chicago Press.
- Götz, S. (2013). *Fluency in native and nonnative english speech*. John Benjamins Publishing Company.
- Gries, S. Th., Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics*, 3, 182–200.

- Guz, E. (2014). Formulaic sequences as fluency devices in the oral production of native speakers of Polish. *Research in Language*, 12 (2), 113–129.
- Gyarmathy D., Neuberger T. (2015). Egy hiánypótló adatbázis: a Tini BEA. *Beszéd kutatás* 2015, 209–222. Web: <http://real-j.mtak.hu/2676/1/Beszedkutatás2015.pdf> (Utoljára megtekintve: 2020.11.29.)
- Gyllstad, H., Wolter, B. (2016). Collocational processing in the light of a phraseological continuum model: Does semantic transparency matter? *Language Learning*, 66 (2), 296–323.
- Halliday M. (1978). *Language as Social Semiotic: The social interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. Edward Arnold.
- Halliday, M. A. K., Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M.A.K. (1959). *Language of the Chinese "Secret history of the Mongols"*. London: Longman.
- Harvard Catalyst (é.n.). Mixed Methods Research. Web: <https://catalyst.harvard.edu/community-engagement/mmr/> (Utoljára megtekintve: 2023.01.18.).
- Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes*, 4, 149–161.
- Hay, J. B., Baayen, R.H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in cognitive science*, 9 (7), 342–348.
- Heltai P. (2004). A fordító és a nyelvi normák I. *Magyar nyelvőr*, 128 (4), 407–434.
- Heltay P., Gósy M. (2005). A terpeszkedő szerkezetek hatása a feldolgozásra. *Magyar Nyelvőr*, 129 (4), 473–484.
- Hernandez, A. E., Fennema-Notestine, Ch., Udell, C., Bates, E. (2001). Lexical and sentential priming in competition: Implications for two-stage theories of lexical access. *Applied Psycholinguistics*, 22, 191–215.
- Hill, J. (2000). Revising priorities: from grammatical failure to collocational success. In: Lewis, M. (Ed.). *Teaching Collocation: Further Developments in the Lexical Approach*. London: Heinle Cengage Learning, 47–67.
- Hoey, M. (2003). Why grammar is beyond belief. *Belgian Essays on Language and Literature* 2003, Liège: University of Liège, 183–196.
- Hoey, M. (2004). The textual priming of lexis. In Bernardini, G.A.S., Stewar D. (Eds.), *Corpora and language learners*. Amsterdam: John Benjamins. 17–21.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abingdon, England: Routledge.
- Hopper, P. J. (1987). *Emergent grammar*. Berkeley Linguistics Society.
- Hornby, A. S. (1954). *A guide to patterns and usage in English*. London: OUP.
- Hunston, S. (2009). The usefulness of corpus-based descriptions of English for learners: The case of relative frequency. In: K. Aijmer (Ed.). *Corpora and Language Teaching*. John Benjamins, Amsterdam/Philadelphia. 141–157.
- Hunston, S. (2022). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S., Francis, G. (2000). *Pattern grammar: A corpus driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

- huTenTen: Corpus of the Hungarian Web (é.n.). Web: <https://www.sketchengine.eu/hutenten-hungarian-corpus/> (Utoljára megtekintve: 2023.02.09.).
- Ibbotson, P. (2013). The Scope of Usage-Based Theory. *Frontiers in Psychology*, 4, 1–15.
- IELTS (2018). *Speaking assessment criteria*. Eeb: <https://www.ielts.org/-/media/documents/online-tutorial/speaking-assessment-criteria.ashx> (Utoljára megtekintve: 2022.04.27.).
- IMS (é.n.). Web: <https://www.ims.uni-stuttgart.de/forschung/projekte/corpus-workbench/> (Utoljára megtekintve: 2023.01.23.).
- Kahng, J. (2017). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39 (3), 569–591.
- Kaltenböck, G., Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: on the potential and limitations of computer corpora in language teaching. *ReCALL*, 171, 65–84.
- Kašikci K., Vermeki B. (2021). Magyar felfedező. Nemzetstratégiai Kutatóintézet. (mintafejezetek egy gyerekeknek készülő magyar nyelvkönyvből) Web: <https://magyarfelfedezo.hu/>
- Katz, L. G., Chard, S. C., Kogan, Y. (2014). *Engaging Children's Mind—The Project Approach*. Praeger, ABC-CLIO, USA: Santa Barbara, California.
- Kilgarriff, A. (2015). Statistics used in Sketch Engine. Web: <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf> (Utoljára megtekintve: 2023. 01. 31.).
- Kirk, S. (2014). Addressing Spoken Fluency in the Classroom. In: Muller, T. (Ed.) *Exploring EFL fluency in Asia*. 101–119.
- Koponen, M., Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In: Riggenbach, H. (Ed.), *Perspectives on Fluency*. Ann Arbor, MI: University of Michigan Press. 5–24.
- Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal*, 59, 322–332.
- Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates Publishers.
- Kovács Á. (2000). Asszimetria a nyelvi rendszereken belül. *Lélekjelenlét*. 1 (1–2), Kolozsvár: Románia. 1–12. Web: [http://acta.bibl.u-szeged.hu/38133/1/lelekjelenlet\\_2000](http://acta.bibl.u-szeged.hu/38133/1/lelekjelenlet_2000) (Utoljára megtekintve: 2023.01.10).
- Kovács, J. (2009). *A gyermek és az idegen nyelv*. Budapest: Eötvös József Könyvkiadó.
- Közös Európai Referenciakeret (2002). Web: <https://nyak.oh.gov.hu/nyat/> (Utoljára megtekintve: 2022.04.26.).
- Krashen, S. (1985). *The Input Hypothesis*. Beverly Hills, California: Laredo.
- Krauss, J., Boss, S. (2013). *Thinking Through Project-Based Learning—Guiding Deeper Inquiry*. Corwin, Sage Publishing, UK: London.
- Kugler, N., Tolcsvai Nagy, G. (Szerk.). (2000). *Nyelvi fogalmak kyszótára*. Budapest: Korona.
- Kuperberg, G. R., Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language Cognition and Neuroscience*, 31 (1), 32–59.
- Langacker R. W. (1987). *Foundations of cognitive grammar*. Stanford University Press.

- Larsen-Freeman, D. (2002). Making sense of frequency. *Studies in Second Language Acquisition*, 24(2), 275–285.
- Larsen-Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. *Applied Linguistics*, 27(4), 590–619.
- Lehtonen, M., Cunillera, T., Rodriguez-Fornells, A., Hulstijn, A., Tuomainen, J., Laine, M. (2007). Recognition of morphologically complex words in Finnish: Evidence from event-related potentials. *Brain Research*, 1148, 123–137.
- Lehtonen, M., Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6(3), 213–225.
- Lennon, P. (2000). The lexical element in spoken language fluency. In: Riggenbach, H. (Ed.), *Perspectives on fluency*. Ann Arbor: University of Michigan Press. 25–42.
- Lester, N. (2015). Book review: Pace-Sigge, Michael: 2013. Lexical priming in spoken English usage. Palgrave MacMillan: New York. *Corpus Linguistics and Linguistic Theory*, 11(2), 341–361.
- Levelt, W. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In Brown, C. M., Hagoort, P. (Eds), *The Neurocognition of Language*. Oxford: Oxford University Press. 83–122.
- Lewis M. (1993). *The lexical approach: the state of elt and a way forward*. Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Hove: Language Teaching Publications.
- Lieven, E. V. M., Tomasello, M. (2008). Children’s first language acquisition from a usage-based perspective. In: Ellis, N. (Ed.). *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York and London: Routledge, 168–196.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157–201.
- MacWhinney, B. (2012). *The CHILDES Project Tools for Analyzing Talk-Electronic Edition Part 1: The CHAT Transcription Format*.
- MacWhinney, B. (2022). *Tools for Analyzing Talk Part 1: The CHAT Transcription Format*. Web: <https://talkbank.org/manuals/CHAT.pdf> (Utolsó megtekintve: 2022.11.26.)
- Martinez, R. (2020). Integrating corpus tools into mixed methods research. In: McKinley, J., Rose H. (Eds.). *The routledge handbook of research methods in applied linguistics*. Routledge.
- Mátyus K., Orosz Gy. (2014). MONYEK–Morfológiailag egyértelműsített óvodai nyelvi korpusz. *Beszédkutatás* 2014. 237–245. Web: <http://real-j.mtak.hu/974/1/Beszedkutatás2014.pdf> (Utolsó megtekintve: 2020.11.29.).
- McCarten, J. (2010). Corpus-informed course book design. In: O’Keefe, A., McCarthy, M. (Eds.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge. 413–427.
- McCarthy M. (2004). *Touchstone–From Corpus to Course Book*, Cambridge: Cambridge University Press.
- McCarthy M. (2009). *Discourse Analysis for Language Teachers*, Cambridge: Cambridge University Press.

- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41 (4), 563–574.
- McCarthy, M., McCarten, J. (2022). Writing corpus-informed materials. In: Norton, J., Buchanan, H. (Eds.). *The Routledge Handbook of Materials Development for Language Teaching*. Routledge. 170–183.
- McDonough, K., Mackey, A. (2008). Syntactic priming and ESL question development. *Studies in Second Language Acquisition*, 30, 31–47.
- McEnery T., Hardie, A. (2012) *Corpus Linguistics - Method, Theory and Practice*. Cambridge: Cambridge University Press. Web: <http://corpora.lancs.ac.uk/clmtp/3-ethics.php> (Utolsó megtekintés: 2022.10.26.).
- McEnery, T., Xiao, R., Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- McRae, K., Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24 (3), 558–572.
- Mersinli, Ü. (2015). Associative measures and multi-word extraction in Turkish. *Mersin University Journal of Linguistics and Literature*, 12, 43–61.
- Mersinli, Ü., Aksan, Y. (2016). A methodology for multi-word unit extraction in Turkish. Proceedings of the first international conference on Turkic computational linguistics, TurCling 2016, 27–31. Web: [https://www.tnc.org.tr/wp-content/uploads/2018/05/MersinliAksan\\_TurCLing.pdf](https://www.tnc.org.tr/wp-content/uploads/2018/05/MersinliAksan_TurCLing.pdf) (Utoljára megtekintve: 2023.01.23.).
- Meunier, F. (2012). Formulaic Language and Language Teaching. *Annual Review of Applied Linguistics*, 32, 111–129.
- Meunier, F., Reppen, R. (2015). Corpus versus non-corpus-informed pedagogical materials: grammar as the focus In: *The Cambridge Handbook of English Corpus Linguistics*, Biber, D., Reppen, R. (Eds.). Cambridge: Cambridge University Press.
- Meyer, D. E., Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90 (2), 227–234.
- Meyer, D. E., Schvaneveldt, R.W. (1976). Meaning, memory structure and mental processes. *Science*, 192 (2), 27–33.
- Mitchell, R., Myles, F., Marsden E. (2019). *Second language learning theories*. Routledge.
- Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford: Oxford University Press.
- MTA Nyelvtudományi Intézet. (1998–2002). *Magyar Nemzeti Szövegtár*. Web: [http://corpus.nyttud.hu/mnsz/bevezeto\\_hun.html](http://corpus.nyttud.hu/mnsz/bevezeto_hun.html) (Utoljára megtekintve: 2023.01.23.).
- Myles, F., Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA: Focusing on Psycholinguistic FSs and their identification. *Studies in Second Language Acquisition*, 39, 3–28.
- Nattinger, J. R., DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition* 4 (5), 648–654.



- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology*, 106, 226–254.
- Nelson, R. (2018). How ‘chunky’ is language? Some estimates based on Sinclair's Idiom Principle. *Corpora*, 13 (3), 431–460.
- Neuberger T. (2014). *A spontán beszéd sajátosságai gyermekkorban*. ELTE Eötvös Kiadó.
- Novák A. (2016). emMorph (Humor) Hungarian morphological analyzer. Web: <https://github.com/nytud/emMorph> (Utoljára megtekintve: 2023.01.23.).
- O’Keeffe, A., McCarthy, M., Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Ortega L., Uno, M., Park, H. I., Tyler A. (2016). *The usage-based study of language learning and multilingualism*. Georgetown University Press.
- Pace-Sigge, M. (2013a). The concept of Lexical Priming in the context of language use. *ICAME Journal*, No.37, April 2013, 149–173.
- Pace-Sigge, M. (2013b). *Lexical priming in spoken english usage*. Palgrave Macmillan.
- Palmer, H. (1925). ‘Conversation: The fundamental guiding principle for the student of conversation’, in Smith, R.C. (1999) *The Writings of Harold E. Palmer: An Overview*, Tokyo: Hon-no-Tomoshia. web: <http://homepages.warwick.ac.uk/~elsdr/WritingsofH.E.Palmer.pdf> (Utoljára megtekintve: 2022.06.12.)
- Paltridge, B., Phakiti, A. (Eds.). (2015). *Research Methods in Applied Linguistics: A Practical Resource*. Bloomsbury Publishing.
- Partington, A. (1998). *Patterns and meanings: using corpora for english language research and teaching*. John Benjamins Publishing.
- Partington, A. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies*. John Benjamins Publishing.
- Pawley, A., Syder, F. (2000). The one clause at a time hypothesis. In: Riggenbach, H. (Ed.), *Perspectives on fluency*. Ann Arbor: University of Michigan Press. 163–191.
- Pawley, A., Syder, F.H. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In: Richards, J., Schmidt, R. (Eds.). *Language and Communication*, London: Longman. 191–226.
- Pelcz K. (2023). Minek is az innováció? Gondolatok a nyelvtanulás hatékonyságának növeléséről. In: Dobos Á (Szerk.) *Aktuális kihívások a szak/nyelvoktatásban–A módszertani megújulás lehetőségei* (Tanulmánykötet). Budapesti Corvinus Egyetem. 9–17.
- Pellicer-Sánchez, A., Siyanova-Chanturia, A., Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers’ eye movements. *Applied Psycholinguistics*, 43 (3), 727–756.
- Pérez-Parades, P., Mark, G., O’Keeffe, A. (2020). *The impact of usage-based approaches on second language learning and teaching*. Cambridge: Cambridge University Press. Web: <https://www.cambridge.org/partnership/research/impact-usage-based-approaches-second-language-learning-and-teaching> (Utoljára megtekintve: 2023.04.08).
- Peters, A. M. (1983). *The Units of Language Acquisition*. Cambridge University Press.

- Pieratt, J. (2019). *CraftED*, web: <https://craftedcurriculum.com/what-is-pbl/> (Utoljára megtekintve:2021.10.09.).
- Pierce, J. E. (1961). A frequency count of Turkish affixes. *Anthropological Linguistics*, 3, 31–42.
- Pinter, A. (2011). Language Learning Processes in Childhood: First and Second Languages. *Children learning second languages*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 49–65.
- Pléh Cs., Lukács, Á. (2001). A szabályok és a kettős disszociációs elv a nyelv agyi reprezentációjában. *Magyar tudomány*, 46 (10), 1202–1206.
- Pléh, Cs. (2014). A módszertan kérdései a pszicholingvisztikában. In: Pléh Cs., Lukács Á. (Szerk.) *Pszicholingvisztika 1–2*. Budapest: Akadémiai Kiadó. 147–163.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1972). A grammar of contemporary English. Harlow: Longman. 1985. *A comprehensive grammar of the English language*. London and New York: Longman.
- Ratcliff, R., McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95(3), 385–408.
- Reményi A. Á. (2010). Kollokációk korpuszalapú vizsgálata. *Fordítástudomány*, XII (2), 67–95.
- Sampson, G. P. (1977). A Real Challenge to ESL Methodology. *TESOL Quarterly*, 11(3), 1977, 241–255.
- Schmidt, I. (2022). Kétnyelvű gyermekek beszéde mint az élőnyelv egy változata. In: Karmacs, Z, Márku, A., Máté, R. (Szerk.) *A határ mint konvergáló és divergáló tényező a nyelvben: Tanulmányok a 21. Élőnyelvi Konferenciáról*. Törökbálint, Magyarország: Termini Egyesület, II. Rákóczi Ferenc Kárpátaljai Magyar Főiskola, Hodinka Antal Intézet, 144-153.
- Schmidt, R. W. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357–385.
- Schooler, L. J., Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, 32, 219–250.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Shelton, J. R., Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18 (6), 1191–1210.
- Simon G. (2012). *Egy kognitív poétikai rímelmélet megalapozása, doktori disszertáció*. Web: <http://doktori.btk.elte.hu/lingv/simongabor/diss.pdf> (Utoljára megtekintve: 2023.01.09.).
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sinclair, J. (2004). The search for units of meaning. In: *Trust the text: language, corpus and discourse*, London: Routledge. 24–48.
- Sinclair, J. (2006). A language landscape. West Word. Web: [www.westword.org.uk/jan2006.html](http://www.westword.org.uk/jan2006.html) (Utoljára megtekintve: 2022.08.02.)
- Sinclair, J. (2006). *Trust the text: language corpus and discourse*. Routledge.
- Siyanova-Chanturia, A., Pellicer-Sanchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. New York, NY: Routledge.

- Siyanova-Chanturia, A., Conklin, K., Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers, *Second Language Research*, 27 (2), 1–22.
- Skehan, P. (2003). Task-Based Instruction. *Language Teaching*, 36, 1–14.
- SkELL (é.n.). Web: <https://www.sketchengine.eu/skell/> (Utoljára megtekintve: 2023.02.03.).
- Sketch Engine (é.n.a). *Build a corpus from the web*. Web: <https://www.sketchengine.eu/guide/create-a-corpus-from-the-web/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.b). *Text type and metadata in a corpus*. Web: [https://www.sketchengine.eu/my\\_keywords/text-type/](https://www.sketchengine.eu/my_keywords/text-type/) (Utoljára megtekintve: 2023.01.31.).
- Sketch Engine (é.n.c). *Wordlist — frequency lists and linguistic databases*. Web: <https://www.sketchengine.eu/guide/wordlist-frequency-lists/> (Utoljára megtekintve: 2023.01.31.).
- Sketch Engine (é.n.d). *N-grams — frequency lists of multiword expressions (MWEs) or lexical bundles*. Web: <https://www.sketchengine.eu/guide/n-grams-multiword-expressions/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.e). *Keywords and term extraction — identifying typical words*. Web: <https://www.sketchengine.eu/guide/keywords-and-term-extraction/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.f). *Word Sketch — collocations and word combinations*. Web: <https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.g). *Word sketch difference — compare words via their collocations*. Web: <https://www.sketchengine.eu/guide/word-sketch-difference-compare-words/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.h). *Text type analysis—statistics of metadata in the corpus*. Web: <https://www.sketchengine.eu/guide/text-type-analysis/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.i). *Concordance — examples of use in context*. Web: <https://www.sketchengine.eu/guide/concordance-a-tool-to-search-a-corpus/> (Utoljára megtekintve: 2023. 01. 31.).
- Sketch Engine (é.n.j). *Trial and paid account limitations*. Web: <https://www.sketchengine.eu/guide/account-limitations/> (Utoljára megtekintve: 2023. 02. 04.).
- Slobin, D. I. (1997). The origins of grammaticizable notions: Beyond the individual mind. In: Slobin, D. I. (Ed.). *The crosslinguistic study of language acquisition*, 5, Mahwah, NJ: Erlbaum. 265–323.
- Smith, S.M., Brown, J.M., Balfour, S.P. (1991). TOTimals: A controlled experimental method for studying tip-of-the-tongue states. *Bulletin of Psychonomic Society*, 29, 445–447.
- Soveri, A., Lehtonen, M., Laine, M. (2007). Word frequency and morphological processing in Finnish revisited. *The mental lexicon*, 2 (3), 359–385.

- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. (Textbooks in Language Sciences 7). Berlin: Language Science Press. Web: <http://langsci-press.org/catalog/book/148> (Utoljára megtekintve: 2023.04.08.).
- Steinkrauss, R. (2017). 5. L1 acquisition beyond input frequency. In: Evers-Vermeul, J., Tribushinina, E. (Eds.). *Usage-Based Approaches to Language Acquisition and Language Teaching*. Berlin, Boston: De Gruyter Mouton. 117–142.
- Stoll, S., Schikowsky, R. (2020). Child-Language Corpora. In: Paquot, M., Gries, S. T. (Eds.). *A practical handbook of corpus linguistics*. Springer. 503–530.
- Stubbs, M. (1996). *Text and corpus analysis. Computer-assisted analysis of language and culture*. Oxford: Basil Blackwell.
- Stubbs, M. (1997). Review of T. McEnery and A. Wilson (1996), *Corpus Linguistics*. *International Journal of Corpus Linguistics*, 2 (2), 296–302.
- Stubbs, M. (2002). *Words and Phrases: Corpus Studies of Lexical Semantics*. Blackwell.
- Szirmai M. (2005). *Bevezetés a korpusznyelvészetbe - A korpusznyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában*. Budapest: Tinta Könyvkiadó.
- Szita Sz. (2014). Invent content, not language– Meaningful interaction and natural language use in the classroom. In: Hegedüs R., Görbe T. (Eds.) *Kleine Sprachen, was nun ...? Studies on Language and Culture in Central and Eastern Europe*. München / Germany: Kubon und Sagner Verlag. 112–127.
- Szita Sz. (2020). Korpuszépítés és korpuszhasználat alacsonyabb nyelvtudási szinteken. *Hungarológiai Évkönyv*, 21 (1-2), 173–179.
- Szita Sz. (2021). A MagyarOK nyílt korpusz használata. *Hungarológiai Évkönyv*, 22 (1–2), 72–88.
- Szita Sz. (2022). *Olvasókönyv a MagyarOK A1-es kötetéhez*. Pécs: Modellalapú Nyelvoktatás Intézete.
- Szita Sz., Pelcz K. (2013). *MagyarOK A1+*, Pécsi Tudományegyetem ÁOK.
- Szita Sz., Pelcz K. (2017). Modellalapú nyelvoktatás–Természetes nyelvhasználat a tanteremben és a tantermen kívül. *THL2*, (1-2), 262–269.
- Szita Sz., Pelcz K. (2020. június 3.). Magyarok A1-től B2-ig: módszertani interjú Szita Szilviával és Pelcz Katalinnal. *Ponthu*. Web: [https://ponthu.blog.hu/2020/06/03/magyarok\\_a1-tol\\_b2-ig\\_modszertani\\_interju\\_szita\\_szilviaval\\_es\\_pelcz\\_katalinnal](https://ponthu.blog.hu/2020/06/03/magyarok_a1-tol_b2-ig_modszertani_interju_szita_szilviaval_es_pelcz_katalinnal) (Utoljára megtekintve: 2023.04.12.).
- Szita Sz., Pelcz K. (é.n.). *MagyarOK A1+ Tanári kézikönyv*, Pécsi Tudományegyetem ÁOK.
- Szita Sz., Pelcz K. (é.n.). *MagyarOK weboldal*. Web: <http://magyar-ok.hu/hu/home.html> (Utoljára megtekintve: 2023.02.03.).
- Taft, M., Forster, K. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning, Verbal Behaviour*. 14, 638–647.
- Tavakoli P., Wright C. (2020). *Second language speech fluency: from research to practice*. Cambridge: Cambridge University Press.
- Teubert, W. (2005). My version of corpus linguistics, *International Journal of Corpus Linguistics*, 10 (1), 1–13.
- Theakston, A., Lieven, E. (2017). Multiunit Sequences in First Language Acquisition, *Topics in Cognitive Science*, 9, 588–603.

- Thompson, P. (2010). Building a specialized audio-visual corpus. In: O’Keeffe, A., McCarthy, M. (Eds.). *The Routledge handbook of corpus linguistics*. London: Routledge. 93–104.
- Thornbury, S. (2022). material for developing speaking skills. In: Norton, J., Buchanan, H. (Eds.). *The Routledge Handbook of Materials Development for Language Teaching*. Routledge. 218–232.
- Thuma O., Pléh Cs. (1995). Lexikális előfeszítés a mentális lexikonban az anyanyelv és egy tanult második nyelv között. *Magyar Pszichológiai Szemle*, 51 (5–6), 293–304.
- Thuma, O. (2014). A mentális szótár. In: Pléh Cs., Lukács Á. (Szerk.). *Pszicholingvisztika 1–2*. Budapest: Akadémiai Kiadó. 167–194.
- Thuma, O. (2014). Kétértelmű szavak a magyarban. In: Pléh Cs., Lukács Á. (Szerk.). *Pszicholingvisztika 1–2*. Budapest: Akadémiai Kiadó. 219–222.
- Timmis, I. (2005). Towards a framework for teaching spoken grammar. *ELT Journal*, 59(2), 117–125.
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. London: Routledge.
- Timmis, I. (2022). Theory and practice in materials development. In: Norton, J., Buchanan, H. (Eds.). *The Routledge Handbook of Materials Development for Language Teaching*. Routledge. 30–46.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2013). The usage-based theory of language acquisition. In: Bavin, E. L. (Ed.). *The Cambridge Handbook of Child Language*. Cambridge: Cambridge University Press. 69–88.
- Tomasello, M. (2015). The usage-based theory of language acquisition. In: Bavin, E., Naigles, L. (Eds.). *The Cambridge Handbook of Child Language*. Cambridge: Cambridge University Press. 89–106.
- Tomasello, M., Dweck, C. S., Silk, J. B., Skyrms, B., Spelke, E. S. (2009). *Why we cooperate*. MIT Press.
- Towell, R., Dewaele, J.M. (2005). The Role of Psycholinguistic Factors in the Development of Fluency Amongst Advanced Learners of French. In: Dewaele, J.M. (Ed.). *Focus on French as a Foreign Language: Multidisciplinary Approaches*. Bristol, Blue Ridge Summit: Multilingual Matters. 210–239.
- Tribushinina, E., Evers-Vermeul, J. (2017). Language acquisition and language teaching in the usage-based framework. In: Evers-Vermeul, J., Tribushinina, E. (Eds.). *Usage-Based Approaches to Language Acquisition and Language Teaching*. Berlin, Boston: De Gruyter Mouton. 1–10.
- Tribushinina, E., Gillis, S. (2017). Advances and lacunas in usage-based studies of first language acquisition. In: Evers-Vermeul, J., Tribushinina, E. (Eds.). *Usage-Based Approaches to Language Acquisition and Language Teaching*. Berlin, Boston: De Gruyter Mouton. 13–46.
- Trofimovich, P. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics*, 26, 479–504.

- Ur, P. (2012). *A course in English language teaching*. Cambridge Teacher Training and Development. UK: Cambridge University Press.
- Vandeweerd, N., Keijzer, M. (2018). J'ai l'impression que: Lexical bundles in the dialogues of beginner French textbooks. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 21 (2), 80–101.
- Vartiainen, J., Aggularo, S., Lehtonen, M., Hulten, A., Laine, M., Salmelin, R. (2009). Neural dynamics of reading morphologically complex words. *Neuroimage*, 47(4), 2064–2072.
- Vermeki B. (2021a). Using a child spontaneous spoken language corpus for informing Hungarian as a foreign language teaching material. *Journal of Applied Languages and Linguistics*, 5 (2), Athens, Greece. 49–58.
- Vermeki B. (2021b). Magyar felfedező, egy gyermekeknek készülő magyar nyelvi tananyag. *Hungarológiai Évkönyv*, 22 (1-2). 156–165.
- Vermeki B. (2022a). Élőnyelvi korpusz-informált tananyagok a MID oktatásában. In: Karmacsi, Z, Márku, A., Máté, R. (Szerk.) *A határ mint konvergáló és divergáló tényező a nyelvben: Tanulmányok a 21. Élőnyelvi Konferenciáról*. Törökbálint, Magyarország: Termini Egyesület, II. Rákóczi Ferenc Kárpátaljai Magyar Főiskola, Hodinka Antal Intézet, 254–268.
- Vermeki B. (2022b, megjelenés alatt). *Formulaic language in child spontaneous speech productions. САРМЕНИ ТОКОВИ У ИЗУЧАВАЊУ ЈЕЗИКА, КЊИЖЕВНОСТИ И КУЛТУ* tanulmánykötet, Belgrádi Egyetem, Szerbia.
- Vermeki B. (2022c, megjelenés alatt) Gyakori többszavas lexikai egységek gyermekekbeszédprodukciónban, *Hungarológiai Évkönyv*, 23.
- Vermeki B. (2023a). *A lexikai jóslás jelentősége a nyelvoktatásban*. In: Dobos Á (Szerk.) *Aktuális kihívások a szak/nyelvoktatásban–A módszertani megújulás lehetőségei* (Tanulmánykötet). Budapesti Corvinus Egyetem. 115–123.
- Vermeki B. (2023b, megjelenés alatt). *Lexico-grammatical patterns of children's language use. Journal of Applied Languages and Linguistics*, Athens, Greece.
- Vermeki B. (2023c, megjelenés alatt). *A gyermekek spontán beszédének jellemzői - szógyakorisági és kulcsszó vizsgálatok a KorSzak Gyermeeknyelvi Korpuszon*. Nyelv és megismerés – Nyelvészeti kutatások határon innen és túl konferencia tanulmányai, Hungarológiai Közlemények. Újvidék, Szerbia
- Vilkaite, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42 (10), 1632–1642.
- Walsh, S. (2010). What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses. In: O'Keeffe, A., McCarthy, M. (Eds.). *The Routledge handbook of corpus linguistics*. London: Routledge. 333–344.
- Widmann, J., Kohn, K. and Ziai, R. (2011). 'The SACODEYL search tool: exploiting corpora for language learning purposes'. In: Frankenberg Garcia, A., Flowerdew, L., Aston, G. (Eds). *New Trends in Corpora and Language Learning*, London: Continuum: 167–180.
- Willis, D. (2003). *Rules, Patterns and Words: Grammar and Lexis in English Language Teaching*. Cambridge University Press.
- Wills, D., Wills, J. (2007). *Doing task-based teaching*. Oxford: University Press.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

- Wray, A. (2019). Concluding Question, Why Don't Second Language Learners More Proactively Target Formulaic Sequences? In: Siyanova-Chanturia, A., Pellicer-Sanchez, A. (Eds.). *Understanding formulaic language: A second language acquisition perspective*. New York, NY: Routledge. 248–269.
- Wray, A., Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorised language as a gauge of nativelike competence. In: Meunier, F., Granger, S. (Eds.). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins. 123–148.

## Korpuszok

- Bíró Andrea (2013). Biro Trilingual Corpus.  
<https://childes.talkbank.org/access/Biling/Biro.html>
- Bodor Péter (1994). Bodor Corpus.  
<https://childes.talkbank.org/access/Other/Hungarian/Bodor.html>
- Brezina, V., Gablasova, D., Reichelt, S. (2018). BNClab. <http://corpora.lancs.ac.uk/bnclab>  
Lancaster University.
- HuTenTen12 (2012). <https://www.sketchengine.eu/hutenten-hungarian-corpus/> (Sketch Engine)
- KorSzak Gyermeeknyelvi Korpusz (2020-2022. október). [ske.li/vfe](http://ske.li/vfe)
- MacWhinney, Brian (1970). MacWhinney Corpus.  
<https://childes.talkbank.org/access/Other/Hungarian/MacWhinney.html>
- MagyarOK korpusz (2013-2019).  
[https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok\\_hp2](https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fmagyarok_hp2)
- Navracsics Judit (1994-1997). Navracsics Bilingual Corpus.  
<https://childes.talkbank.org/access/Biling/Navracsics.html>
- Réger Zita (1990). Réger Corpus.  
<https://childes.talkbank.org/access/Other/Hungarian/Reger.html>
- SACODEYL (2004-2008). <http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp>
- SKELL (é.n.). <https://skell.sketchengine.eu/#home?lang=en>
- Türkçe Ulusal Derlemi (2017). Web: <https://www.tnc.org.tr/>
- CorpusMate (2023). Web: <https://corpusmate.baisa.cz/search>

## Ábrák és diagrammok jegyzéke

|                 |   |     |
|-----------------|---|-----|
| <b>1. ábra</b>  | A formulaszerű egységek jellemzői .....   | 44  |
| <b>2. ábra</b>  | Levelt beszédprodukción-modellje Levelt (1989: 9) alapján.....                            | 55  |
| <b>3. ábra</b>  | Kormos beszédprodukción-modellje Kormos (2006: 168) alapján .....                         | 56  |
| <b>4. ábra</b>  | Tavakoli és Hunter piramis modellje Tavakoli (2020: 7-8) alapján .....                    | 59  |
| <b>5. ábra</b>  | Nyelvtanárok véleménye a beszédflowkonyságról .....                                       | 66  |
| <b>6. ábra</b>  | Egy spanyol kisfiú és szüleinek felvétele a CHILDES adatbázisban .....                    | 76  |
| <b>7. ábra</b>  | CHILDES magyar nyelvű korpusz a Sketch Engine felületén a korpusz készítői szerint .....  | 76  |
| <b>8. ábra</b>  | CHILDES magyar nyelvű korpusz a Sketch Engine felületén a szószámok szerint ..            | 77  |
| <b>9. ábra</b>  | A SACODEYL angol nyelvű alkorpuszának kezdőképernyője .....                               | 84  |
| <b>10. ábra</b> | A SACODEYL keresője: a leggyakoribb szavak az állatok témakörében .....                   | 85  |
| <b>11. ábra</b> | A CorpusMate platform kezdőlapja .....  | 85  |
| <b>12. ábra</b> | A BNClab keresőfelületének kezdőlapja .....   | 90  |
| <b>13. ábra</b> | Konkordanciasor a BNClab felületén .....  | 91  |
| <b>14. ábra</b> | Jó szótári példamondatok [GDEX] a SKELL felületén .....                                   | 92  |
| <b>15. ábra</b> | A szavak kollokációinak és grammatikai mintázatainak keresése a SkELL felületén .....     | 92  |
| <b>16. ábra</b> | A hasonló szavak keresése a SkELL felületén.....  | 92  |
| <b>17. ábra</b> | A MagyarOK nyílt korpusznak a kezdőoldala a Sketch Engine-en.....                         | 93  |
| <b>18. ábra</b> | Scott Thornbury: Natural Grammar (2004: 2) .....  | 96  |
| <b>19. ábra</b> | Példa a mondatmodellek bemutatására, variációk egy témára (Szita – Pelcz, 2013: 177)..... | 105 |
| <b>20. ábra</b> | Példa a Szavak kontextusban című részből (Szita, 2022: 16).....                           | 106 |
| <b>21. ábra</b> | A lányok és fiúk aránya a KorSzak Gyermeknyelvi Korpuszban (2022. október). 110           |     |
| <b>22. ábra</b> | A KorSzak Gyermeknyelvi Korpuszban szereplő szövegek témakörei.....                       | 110 |
| <b>23. ábra</b> | A felmérés eredményeinek összesítése közben készült Excel tábla részlet .....             | 112 |
| <b>24. ábra</b> | A hobbi témakörhöz készült gondolattérkép.....  | 113 |
| <b>25. ábra</b> | A felvételek tokenek száma szerinti évenkénti eloszlása.....                              | 114 |
| <b>26. ábra</b> | A felvételek számának évenkénti eloszlása .....   | 114 |
| <b>27. ábra</b> | Beszélők számának felvételenkénti eloszlása a KorSzak Gyermeknyelvi Korpuszban .....      | 116 |
| <b>28. ábra</b> | A KorSzak Gyermeknyelvi Korpusz felvételein szereplő gyermekek életkori megoszlása .....  | 117 |
| <b>29. ábra</b> | A korpuszalapú kutatások három elemének interakciója Martinez (2019: 214) alapján .....   | 119 |
| <b>30. ábra</b> | A kevert módszertanú kutatások alap típusai (Harvard, é.n.) alapján.....                  | 121 |
| <b>31. ábra</b> | Korpuszépítési lehetőségek a Sketch Engine felületén .....                                | 123 |
| <b>32. ábra</b> | Korpuszépítés a Sketch Engine felületén interneten található szövegek segítségével .....  | 124 |
| <b>33. ábra</b> | Metaadatok megadása a Sketch Engine felületén .....                                       | 125 |
| <b>34. ábra</b> | A KorSzak Gyermeknyelvi Korpusz kezdőképernyője a Sketch Engine felületén                 | 127 |
| <b>35. ábra</b> | A Wordlist egyszerű keresés funkciója .....   | 128 |



|                 |  |     |
|-----------------|--|-----|
| <b>36. ábra</b> | A Wordlist összetett keresés funkciója .....   | 129 |
| <b>37. ábra</b> | A Wordlist funkció során alkalmazható gyakorisági mutatók .....  | 130 |
| <b>38. ábra</b> | Az N-grams egyszerű keresés funkciója .....  | 131 |
| <b>39. ábra</b> | Az N-grams összetett keresés funkciója.....  | 132 |
| <b>40. ábra</b> | A Word Sketch egyszerű keresés funkciója.....  | 133 |
| <b>41. ábra</b> | A Word Sketch összetett keresés funkciója .....  | 134 |
| <b>42. ábra</b> | A Word Sketch második összetett keresés funkciója.....   | 135 |
| <b>43. ábra</b> | A Word Sketch keresési eredményei (keresett lemma: KUTYA) .....  | 135 |
| <b>44. ábra</b> | A Word Sketch keresési eredményei (keresett lemmák: KUTYA + NAGY).....   | 136 |
| <b>45. ábra</b> | A Word Sketch találatainak vizualizációja (keresett lemma: KUTYA) .....  | 136 |
| <b>46. ábra</b> | A Word Sketch Difference egyszerű kereső funkciója .....   | 137 |
| <b>47. ábra</b> | A Word Sketch Difference összetett kereső funkciója.....   | 138 |
| <b>48. ábra</b> | A Word Sketch Difference keresési eredményei (keresett lemmák: JÖN – MEGY)<br>.....                                    | 138 |
| <b>49. ábra</b> | A Word Sketch Difference találatainak vizualizációja (keresett lemmák: JÖN –<br>MEGY).....                             | 139 |
| <b>50. ábra</b> | A Word Sketch és a Word Sketch Difference korpuszeszközök által használt<br>logDice képlet (Kilgarriff, 2015: 2) ..... | 139 |
| <b>51. ábra</b> | A Teyt type analysis kezdőoldal .....  | 141 |
| <b>52. ábra</b> | A Concordance egyszerű keresés funkciója.....  | 142 |
| <b>53. ábra</b> | A Concordance összetett keresés funkciója .....  | 143 |
| <b>54. ábra</b> | a Sketch Engine keresőeszközeivel használható jelek [RegEx] (Sketch Engine, é.n.i)<br>.....                            | 143 |
| <b>55. ábra</b> | A k betűvel kezdődő lemmák a KorSzak Gyerme nyelvi Korpuszban.....   | 144 |
| <b>56. ábra</b> | Példa a kapott eredményekre lekérdezési típusonként (Sketch Engine, é.n.i) .....                                       | 144 |
| <b>57. ábra</b> | Találatok a két b betűt tartalmazó tokenekre a KorSzak Gyerme nyelvi Korpuszban<br>.....                               | 145 |
| <b>58. ábra</b> | Találatok a két b betűt tartalmazó melléknevekre a KorSzak Gyerme nyelvi<br>Korpuszban.....                            | 146 |
| <b>59. ábra</b> | Találatok a középfokú melléknevekre a KorSzak Gyerme nyelvi Korpuszban.....  | 147 |
| <b>60. ábra</b> | Találatok a felsőfokú melléknevekre a KorSzak Gyerme nyelvi Korpuszban .....   | 147 |
| <b>61. ábra</b> | Találatok a középfokú melléknevekre a KorSzak Gyerme nyelvi Korpuszban<br>mondatonkénti elrendezésben .....            | 148 |
| <b>62. ábra</b> | A Concordance eszköz nézetének beállítási lehetőségei.....   | 149 |
| <b>63. ábra</b> | A Concordance eszköz nézetének beállítási lehetőségei, metaadatok megjelenítése<br>.....                               | 150 |
| <b>64. ábra</b> | A konkordanciasorok keresési beállításainak makró rögzítése .....  | 151 |
| <b>65. ábra</b> | A Praat programhoz tartozó weboldal kezdőlapja .....   | 152 |
| <b>66. ábra</b> | Hangfájl importálása külső forrásból a Praat szoftverben.....  | 153 |
| <b>67. ábra</b> | A beszéd szünetek automatikus jelölésének módja a Praat szoftverben .....  | 154 |
| <b>68. ábra</b> | A kutatás kvantitatív és kvalitatív vizsgálatok szekvenciái .....  | 173 |
| <b>69. ábra</b> | A huTenTen12 korpusz adatai .....  | 176 |
| <b>70. ábra</b> | A szavak szófajonkénti előfordulása a KorSzak Gyerme nyelvi Korpuszban (2020.<br>október) .....                        | 183 |
| <b>71. ábra</b> | Példa az én is így vagyok ezzel/vele rövidített formulaszerű szekvencia<br>konkordanciasorára .....                    | 191 |

|   |     |
|---|-----|
| <b>72. ábra</b> Szünetek helye az ezzel én is így vagyok kimondása közben (doc#2: Mon_Állatok_003).....                 | 193 |
| <b>73. ábra</b> Szünetek helye az én is így vagyok ezzel kimondása közben (doc#54: Mon_Kirándulás_001) .....            | 193 |
| <b>74. ábra</b> Szünetek helye az én is így vagyok vele kimondása közben (doc#73: Mon_Utazás_001) .....                 | 194 |
| <b>75. ábra</b> Az UGYE pozíciója a megnyilatkozásokban .....   | 204 |
| <b>76. ábra</b> Találatok a nem tudom + kérdőszó mintázatra a KorSzak Gyermeknyelvi Korpuszban.....                     | 207 |
| <b>77. ábra</b> A nem tudom + Inf. mintázat megjelenése a KorSzak Gyermeknyelvi Korpuszban .....                        | 207 |
| <b>78. ábra</b> A TUD lemma kollokációi.....  | 208 |
| <b>79. ábra</b> A TUD lemma kolligációi .....   | 209 |
| <b>80. ábra</b> A mit tudom én háromelemű lexikai egység leggyakoribb megjelenéseinek konkordanciasora .....            | 211 |
| <b>81. ábra</b> Példák az IGAZÁBÓL + NEM használatára.....  | 224 |
| <b>82. ábra</b> A HÁT IGAZÁBÓL ÉN konkordanciasora .....  | 224 |
| <b>83. ábra</b> Szünetek helye a hát igazából nálam is hasonló a helyzet kimondása közben (doc#68: Mon_Sport_001) ..... | 225 |
| <b>84. ábra</b> Az én is használata rövid válaszként a KorSzak Gyermeknyelvi Korpuszban.....                            | 226 |
| <b>85. ábra</b> Milyen állatok élnek az állatkertben? (A1) .....  | 246 |
| <b>86. ábra</b> Az állatok bemutatása (A1).....   | 247 |
| <b>87. ábra</b> Az állatok bemutatása (A2) 1.....   | 248 |
| <b>88. ábra</b> Az állatok bemutatása (A2) 2.....   | 248 |
| <b>89. ábra</b> Állati rekordok (A2).....   | 249 |
| <b>90. ábra</b> Milyen lehet egy kutya? (A2).....   | 250 |
| <b>91. ábra</b> Milyen lehet egy háziállat? (B1) .....  | 250 |
| <b>92. ábra</b> Vélemény kifejtése (B2).....  | 253 |
| <b>93. ábra</b> Vélemény kifejtése, információ kérése, válaszadás késleltetése (C1) .....                               | 254 |
| <b>94. ábra</b> Érvelés és meggyőzés (A2) .....   | 254 |
| <b>95. ábra</b> Érvelés és meggyőzés (B1) .....   | 255 |
| <b>96. ábra</b> Beszélgetés egy elveszett állatról és plakát készítése (A1).....  | 256 |
| <b>97. ábra</b> Az ILYEN és az OLYAN mellett megjelenő leggyakoribb főnevek a Magyar felfedező szövegeiben .....        | 257 |
| <b>98. ábra</b> Szereted az állatokat? (A1) .....   | 258 |
| <b>99. ábra</b> Nektek mi a kedvenc kutyafajtátok? (A2).....  | 260 |
| <b>100. ábra</b> És milyen volt az állatkert? (B2).....   | 260 |
| <b>101. ábra</b> A Magyar felfedező 4. fejezetében lévő kapcsolódó feladatok összefoglaló ábrája .....                  | 265 |

## Táblázatok jegyzéke

|  |     |
|--|-----|
| <b>1. táblázat</b> Magyar mint idegen nyelvi és származásnyelvi tananyagok életkoronként és nyelvi szintenként.....  | 5   |
| <b>2. táblázat</b> A lexikai előfeszítés elméletének tíz pontja (Hoey, 2005: 13) alapján .....   | 34  |
| <b>3. táblázat</b> A formulaszerű nyelvhasználat aspektusai (Wray, 2002:9) alapján .....   | 41  |
| <b>4. táblázat</b> A beszédfolyékonyság mérhető egységei Kormos (2006: 163) alapján .....  | 49  |
| <b>5. táblázat</b> A Segalowitz-féle hármass modellt és annak meghatározó tényezői .....   | 53  |
| <b>6. táblázat</b> Az automatizálás értelmezéseinek összefoglalása Dörnyei (2009: 270–274) alapján .....   | 58  |
| <b>7. táblázat</b> A folyékonyság meghatározása (KER, 2002: 36–37) .....   | 63  |
| <b>8. táblázat</b> Az ECL beszédképesség összetevőjének értékelési kritériumai (ECL, é.n.) .....   | 64  |
| <b>9. táblázat</b> Az beszédképesség értékeléseinek kritériumai (IELTS, 2018).....   | 65  |
| <b>10. táblázat</b> A korpuszok típusai felhasználási módjuk szerint Hunston (2022: 21–30) és saját gyűjtésű példák alapján .....  | 70  |
| <b>11. táblázat</b> A gyermeknyelvi korpuszvizsgálatok típusai és azok jellemzői.....  | 73  |
| <b>12. táblázat</b> A Sketch Engine felületen megtalálható CHILDES korpuszok.....  | 75  |
| <b>13. táblázat</b> A CHILDES adatbázisban megtalálható magyar vonatkozású gyermeknyelvi korpuszok összehasonlítva a KorSzak Gyermeknyelvi Korpuszsal .....                | 79  |
| <b>14. táblázat</b> Példa az ismétlődő előfordulásokra a MagyarOK tananyagaiban.....   | 102 |
| <b>15. táblázat</b> A születésnap meghíváshoz, valamint a telefonbeszélgetésekhez tartozó lexikai egységek az előző párbeszédben (Szita – Pelcz, 2013: 174. alapján) ..... | 104 |
| <b>16. táblázat</b> A KorSzak Gyermeknyelvi Korpuszban lévő témakörök tokenszám szerint ...  | 111 |
| <b>17. táblázat</b> A KorSzak Gyermeknyelvi Korpuszban lévő szövegek típusai.....  | 115 |
| <b>18. táblázat</b> A lexikai egységek szövegben betöltött szerepük szerint Biber (2006: 139-146) alapján .....  | 167 |
| <b>19. táblázat</b> A KorSzak Gyermeknyelvi Korpusz és a huTenTen12 első nyolc leggyakoribb lemmája .....  | 176 |
| <b>20. táblázat</b> A KorSzak Gyermeknyelvi Korpusz és a huTenTen12 9-20. leggyakoribb lemmája .....   | 177 |
| <b>21. táblázat</b> A szignifikánsan nagyobb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező itemek a 9–20. leggyakoribb lemma közül.....                  | 178 |
| <b>22. táblázat</b> A TUD, NAGYON, ILYEN és az AKKOR lemmák gyakorisága a többemlékű lexikai egységekben .....   | 179 |
| <b>23. táblázat</b> A KorSzak Gyermeknyelvi Korpusz első száz lemmája között (21–100. hely) szereplő, de a huTenTen12 első száz lemmája között nem megjelenő itemek .....  | 180 |
| <b>24. táblázat</b> A TÖK, GONDOL, EGYSZER, IGAZÁBÓL és a SZOKIK lemmák gyakorisága a többemlékű lexikai egységekben .....   | 181 |
| <b>25. táblázat</b> A szignifikánsan nagyobb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező itemek a 21–100. leggyakoribb lemma közül.....                | 181 |
| <b>26. táblázat</b> Az ÍGY, SZERET, OLYAN, SZERINT és a MOND lemmák gyakorisága a többemlékű lexikai egységekben .....   | 182 |
| <b>27. táblázat</b> Az 50 leggyakoribb főnév a KorSzak Gyermeknyelvi Korpuszban .....  | 184 |
| <b>28. táblázat</b> Az 50 leggyakoribb melléknév a KorSzak Gyermeknyelvi Korpuszban.....   | 185 |

|  |     |
|--|-----|
| <b>29. táblázat</b> A szignifikánsan alacsonyabb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező melléknevek .....                             | 187 |
| <b>30. táblázat</b> Az 50 leggyakoribb ige a KorSzak Gyermeknyelvi Korpuszban .....  | 187 |
| <b>31. táblázat</b> A szignifikánsan magasabb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkező igék.....  | 189 |
| <b>32. táblázat</b> Az én is így vagyok + vele/ezzel előfordulásai a KorSzak Gyermeknyelvi Korpuszban.....   | 192 |
| <b>33. táblázat</b> Az én is így vagyok + vele/ezzel közben tartott szünetek összefoglaló táblázata .....  | 194 |
| <b>34. táblázat</b> A lexikai egységek kiválasztásának meghatározott kritériumai.....  | 196 |
| <b>35. táblázat</b> A tíz leggyakoribb kételemű lexikai egység a KorSzak Gyermeknyelvi Korpuszban.....   | 197 |
| <b>36. táblázat</b> A tíz leggyakoribb háromelemű lexikai egység a KorSzak Gyermeknyelvi Korpuszban.....   | 198 |
| <b>37. táblázat</b> A tíz leggyakoribb négyelemű lexikai egység a KorSzak Gyermeknyelvi Korpuszban.....  | 198 |
| <b>38. táblázat</b> A leggyakoribb kételemű, a háromelemű és négyelemű lexikai egységek összehasonlító táblázata .....   | 200 |
| <b>39. táblázat</b> Az ugye kollokációi a KorSzak Gyermeknyelvi Korpuszban.....  | 202 |
| <b>40. táblázat</b> A NEM TUD kételemű lexikai egység előfordulása a KorSzak Gyermeknyelvi Korpuszban.....   | 206 |
| <b>41. táblázat</b> A MI + TUD együttes megjelenéseinek használati mintázatai a KorSzak Gyermeknyelvi Korpuszban.....  | 211 |
| <b>42. táblázat</b> Az annyira nem kételemű lexikai egység előfordulása a KorSzak Gyermeknyelvi Korpuszban.....  | 214 |
| <b>43. táblázat</b> Az ANNYIRA és a NEM együttes előfordulásai a KorSzak Gyermeknyelvi Korpuszban.....   | 214 |
| <b>44. táblázat</b> Az ILYEN és az OLYAN lemma megjelenése a KorSzak Gyermeknyelvi Korpuszban.....   | 216 |
| <b>45. táblázat</b> Az ILYEN használati mintázatai a többelemű lexikai egységekben .....   | 217 |
| <b>46. táblázat</b> Az OLYAN használati mintázatai a többelemű lexikai egységekben.....  | 218 |
| <b>47. táblázat</b> Az ILYEN és az OLYAN mellett megjelenő leggyakoribb főnevek .....  | 219 |
| <b>48. táblázat</b> Az ILYEN és az OLYAN mellett megjelenő leggyakoribb melléknevek, határozószók, névmások, kötőszók .....                                    | 220 |
| <b>49. táblázat</b> Az én is beágyazódásai hosszabb lexikai egységekbe a KorSzak Gyermeknyelvi Korpuszban.....   | 227 |
| <b>50. táblázat</b> Az így vagyok és az úgy vagyok kollokációi a KorSzak Gyermeknyelvi Korpuszban.....   | 228 |
| <b>51. táblázat</b> Szünetek helye az én úgy vagyok (vele) kimondása közben .....  | 230 |
| <b>52. táblázat</b> Az én is így vagyok ezzel / vele és az én úgy vagyok (vele) lexikai egységeket követő kötőszók és előfordulásaiknak százalékos aránya..... | 231 |
| <b>53. táblázat</b> Szünetek helye a hogy mit tudom én kimondása közben.....   | 233 |
| <b>54. táblázat</b> Szünetek helye a vagy mit tudom én kimondása közben .....  | 233 |
| <b>55. táblázat</b> Szünetek helye a mit tudom én kifejezést tartalmazó egyéb lexikai egységekben kimondása közben .....                                       | 233 |
| <b>56. táblázat</b> Vélemény kérésére és vélemény kifejezésére szolgáló lexikai egységek a KorSzak Gyermeknyelvi Korpuszban .....                              | 238 |

|  |     |
|--|-----|
| <b>57. táblázat</b> A lexikai egységek szövegben betöltött szerepük szerinti kategóriáinak összefoglaló táblázata..... | 243 |
| <b>58. táblázat</b> A Magyar felfedező 4. fejezetének lexikogrammatikai tanmenetének részlete                          | 251 |
| <b>59. táblázat</b> A projekt elkészítésének lépései.....  | 266 |

## **English summary**

### **From corpus building to teaching material development**

#### **Corpus linguistic studies and the use of the results in teaching material for children**

Doctoral thesis (PhD)

### **1. Introduction**

Corpus linguistics research has received increasing attention in recent decades and is being used more and more widely in various fields of applied linguistics. These analyses, which rely on large amounts of data, are of great use in research to observe the natural use of language.

At the same time, the use of corpora is not limited to linguistics alone, and their importance in language teaching has also become apparent. Among other reasons, by learning about the usage patterns they contain, students can learn linguistic structures that native speakers use in their everyday communication. Using real-life language examples, they can better understand how and in what context certain language elements are used. It will also make their use of language more natural and accurate.

The corpus provides access to a wide range of real language data, and the observation and empirical analysis of which helps to create teaching materials that are more relevant to language learners in terms of their subject matter and the texts they contain reflect natural language use. Such materials also contribute to language learners' motivation and can make the learning process more attractive. For the above reasons, using corpora and corpus-informed materials can also benefit children's language learning.

Applied linguistics research, including corpus linguistics research, aims to use research results to solve real-life problems. Accordingly, my objective from the beginning of my doctoral studies was to conduct research that could be useful in language teaching and produce visible outcomes. Consequently, corpus linguistics research aims to present its results in the form of Hungarian as a foreign language teaching material that can help both children and language teachers.

## **2. Hypotheses of the thesis**

The doctoral thesis is assumed to contribute to corpus linguistics and corpus-based research on agglutinating languages. Research in applied linguistics, particularly corpus linguistics, is conducted most commonly in English, including the use of formulaic language, which is the subject of the present research. This research can also serve as an example of how research methods and paradigms developed for English can be reconciled and applied to studying an agglutinating language.

Research on children's language use focuses on language acquisition and development. However, curriculum-development-oriented research on using formulaic units in the language of 11–15-year-olds has not yet been carried out in Hungarian. Therefore, the present research can significantly contribute to the extension of such studies. In addition, the development and description of building the pedagogical corpus presented in this thesis may support further research.

From the point of view of language teaching, it is assumed that relying on our intuitions and knowledge of the language is insufficient to create a pedagogical corpus that reflects the natural language use of competent speakers. Furthermore, for a language textbook to be effective, it needs to be based on corpus-based research. In order to achieve this, we must first build a pedagogical corpus specifically for the purpose of creating a curriculum. Before creating teaching materials, it is paramount to observe the language use of competent speakers of a similar age to the target group and learn about their preferred topics of conversation, which may also be relevant to language learners. The selection of relevant topics is crucial to maintain language learners' motivation, in this case, children, and to ensure an effective learning process. Using research methods that meet the above assumptions can contribute to creating teaching material that reflects natural language use and is adapted to language learners' needs.

## **3. Research objectives**

In the corpus-based research presented in this thesis, I aim to observe and investigate the natural language use of native Hungarian-speaking children in order to produce a corpus-informed teaching material for children learning Hungarian. Beside drawing on Hoey's (2003, 2004,

2005) theory of lexical priming, the research will focus on formulaic lexical units (cf. Wray, 2002, 2019) that appear in children's spontaneous speech and are of paramount importance in first language acquisition and are also very useful for language learners. The research aims to map the overall use of formulaic units in the speech productions of the children who provided the data and to find out the usage patterns of lexical units specific to their language use in the focus corpus.

The present research relied on a pedagogical spoken language corpus, the KorSzak Child Language Corpus, which was specifically designed and created for the purpose of curriculum development by the observation and analysis of children's language use, and thus the description of the process of corpus construction, which started in 2020, is an important part of the thesis. However, it should also be noted that corpus building is not the aim of the research but merely a tool to use the results of the underlying research to produce effective teaching material for children. It is also important to note that in teaching Hungarian as a foreign language and teaching other languages, the production of teaching materials based on corpus-based research on children's language use is rare. There is also a need for pedagogical corpora that could be used in teaching children. Although there are pedagogical corpora specifically designed for children (e.g. SACODEYL, CorpusMate), there are either no or only limited worksheets available for language teaching using these corpora. The texts in the corpora have not been didacticised and have not been used to produce complex teaching materials.

#### **4. Theoretical background**

The research aimed to identify the lexical units and their usage patterns in children's spontaneous speech useful for language learning. The theoretical framework was built around five pillars.

The first pillar is the usage-based language theory models (cf. Tomasello, 2003, 2013, 2015; Bybee, 2008; Ortega et al., 2016; Ellis et al., 2016; Pelcz, 2023), which mainly focus on children's language acquisition process, but their relevant findings should also be taken into account in language learning. According to usage-based language theories, language development results from associations formed while using the language. Such associations consist of available examples of language use that can be accessed more quickly and efficiently from memory by repeated use. To this end, children – and language learners – need to be



provided with opportunities for frequent encounters and a complex language environment (Tomasello, 2013).

The second pillar is the theory of lexical priming (cf. Hoey, 2005; Pace-Sigge, 2013). The main idea of the theory developed by Hoey (2005) is based on the observation that certain words and word forms promote or prevent the occurrence of each other. According to the theory, everything we hear, read, write or say reinforces how we use words or how we do not. When we come across a word, our mind automatically starts to form associations with it. Over repeated encounters, previously formed associations may be strengthened or weakened (Hoey, 2005: 9).

The third theoretical pillar of the research is formulaic language use (cf. Wray, 2002, 2013). Corpus linguistics research has shown that competent language users rely heavily on formulaic language use in their speech production (Wray, 2002; Wray - Fitzpatrick, 2008; Wray, 2019) and do not take advantage of the unlimited variation possibilities theoretically provided by grammar (Pawley - Syder, 1983). Such formulaic units help facilitate comprehension and make communication rapid and thus contribute to both productive and receptive fluency.

The fourth pillar is theories related to fluency. These include Segalowitz's framework, the main models of speech production (Levelt, 1999; Kormos, 2006), Tavakoli and Hunter's pyramid model (Tavakoli, 2020), and the understanding of the concept of automaticity.

Finally, the fifth pillar of the research is the corpus-based approach (cf. Szirmai, 2005; Stefanowitsch, 2020; Hunston, 2022), which is based on the idea that empirical research using large amounts of linguistic data is essential for the analysis of linguistic phenomena observable in natural language use.

## **5. About the Association for Corpus Linguistics and Didactics (KorSzak)**

The KorSzak Child Language Corpus is part of the corpora edited by the Association for Corpus Linguistics and Didactics (KorSzak), which was founded in February 2020 to bridge the gap between corpus linguistics research and the methodology of teaching Hungarian as a foreign language (Baumann et al., 2020, p. 32) The founding members recognised that corpus-based empirical research findings are often not integrated into language teaching methodology, which is essential for sound methodological development and successful language teaching. Indeed,

corpus research has many advantages in language teaching and curriculum development. It provides insights into the frequency of use of language items, the typical contexts of their use and the grammatical features associated with them (cf. Hoey, 2005). This data thus allows us to make more informed decisions about natural language use rather than relying solely on our intuition (Hoey, 2005; Hunston, 2002; O'Keeffe et al., 2007). The Association for Corpus Linguistics and Didactics aims to contribute reliable, evidence-based information to the work of language teachers, learners and curriculum planners. In order to achieve these goals, several sub-working groups have been formed to produce a variety of corpora (Baumann et al., 2020, pp. 32–33), such as:

1. spontaneous speech corpus led by Pelcz Katalin,
2. learner corpus led by Tímea Baumann,
3. child language corpus led by Boglárka Vermeki.

After the inaugural meeting of the KorSzak working group in February 2020 at the International Education Centre of the University of Pécs, the creation of the corpora mentioned above started almost immediately. Since then, extensive data has been collected, and the corpus is continuously expanded.

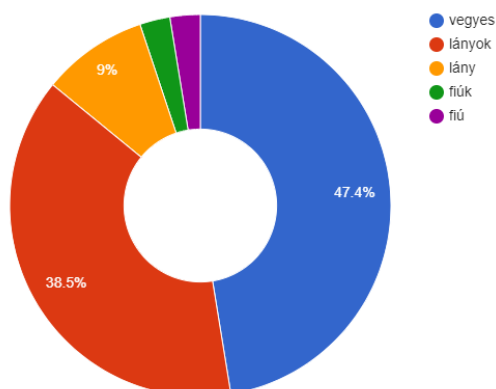
### **5.1 KorSzak Child Language Corpus**

The KorSzak Child Language Subgroup aims to build a pedagogical corpus that deals with a variety of relevant and interesting topics for children. Its primary objective is to provide a corpus that can be used to produce teaching materials for children learning Hungarian as a foreign or heritage language (Baumann et al., 2020, p. 33).

Since the objective of corpus building is to produce corpus-informed teaching materials, the KorSzak Child Language Corpus is dynamic in the way it is sampled and a specialised corpus for pedagogical purposes in the way it is used. In the present doctoral thesis, the notion of pedagogical corpus is interpreted as a set of authentic texts for language teaching. The texts in the corpus have been consciously compiled with curriculum development as the primary goal. The corpus construction was designed solely on the basis of the definition of topics, which did not affect the linguistic devices used by the children in the conversations so that natural language use is well observed in the corpus. The choice of topics was made with the

involvement of the children. The corpus is constantly growing. Currently, in October 2022, it contains seventy-eight records of twenty-nine children aged 11–15 (nine boys and twenty girls).

**102. Figure** Proportion of girls and boys in the KorSzak Children's Language Corpus (October 2022)



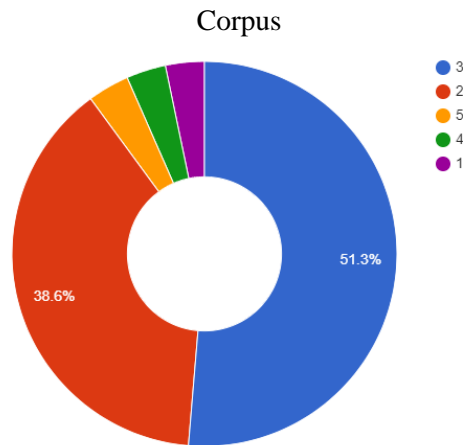
The KorSzak Child Language Corpus currently contains more than 71,000 tokens as of October 2022. The video and audio recordings feature children in pairs or small groups freely discussing a topic. The task's duration was not specified, so there are 1-2 minutes and much longer recordings of up to 45 minutes using the Zoom video conferencing programme. The corpus consists of sixty-eight dialogues and ten monologues.

**60. Table** Types of texts in the KorSzak Child Language Corpus

| Type      | Number of recordings | Number of tokens |
|-----------|----------------------|------------------|
| Dialogue  | 68                   | 69101            |
| Monologue | 10                   | 2445             |

The monologues were first produced in 2021 because of the pandemic. Later, the corpus was enriched with additional recordings because of the introduction of the topic of a description of a person. During the dialogues, 2-5 respondents talk to each other about the topic. A high proportion of the recordings feature two or three children talking to each other. 3.3 per cent of them have four speakers, and 3.5 per cent have five.

**103. Figure** Distribution of the number of speakers per recording in the KorSzak Child Language

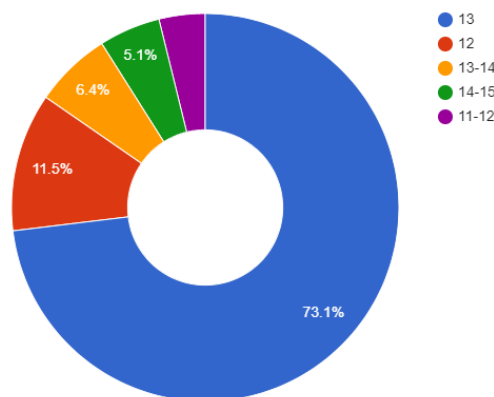


The twenty-nine respondents are classmates and friends, or just friends. Children in rural schools in Hungary. 64.1% of the recordings were made in Pest County and 35.9% in Baranya County. The first recording phase, in 2020, mainly involved schoolchildren from Baranya County. It is when the longer recordings were made, so if we look at the coverage of the number of tokens, 84.1% of the corpus is from Baranya County.

Most of the children are monolingual Hungarian native speakers, two children (respondents 5 and 16) have a father of non-Hungarian origin (Dutch, Italian) but communicate in Hungarian at home. In the upper primary school level, the reporting children study English (83%) or German (17%) as a foreign language.

Most records are from the age group of 13, but the age distribution of the children in the corpus ranges from 11 to 15 years. There may be age differences in the age of classmates in some records, and there are also respondents who provided samples more than once over the three years.

**104. Figure** Age distribution of children included in the KorSzak Children's Language Corpus



## 6. Research questions

The corpus-based research presented in the doctoral thesis (cf. Szirmai, 2005; Stefanowitsch, 2020; Hunston, 2022) is thus based on the KorSzak Children's Language Corpus, and its main objective is to map the most frequent formulaic units in the language use of the children who provide the data (cf. Wray, 2002; Wray - Fitzpatrick, 2008; Wray, 2019) to study the usage patterns of these lexical units (cf. Hoey, 2005) and to use the results to produce corpus-informed Hungarian language teaching materials for children. Observing lexical units in the corpus is necessary because these frequent lexical units play an important role not only in language acquisition but also in language learning (cf. Tomasello, 2003, 2013, 2015; Bybee, 2008; Ortega et al., 2016; Ellis et al., 2016; Pelcz, 2023), since their knowledge and appropriate use make our language use more fluent (cf. Tavakoli, 2020).

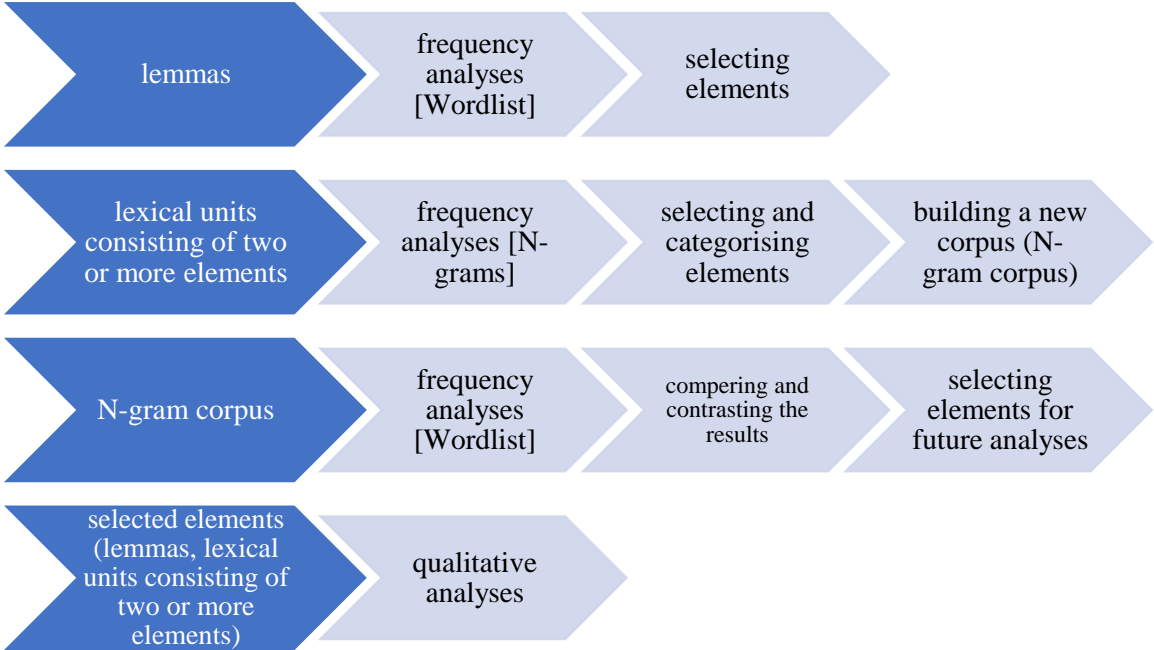
To achieve this goal, the following research questions are addressed:

1. What are the most common lemmas in the KorSzak Child Language Corpus?
2. What is the part of speech (noun, adjective, verb) distribution of the most frequent lemmas in the KorSzak Child Language Corpus?
3. What are the most frequent two- or multi-word lexical units in the KorSzak Child Language Corpus?
4. How do children in the KorSzak Child Language Corpus use the selected lexical units? What usage patterns can be observed in the corpus?
5. How are the two- or multi-word lexical units nested in longer sequences? What usage patterns do the nested lexical units show?
6. What is the role in the text of the two- and multi-word lexical units in the KorSzak Child Language Corpus?

The research on the KorSzak Child Language Corpus thus focuses on the above six research questions. Corpus-based research has traditionally used a quantitative research methodology. Its primary aim has been to provide linguists with numerically verifiable data beyond the limits of human intuition (see Stefanowitsch, 2020, pp. 15–17). However, in order to correctly interpret and effectively use the results obtained, it is also necessary to apply qualitative

research methodology (Martinez, 2019, p. 211). With this in mind, in order to achieve a more comprehensive picture, the research was based on an explanatory sequential typology (cf. Dörnyei, 2007), in which qualitative analyses followed quantitative analyses. The research consisted of several sequences, which started from two specific points: the analysis of lemmas and the analysis of lexical units consisting of two or more words.

**4. Figure** Quantitative and qualitative analyses in the research



**7. The conditions set out in this research**

**7.1 Frequency**

Taking into account Biber's (2006) study, Dóla's (2016, 2018) research on Hungarian and Mersinli – Aksan's (2016) research on Turkish, I have defined the frequency indicators for the selected lexical units of the present research. For the absolute frequency indicator, a minimum of 100 occurrences per million words was required for two-grams, 14 occurrences per million words for three-grams, and at least 4 occurrences per million words for four-grams to be selected (Vermeki, 2023b, to appear).

| Type of lexical unit | Minimum frequency index           |
|----------------------|-----------------------------------|
| 2-gram               | 100 occurrences / 1 million words |
| 3-gram               | 14 occurrences / 1 million words  |
| 4-gram               | 4 occurrences / 1 million words   |

In addition to the above-mentioned absolute frequency indicator, two other indicators were considered for selecting lexical units: document frequency and relative document frequency. The first one shows the number of uploaded documents containing the items in the list. In contrast, the relative document frequency shows the percentage of documents containing the items in the list. I aimed to ensure that the results reflect the language use of different data providers and not limit the results to the language use of a single speaker. The number of items selected based on absolute frequency indicators was thus reduced as follows. The selected items had to appear in at least half of the documents for lemmas. Half of this number was determined in the case of a two-grams, i.e. they had to appear in at least a quarter of the documents. I specified that three-grams should appear in at least five documents, while I set this number at three for four-grams.

| Type of lexical unit | Minimum document frequency | Minimum relative document frequency (%) |
|----------------------|----------------------------|---|
| lemma                | 39                         | 50                                      |
| 2-gram               | 20                         | 25                                      |
| 3-gram               | 5                          | 6,4                                     |
| 4-gram               | 3                          | 3,8                                     |

## 7.2 Size of the formulaic lexical units

Of the two methods mentioned by Dóla (2016: 98) – the inclusion of smaller lexical units; the inclusion of verb conjunctions and suffixes in the definition – I chose the first one for determining the size of formulaic units. Therefore, the multi-word lexical units sought in the present research are composed of two or more words (see Biber’s (2006) three or more words).

However, following the suggestion of Siyanova-Chanturia and Pellicer-Sánchez (2019, p. 2), I do not exclude lemmas from the formulaic units either.

### 7.3 Function and Meaning

The lexical units selected based on frequency are classified according to their role in the text, as defined by Biber (2006).

**2. Table** Lexical units according to their role in the text (Biber, 2006, pp. 139–146) (Vermekei, 2023b, to appear)

| Main Category                | Function  | Subcategory   | Biber's examples   | KorSzak Child Language Corpus                                 |
|------------------------------|---|---|--|---|
| Stance bundles               | to give a framework, the speaker can express that something is certain, uncertain or possible, his or her attitude to the action, express wishes, obligations, intentions, predictions, abilities | epistemic terms   | <i>I don't know what, I think it was, are more likely to</i>                                   | <i>nem tudom, nem is tudom</i>                                |
|                              |   | lexical units expressing attitude, modality (desire, wish, instruction, intention, effort, ability, obligation) | <i>you need to know, I don't want to, you might want to, it is important to, to be able to</i> | <i>nagyon jó lenne, nekem az a kedvencem, én is szeretnék</i> |
| Discourse organising bundles | introducing the topic, achieving focus on the topic, and clarifying and elaborating on the topic  | topic introduction  | <i>want to talk about, going to talk about, if you look at</i>                                 | <i>az lenne a kérdésem, akkor ott van a</i>                   |
|                              |   | clarification of the subject  | <i>has to do with the, you know I mean, as well as the</i>                                     | <i>ez a, erről a, volt az az</i>                              |
|                              |   | focus   | <i>those of you who, that's one of the, one of the things</i>                                  | <i>az volt a, neki volt egy</i>                               |
| Referential bundles          | the speaker identifies something or an attribute of   | indication of inaccuracy  | <i>or something like that, and things like that</i>  | <i>vagy hogy mondjam, vagy nem is tudom,</i>                  |



| Main Category | Function                           | Subcategory                | Biber's examples  | KorSzak Child Language Corpus                            |
|---------------|------------------------------------|----------------------------|---|--|
|               | something they think is important. |                            |   | <i>vagy mit tudom én</i>                                 |
|               |                                    | identification             | <i>have a lot of, the size of the, the nature of the, in terms of the</i> | <i>az ilyen nagy(obb), de ő is nagyon +Adj.</i>          |
|               |                                    | indicating time and place  | <i>at the end of the year, as shown in figure</i>                         | <i>egy idő után, amikor kicsi voltam, akkor ott volt</i> |
|               |                                    | Multi-functional reference | <i>at the end of, the end of the</i>                                      | <i>az volt az első, és akkor így</i>                     |

## 7.4 Processing and Production

Among the selected multi-word lexical units, several are characterised by nesting (cf. Hoey, 2005). Thus, a smaller lexical unit form part of a larger one. For these units and the partially fixed formulaic units, where even the word order differs in their occurrence in the corpus, I examined where the children in the study paused during speech production. Among them, those items were selected for which children did not pause during the utterance of the unit in question, the pause taking place before and after the unit (Vermeki, 2023b, to appear).

## 8. Results

### 8.1 What are the most common lemmas in the KorSzak Child Language Corpus?

Based on the predefined criteria of the research, the first hundred of the lemmas with the highest frequency index were compared with the huTenTen12 corpus. The results show a difference between the lemmas with the highest frequency index in children's language use and the most frequent items in the huTenTen12 corpus. When comparing the top 20 most frequent items, it

was clear that of the four lemmas with the highest frequency differences, ILYEN had a higher score of 88%, NAGYON 80%, AKKOR 75% and TUD 71%. In the N-gram corpus constructed from the lexical units selected according to the criteria, we also noticed that all appear hundreds of times nested in the multi-word lexical units of the focus corpus.

Of the lemmas examined, several elements in the 20<sup>th</sup>–100<sup>th</sup> positions do not appear at all among the first hundred lemmas of the reference corpus. For example, 50% of the documents contain the lemma IGAZÁBÓL, the usage pattern discussed later in more detail. In the children's conversations, spoken language turns such as UGYE, TÖK, ÚGYHOGY, TÉNYLEG, NA, SZÓVAL, as well as AMÚGY, naturally appear. The multi-word lexical units selected from the focus corpus do not include the lemmas UGYE, INKÁBB, ÚGYHOGY, TÉNYLEG, NA, HISZ, SZÓVAL, AMÚGY, AZTÁN, as well as the lemma BESZÉL. It means that although they appear in the corpus as part of the multi-word lexical units, they are not present in significant numbers. Most of them instead form a single formulaic unit independently, functioning mainly as a discourse marker.

## **8.2 What is the part of speech (noun, adjective, verb) distribution of the most frequent lemmas in the KorSzak Child Language Corpus?**

The KorSzak Child Language Corpus contains 55.143 words, 17.180 nouns, 10.259 verbs and 933 adjectives.

Regarding nouns, the lemmas with the highest frequency index represent the topics of conversation (for example, animal, sport, movies, friends, family, hobbi, or the name of dogs, e.g., Bubu). Excluding these, and looking at the first fifty items, we have the lemma ÉSZ and ÉRZÉS, which do not appear at all in the first hundred hits of huTenTen12, and the lemma RÉSZ, BAJ, KÉRDÉS, which have slightly different frequency rates. Of the lemma listed, ÉSZ, ÉRZÉS and RÉSZ do not form part of any of the multi-word lexical units in the focus corpus and do not appear as lemmas per se in half of the documents, so their use is not presented in this research.

The frequency lists of adjectives show the biggest difference between the focus and the reference corpus. Among the adjectives, the most frequently appearing lemma in the KorSzak Child Language Corpus is JÓ, which ranks third in the reference corpus. Most of the adjectives used by the children do not appear in the top 100 lemma of huTenTen12. These include the following:

ARANYOS, BOLDOG, BÜDÖS, CUKI, CSODÁLATOS, CSODÁS, CSÚNYA, FURA, HÍRES, IDEÁLIS, IDŐS, IZGALMAS, KEDVENC, KREATÍV, MENŐ, NORMÁLIS, PICI, SZERENCSÉTLÉN, SZÖRNYŰ, VICCES.

In addition, the items in the huTenTen12 corpus, which also appear in the focus corpus, have a different and, in most cases, much lower frequency index (frequency per million words). These include, for example, ÚJ or UTOLSÓ, which are significantly less frequent in the texts. For the lemma ÚJ, I found almost 80% fewer occurrences in the KorSzak Child Language Corpus, while for the UTOLSÓ I recorded 60% fewer occurrences.

In the category of verbs, there are lemmas closely related to the subjects, like in the case of nouns, but to a lesser extent. For example, among the leisure-related verbs: SPORT, RAJZOL, SÉTÁL, ÉLVEZ, which are not among the 100 most frequent verbs in the reference corpus, are extremely useful for the preparation of teaching materials and for the selection of vocabulary for the topics. Verbs with prefixes such as ELMEGY, ELKEZD, MEGNÉZ or KIMEGY, which are used by child respondents many times in dialogues, do not appear in the reference corpus at all, nor does ESZIK, which is ranked 41<sup>st</sup> in the focus corpus, appear in the 100 most frequent verbs of the reference corpus. Unlike the adjectives, this category is characterised by the fact that verbs that appear in both corpora and are not in the same position in the word frequency list have a higher frequency index in the focus corpus. These are the verbs *megy* (72% higher), *CSINÁL* (81% higher) and *EMLÉKSZIK* (78% higher), already mentioned in the previous chapter.

### **8.3 What are the most frequent two- or multi-word lexical units in the KorSzak Child Language Corpus?**

Among the lexical units selected based on the set of criteria, the largest category is that of three-grams (120 items), followed by two-grams (60 items) and finally, four-grams (20 items). A total of 200 lexical units were selected, of which the ten with the highest frequency index are shown in the list below.

- Two-grams: *nem tudom, és akkor, hogy a / az, én is, nem is, az a, egy ilyen, az ilyen, meg a, és a.*
- Three-grams: *nem is tudom, mit tudom én, vagy nem tudom, vagy hogy mondjam, ott van a, de nem tudom, és akkor így, mi a kedvenc, hát nem tudom, van egy ilyen.*

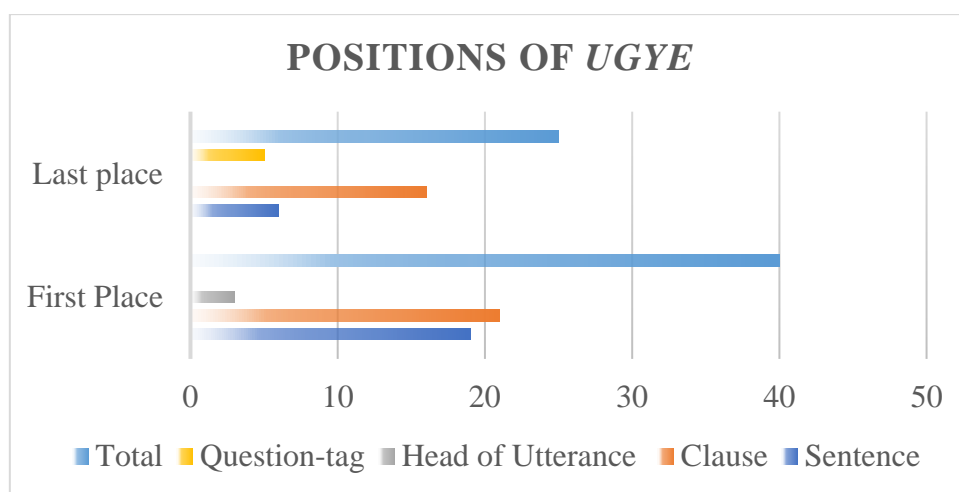
- Four-grams: *vagy nem is tudom, én is így vagyok, hogy mit tudom én, nagyon jól éreztük magunkat, én is egyet tudok, én úgy vagyok veled, vagy mit tudom én, az volt az első, az nagyon jó volt, de ő is nagyon.*

## 8.4 How do children in the KorSzak Child Language Corpus use the selected lexical units? What usage patterns can be observed in the corpus?

### 8.4.1 Usage patterns of *ugye*

*Ugye* appears in the corpus with four different functions. The first is when children use the word as a time-filler while thinking. The following function is reinforcement, when the speaker is confident that the listener agrees with them, assuming some shared background knowledge or understanding. The third function is the confirmation request, where the speaker asks for confirmation that the statement they are referring to is true. We can also observe the positions taken in the utterances. A significantly high percentage (71%, 166 occurrences) of the corpus appears as the second or third element of a sentence or clause. Of the remaining 65 occurrences, it appears in the position of a sentence or clause head 40 times and at the end of a sentence or clause 25 times. Children use the latter as a question tag in 5 cases. Regarding colligation, it is generally (95%) used in declarative sentences, and it occurs 10 times out of 231 occurrences in questions. Of these, a total of 6 occur as a follow-up question, of which five, as already mentioned, occur as the last element of the sentence, and one occurs in the penultimate position as a negative question tag.

5. Figure Positions of UGYE in the speech productions



## 8.4.2 Usage patterns of NEM TUD

The most frequent two-gram is from the lemma NEM TUD, which appears 292 times in the KorSzak Child Language Corpus. The most frequently occurring form is the first person singular definite form, *nem tudom* [‘I don’t know’], which accounts for 77.74% of the hits. Next, the already much lower hit rate is the indefinite form *nem tudok* [‘I cannot’], which appears 17 times, accounting for 5.82% of the hits. The other inflected forms have a significantly lower frequency than the first two hits. The *nem tudom* forms a unit in 84% of the hits (192 occurrences) or is part of more extended lexical units. Its occurrences are often followed by question words (Vermeki, 2023a). The concordance lines below provide some examples. In addition, there are 5 instances of the infinitive structure *nem tudom* + infinitive.

### 105. Figure Examples of *nem tudom* + question words in the KorSzak Child Language Corpus

| Left context   | KWIC                                    | Right context ↑   |
|--|---|---|
| és akkor hú, de menő, hogy ott állt, jött az állat a | <b>nem tudom</b><br>N.Nom V.Prs.Def.1Sg | <b>hány</b> négyzetméteres kennelbe, és akkor de menő, hogy levadászta, amikor n          |
| vesen gitározniék, és akkor még ott van a másik,     | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>hány</b> elfoglaltság, vagy éppen hegedüt kell gyakorolni, zongorát kell gyakoroln     |
| s>Ő biztos, hogy ott van, mert nem férne el az a     | <b>nem tudom</b><br>N.Nom V.Prs.Def.1Sg | <b>hány</b> tonna régiség, ami még a szüleié nagyszüleié volt, és azokat úgy őrizget      |
| m.</s><s>Így őszintén megmondom, hogy most           | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>hol</b> van, de biztos, hogy megvan még valahol. - És Bubu hogyan teljesített?</s><s>  |
| : YouTube-on, hogy bement valami kínai erdőbe,       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>hol</b> , és ott azt hiszem talált egy halottat és kiröhögte.</s><s>Pontosabban nei    |
| sen van megformálva.</s><s>Halálfaló.</s><s>         | <b>Nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>milyen</b> részben van először, de egy a Voldemort leghűségesebb szolgája, és e        |
| szeretem a meglepetéseket, mert idegesít, hogy       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>mi</b> az, de persze, amikor megkapom, örülök neki, és van, amikor jó érzés meg        |
| semelyik. - Hát én inkább kutya.</s><s>Igazából      | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> , de valahogy én a kutyákat talán jobban kedvelem.</s><s>Nem azt mor         |
| neg a macskát, pedig egyébként nem is.</s><s>        | <b>Nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> , de nem tudom.</s><s>Múltkor jöttem rá, hogy igazából annyira nem is        |
| tt.</s><s>Hát valamiért a szülei ellenezték, de      | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> . - Nem tudom, én nem úgy talán azért félek háziállatot tartani, mert ha n   |
| csótányok, illetve a poloskától, így undorodom,      | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> így, nem azoktól is, mint a csótányoktól. - Én a kigyóktól félek, mert ha v: |
| </s><s>Tessék. - Nekem a Grindelwald.</s><s>         | <b>Nem tudom</b><br>Adv V.Prs.Def.1Sg   | <b>miért</b> , de a Grindelwald. - De miért? - Nekem azért tetszik a Grindelwald karal    |

### 106. Figure Examples of *nem tudom* + Inf. in the KorSzak Child Language Corpus

CONCORDANCE

KorSzak Gyermekegyelvi Korpusz

simple *nem tudom* • 227  
3,172.78 per million tokens • 0.32%

filter ( [word="abbahagyni"] ) ( [word="kikefélni"] ) ( [word="el... • 5  
69.89 per million tokens • 0.007%

Sort word x

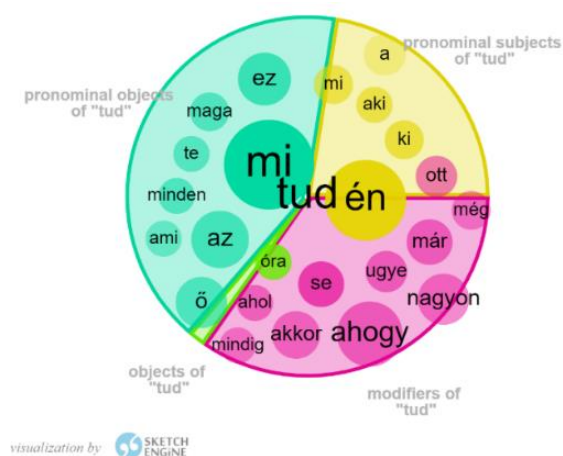
Details

|   | Left context   | KWIC                                  | Right context ↑   |
|---|--|---------------------------------------|---|
| 1 | doc#65 dom, én egy éve nézem körülbelül, és már hatszor végignéztém az egészet és        | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>abbahagyni</b> .</s><s>Nekem ez a kedvenc sorozatom. - Nekem nem nagyon v            |
| 2 | doc#66 .</s><s>Hát szerintem azért, mert úgy hogyha elkezdünk egy sorozatot, akkor       | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>abbahagyni</b> , mivel izgalmas. - Hű ezzel én is így vagyok.</s><s>Én függővé       |
| 3 | doc#0 k a szarvashúst étteremben.</s><s>Ez így, hát így határeset.</s><s>Hogy így        | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>eldönteni</b> , hogy ez most így jó vagy nem. - Meg ugye te is beszéltél arról, hog  |
| 4 | doc#5 .tutya a legjobb, vagy legideálisabb háziállat, mert én egy macskáról egyáltalán   | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>elképzeln</b> , hogy ő akkor most szuperhős, meg és megment, hanem inkább €          |
| 5 | doc#6 jttje, hogy le kell nyírni a szürét. - Igen.</s><s>A Rozival az a baj, hogy nagyon | <b>nem tudom</b><br>Adv V.Prs.Def.1Sg | <b>kikefélni</b> , mert ugye ilyen nagyon. 5 percig nem bír nyugton lenni, de a Benit i |

The second most frequent form is *nem tudok*, with 17 occurrences in the corpus. The unit *nem tudok* is most often followed by an infinitive (10 occurrences), either immediately after it (e.g. *nem tudok + beszélni / hegedűlni / leállni / síelni / választani*) or after a nested *mit*.

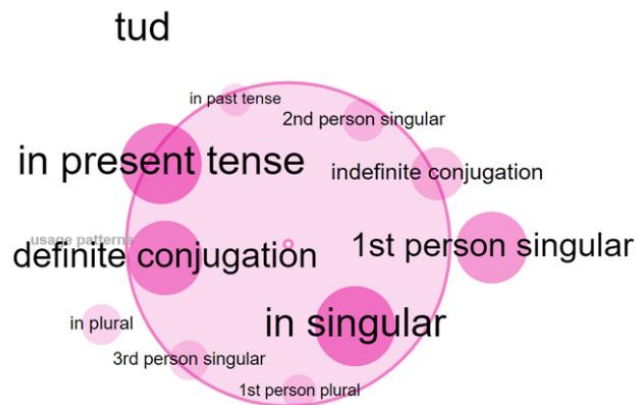
It is also worth comparing the results for NEM TUD with those for the lemma TUD to draw further conclusions about its use. The lemma TUD is the 13<sup>th</sup> most frequent lemma, appearing 650 times in the corpus. Out of the 650 hits, it is part of a multi-word lexical unit 591 times in 90% of the hits, of which 346 times it appears in the form *tudom* [‘I know’], and as we have seen above, 227 times of these, in 65% of the hits, it appears in the structure *nem tudom* [‘I do not know’]. The figure below, which shows the collocations of TUD, also shows that the TUD lemma is most closely related to the first-person singular subject (*én*) and the subject MI (Vermeki, 2023a).

**6. Figure** Collocations of TUD (Vermeki, 2023a, p. 120)



Considering the colligation of the TUD lemma, its use is associated with limited grammatical patterns. The most frequent is present tense, definite conjunction in the first person singular, the form *tudom* (frequency index 346). It is much less frequent in other grammatical patterns. In the children's dialogues, the present tense, indefinite, first-person singular (*tudok*) is also found repeatedly, much less frequently in the second person (*tudsz*) and third person (*tud*) and its only 27 times in the past tense (*tudtam, tudta*). We can see, therefore, that the results obtained when examining the TUD lemma are similar to those obtained when examining the NEM TUD two-gram (Vermeki, 2023a).

**7. Figure** Colligations of TUD (Vermeki, 2023a, p. 120)



In the KorSzak Child Language Corpus, the lexical unit *nem tudom* has four functions. In addition to expressing a lack of knowledge, the use of *nem tudom* can be used to avoid questions, as a time-filler or to let the interlocutor know that you have heard and understand what he or she is talking about.

### 8.4.3 Usage patterns of MI és TUD

Most lemmas, two-grams and three-grams with high-frequency indicators in the KorSzak Child Language Corpus are nested in a more extended lexical unit. For example, if we go back to the lemma TUD and look in more detail at the collocations it shares with the lemma MI, which is its most frequent collocation, we observe that the most frequent lexical unit in which it appears is the three-gram *mit tudom én*, which is the second most frequent three-gram in the corpus.

8. Figure Concordance of the three-gram *mit tudom én*

| Left context   | KWIC ↑                        | Right context   |
|--|-------------------------------|---|
| nvalóan a haszonállatokat azokat munkára tartják, <b>mit</b>       | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , ígás lovat szállításra használnak, és. - Vagy h |
| gy mennyi ideje van vele foglalkozni, mármint egy <b>mit</b>       | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , egy hörcsög nem feltétlenül igényli, hogy fogl  |
| z állnak a ketrecen, és sosem látják a napot, vagy <b>mit</b>      | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , Magdi néni udvarán kaparászik a csirke, és b    |
| most egy jobb gazdája is, ha nem hoztam el volna <b>mit</b>        | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , a menhelyről, vagy bármi, és igazából én talá   |
| ít tartani. </s><s>Ilyen kis pici kutyát. </s><s>Ilyen, <b>mit</b> | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> pomerániai törpe spicc, az még úgy elmegy tal     |
| állatok ezrei, vagy direkt nem szeretik az állatokat, <b>mit</b>   | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , most tudok egy ilyen példát, hogy nem mond      |
| <or már elegend van belőle, akkor egyszerűen csak <b>mit</b>       | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , kiviszed az országútra, és otthagyt, de el se   |
| az oroszlán levadássza a leggyengébb gnút, vagy <b>mit</b>         | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , zebrát akkor, nem arra törekszik, hogy az az :  |
| <s>Igen. - Meg ez is, hogy mindenki mondja, hogy <b>mit</b>        | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , az emberek mennyivel intelligensebbek az áll    |
| és tudod, ez a mennek a reptéren, és akkor jön a <b>mit</b>        | <b>tudom</b><br>V.Prs.Def.1Sg | <b>én</b> , a vakvezető kutya, éppen ő is megy föl a gép    |

Out of 46 hits, it appears in this unit 38 times or 82%. We can also see that MI + TUD is in the present tense in all 46 occurrences, of which 38 occur in the definite conjugation and 8 in the indefinite conjugation. It is found in three persons in the corpus: first of all, in the first-person singular (40 occurrences), in the first-person plural (4 occurrences) and finally, in the second-person singular (2 occurrences).

**3. Table** Appearances and usage patterns of MI + TUD in the KorSzak Child Language Corpus

| <b>Lexical unit</b>                           | <b>Hits</b> | <b>Frequency of forms (%)</b> |
|---|-------------|-------------------------------|
| <b>MI + TUD</b>                               | <b>46</b>   |                               |
| <b>MI + TUD + Def: 1Sg</b>                    | <b>38</b>   | <b>82</b>                     |
| mit tudom én                                  | 29          | 63                            |
| hogy mit tudom én                             | 5           | 10.86                         |
| vagy mit tudom én                             | 4           | 8.69                          |
| <b>MI + TUD + NDef: 1Sg (+ Adv/Pro) + Inf</b> | <b>2</b>    | <b>4.34</b>                   |
| hogy mit tudok belőle kihozni                 | 1           | 2.17                          |
| mit tudok még erre mondani                    | 1           | 2.17                          |
| <b>MIT + TUD + NDef: 1Pl (+ Adv) + Inf</b>    | <b>4</b>    | <b>8.69</b>                   |
| nem tudunk mit csinálni                       | 1           | 2.17                          |
| nem igazán tudunk mit csinálni                | 1           | 2.17                          |
| hát nem tudunk mit csinálni                   | 1           | 2.17                          |
| már nem tudunk mit szidni                     | 1           | 2.17                          |
| <b>MIT + TUD + NDef: 1Sg + Inf</b>            | <b>2</b>    | <b>4.34</b>                   |
| nem tudsz mit kezdeni                         | 1           | 2.17                          |
| nem tudsz mit mondani                         | 1           | 2.17                          |

The table above shows that the most common of the co-occurrences of the lemmas MI + TUD is *mit tudom én*, with three variations. In the first 29 occurrences, it forms a lexical unit and is used to express imprecision, doubt or uncertainty or in a time-filling function by the children in the corpus.



#### 8.4.4 Usage Patterns of *annyira nem*

The lexical unit *annyira nem* has a much lower frequency index than the unit mentioned above, *nem tudom*. With 39 occurrences in the corpus and 28% of the documents contain these two words side by side. On the other hand, if we look at the collocations of *annyira*, we can see that *nem* appears 81 times in total in the corpus, not directly next to it, but close to it.

**4. Table** Occurrences of *annyira nem* in the KorSzak Child Language Corpus

| Lexical unit | Hits | Frequency of forms (%) |
|--------------|------|------------------------|
| ANNYIRA NEM  | 39   |                        |
| annyira nem  | 39   | 100                    |

**5. Table** Co-occurrences of ANNYIRA and NEM in the KorSzak Child Language Corpus

| Rank | Word | Co-occurrences (-5 /+5) | Total occurrences in the corpus | LogDice |
|------|------|-------------------------|---------------------------------|---------|
| 1    | nem  | 81                      | 1412                            | 10.83   |

In terms of its occurrences in the corpus, *annyira nem* is most often followed by a verb (*annyira nem szeretem*, *annyira nem támogatom*) or an adjective (*annyira nem egészséges*, *annyira nem ugrálós*), and appears in three different functions. The expression emphasises and reinforces negation in the KorSzak Child Language Corpus. With this function, it occurs in 41% (16 occurrences) of the appearances in the corpus. The other function, which is frequent, is the expression of opinion, where the opposite of the first function occurs. The children use this function in 59% of the appearances in the corpus. Finally, the third function of *annyira nem* is comparison, which appears a total of one time in the corpus.

#### 8.4.5 Usage patterns of ILYEN és az OLYAN

The lemmas ILYEN and OLYAN in the KorSzak Child Language Corpus have significant frequency rates, which means that children regularly use them in conversation. These lemmas

also occur as part of several multi-word lexical units. A comparative analysis of the corpus revealed that the children use the lemma ILYEN much more frequently, while the lemma OLYAN is used more widely. The lemma OLYAN is found in 57 documents and 57 conversations. Because of their frequent use and versatility, these words and the formulaic expressions in which they appear are essential building blocks of spontaneous speech. They are, therefore, of great importance for language teaching.

**6. Table** Appearances of the lemmas ILYEN és az OLYAN in the KorSzak Child Language Corpus

| Rank | Lemma | Freq. | Freq. / million | Doc. freq. | Relative Doc. freq. % |
|------|-------|-------|-----------------|------------|-----------------------|
| 1    | ILYEN | 541   | 7,561.57        | 49         | 62.82                 |
| 2    | OLYAN | 298   | 4,165.15        | 57         | 73.08                 |

The comparative analysis also showed that ILYEN is part of a multi-word lexical unit selected for analysis from the corpus in almost half of the hits (266 occurrences). In addition, it appears several times before adjectives and nouns. Among the adjectives following ILYEN, the most frequent are *kis* (16 occurrences), *nagy(obb)* (15 occurrences), *kicsi, jó, érdekes* (5 occurrences), *menő, vicces, kreatív* (4 occurrences) and *furcsa* (3 occurrences). Among the nouns, the most frequent are *szempont* (8 occurrences), *dolog* (7 occurrences), *ember* (5 occurrences) and *hobbi* (3 occurrences).

Like the lemma ILYEN, OLYAN appears several times before adjectives and nouns. Among the adjectives, *jó* (13 occurrences), *sok* (8 occurrences), *nagy* (5 occurrences), *vicces* (4 occurrences) appear most often next to it. As a noun, it is followed by *ember* (12 occurrences), *dolog* (6 occurrences), *kutya* (5 occurrences), *tulajdonság* and *sorozat* (4 occurrences). As can be seen from the list, some words tend to occur with ILYEN, and some tend to occur with OLYAN.

**7. Table** Nouns appearing with ILYEN and OLYAN

| Appearing only with ILYEN | Appearing with both ILYEN and OLYAN | Appearing only with OLYAN |
|---------------------------|-------------------------------------|---------------------------|
| szempont 8                | dolog 18                            | nap 2                     |
| mindenféle 4              | videó 10                            | szereplő 2                |
| telefon 4                 | sorozat 7                           | fajta 2                   |

|       |   |             |    |    |             |   |
|-------|---|-------------|----|----|-------------|---|
| edzés | 4 | kutya       | 12 | 8  | varázsvilág | 2 |
| játék | 4 | rész        | 4  | 4  | kreativitás | 2 |
| hobbi | 4 | ajándék     | 2  | 2  | terület     | 2 |
| csoda | 3 | tulajdonság | 4  | 4  | helyzet     | 3 |
| alap  | 3 | állat       | 3  | 4  | fej         | 3 |
| smink | 3 | sport       | 2  | 3  | szint       | 4 |
| kép   | 3 | kérdés      | 2  | 3  | helyszín    | 5 |
| tanár | 3 | ember       | 7  | 12 |             |   |

**8. Table** Adjectives, adverbs, pronouns and conjunctions appearing with ILYEN and OLYAN

| <b>Appearing only with ILYEN</b> | <b>only</b> | <b>Appearing with both ILYEN and OLYAN</b> | <b>Appearing only with OLYAN</b> |   |          |   |
|----------------------------------|-------------|--|----------------------------------|---|----------|---|
| inkább                           | 7           | hogy                                       | 13                               | 6 | igazából | 2 |
| akár                             | 2           | mindig                                     | 2                                | 2 | még      | 4 |
| teljes                           | 2           | már  | 2                                | 6 | egyszer  | 2 |
| sok                              | 2           |  |                                  |   | mégis    | 2 |
| csak                             | 5           |  |                                  |   | konkrét  | 2 |
| tényleg                          | 2           |  |                                  |   |          |   |
| ahol                             | 2           |  |                                  |   |          |   |
| nagyon                           | 7           |  |                                  |   |          |   |
| például                          | 2           |  |                                  |   |          |   |
| ugye                             | 2           |  |                                  |   |          |   |
| amikor                           | 2           |  |                                  |   |          |   |
| ott                              | 2           |  |                                  |   |          |   |

In the analysis, we have highlighted the multi-word lexical units VAN EGY ILYEN (20 occurrences) and VAN EGY OLYAN (11 occurrences), as both play an important role in children's dialogues. VAN EGY ILYEN appears in the corpus in two different functions. This unit is a word combination often used in spontaneous speech, usually when telling a story or anecdote. The primary function of this multi-word lexical unit is to highlight or prepare something (Vermeki, 2023b, to appear). However, the children in the corpus also use the term a few times to indicate inaccuracy. The other multi-word lexical unit VAN EGY OLYAN occurs in a total of three functions in the corpus. The expression VAN EGY OLYAN, like the unit VAN EGY ILYEN, is also a feature of the narrative speech of the respondents. However, in contrast to the previous expression, this unit is used by children when they want to name a specific thing or situation but either cannot remember its exact name or do not want to say it (Vermeki, 2023b, to appear).

(doc#07 Mon\_Állatok\_Kutya\_004: Children talk about their memories connected with dogs.)

- (1) A: Ő eléggé hírhedt lett ott, és ő egy olyan kutyas, aki igazából ő is bichon, kicsi, de nagyon aktív és nagyon ilyen kis bolondos kutya. Ezért szerették őt nagyon a gyerekek.

(He has become quite notorious there, and he is a doggy who is actually a bichon, small but very active and very much the kind of silly little dog. So, the kids loved him very much.)

- B: Hát nekünk pont a szomszédban **van egy ilyen** kutyas, úgy hívják, hogy Maci. Ő egy nagyobb testű kutya, de nem tudjuk, hogy milyen fajta, keverék. És tényleg nagyon lökött, de én nagyon sokat játszok vele.

(Well, **we have a** puppy just next door called Maci. He is a bigger dog, but we don't know what kind of breed he is. And he's really crazy, but I play with him a lot.)

(KorSzak Child Language Corpus, 2020)

(doc#53 Mon\_Influenszerek\_001: Children talk about influencers. One of the speakers told a negative story about one of them. The other reacts.)

- (2) A: Ha **van egy olyan** influencers, akiről nem vagy jó véleménnyel, meg hülyeségeket posztol, akkor miért követed?

(If **there is an** influencer you don't like and they post stupid things, why do you follow them?)

(KorSzak Child Language Corpus, 2020)

#### 8.4.6 Usage patterns of *igazából*

The lemma *igazából* [actually] appears a total of 170 times in the corpus, 6 of which occur in HÁT IGAZÁBÓL ÉN the three-gram. It is used in functions by the children in the corpus. In a high proportion of its occurrences (46 times), it is used by children as an opening or introductory word, to introduce a statement or to confirm a statement. After conjunctions and *hát*, it also performs a similar function in the second position of a sentence or clause (82 occurrences).

Another function is to express uncertainty or doubt about a statement or situation. The third function is discourse-organising, whereby children use the word as a synonym for actually or in fact.

Regarding collocations, the *igazából* occurs more often in statements with negative connotations than in those with positive connotations in the corpus. It is often accompanied by the negation *nem* or *nincs* (56 occurrences).

### 9. Figure Examples of *igazából* + *nem*

| Left context   | KWIC                   | Right context     |
|--|------------------------|-------------------|
| :ettő.</s><s>Mindkettőt imádom. - Én semelyik. - Hát én inkább kutya.</s><s>       | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| tudom miért, de valahogy én a kutyákat talán jobban kedvelem.</s><s>               |                        |                   |
| szűnkbe jut. - Igen. -És Zsófi néni is mondta, hogy ő is olyat írt fel, hogy utána | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| is annyira szereti, csak az jutott eszébe.</s><s>És én is, ha én megkér            |                        |                   |
| </s><s>Olyan, mintha látná a jövőt. - Hát már nem sokáig.</s><s>Nem tudom,         | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| szeretnék ilyen lenni, meg mondjuk egy szünyog is hiába él mondjuk va              |                        |                   |
| mondjuk egy ilyen tudományos, hogy mondjuk, tehát az ilyen rosszindulatú, ami      | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| jó semmire, az mondjuk nem, de ami így nem árt az állatnak sem.</s><               |                        |                   |
| jes mértékben, tehát volt esélye arra, hogy túléljen. - Ez tök jó.</s><s>Nekünk    | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| volt madarunk, csak egyszer, de anyuék azzal vigasztaltak, hogy elrept             |                        |                   |
| g minden, már fű alig evett, meg alig ivott, az nagyon szörnyű volt.</s><s>Meg     | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| is tudtuk végül, hogy mi lett a baja.</s><s>Szóval ez a baj az ilyen állat         |                        |                   |
| </s><s>Úgy bent. - Hát nálam egy megint tapasztalatból mondott sztori.</s><s>      | <b>igazából</b><br>Adv | <b>nem</b><br>Adv |
| teljesen szilveszterkor történt, de egyszer elkezdtek tűzijátékozni nálun          |                        |                   |

When it is part of the three-gram, HÁT IGAZÁBÓL ÉN, it appears in the corpus with only one function, which is the last mentioned discourse organisation, the introduction of a topic. Since there are only 6 occurrences of this unit in the corpus, it is impossible to draw far-reaching conclusions. However, it can be concluded that in all these occurrences, children use it similarly.

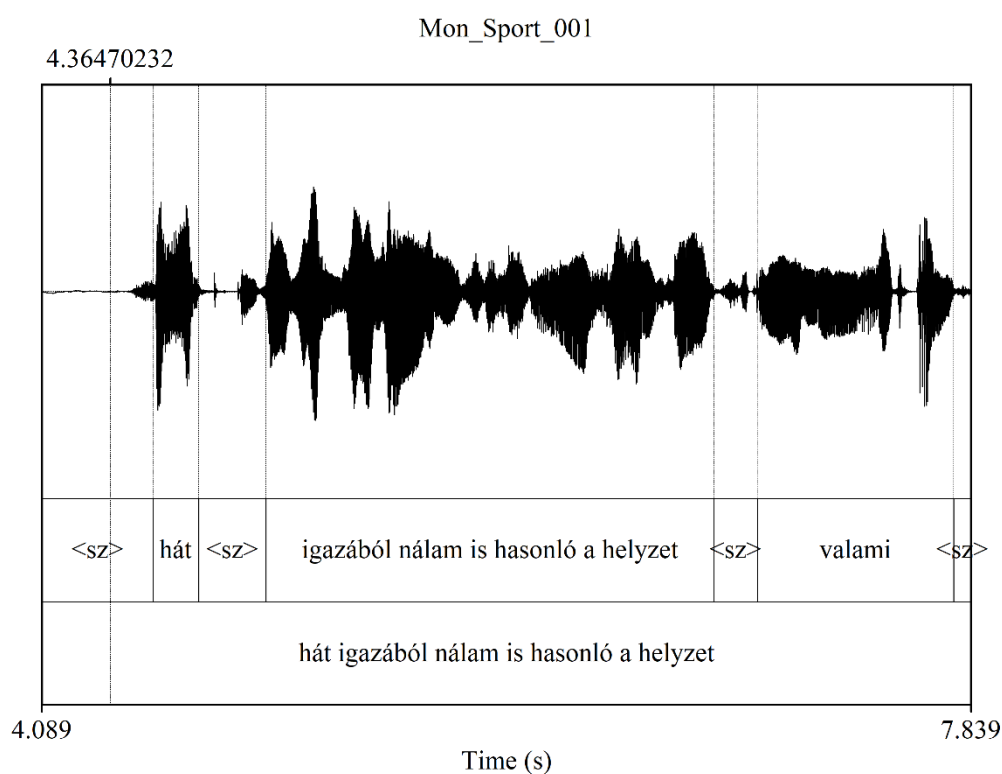
### 10. Figure Concordance of HÁT IGAZÁBÓL ÉN

| Left context                                      | KWIC   | Right context                                      |
|---|--|--|
| , mert félek, hogy nem jól gondoskodik róla. -    | <b>Hát igazából én</b><br>Inj-Utt Adv N_Pro:1Sg:Nom    | pont attól félek, hogy nem tudok róla gondoskodni, |
| t, hogy szegényke még előbb meghalt volna. -      | <b>Hát igazából nekem</b><br>Inj-Utt Adv N_Pro:Dat:1Sg | a hörcsögömmel az volt, hogy én így az elején mo   |
| lis háziállat? - Szerintem teljesen ideális, mert | <b>hát igazából én</b><br>N:Nom Adv N_Pro:1Sg:Nom      | úgy tudom, hogy a kutyák érzik, hogyha bármi bajc  |
| jére.</s><s>Mesélj a kedvenc mozifilmedről. -     | <b>Hát igazából nekem</b><br>Inj-Utt Adv N_Pro:Dat:1Sg | a kedvencem az talán a Star Wars, amit még kiset   |
| edvem se.</s><s>Én ezt a témát elhagytam. -       | <b>Hát igazából én</b><br>Inj-Utt Adv N_Pro:1Sg:Nom    | is egyszer kezdtem bele a Riverdale-be, de aztán : |
| oardoznak, és én ezt szívesen kipróbálnám. -      | <b>Hát igazából nálam</b><br>Inj-Utt Adv N_Pro:Adv:1Sg | is hasonló a helyzet.</s><s>Valami extrémebb sgc   |

The phrase HÁT IGAZÁBÓL ÉN appears in the corpus at the beginning of a statement and the beginning of a response. It helps the speaker make it clear to the listener that they communicate honestly, openly and directly. So, in this structure, the meaning is different. Instead, it suggests that the speaker is about to answer honestly or say something important to him. Among the above examples of language use, one differs from this and may even be of interest for language

teaching and learning. The lexical unit *nálam is hasonló a helyzet*, which is nested with *hát igazából* and can also be used to express agreement. Since it is unclear in the concordance lines whether the two phrases form a unit, I looked at it in Praat to see when the children pause during their speech productions. As can be seen in the figure below, the pause is between *hát* and *igazából nálam is hasonló a helyzet*. In this case, we can conclude that *hát* is used as a time filler. And the unit itself is *igazából nálam is hasonló a helyzet*.

**107. Figure** Pauses during the production of *hát igazából nálam is hasonló a helyzet* (doc#68: Mon\_Sport\_001)



## 8.5 How are the two- or multi-element lexical units nested in longer sequences? What usage patterns do the nested lexical units show?

### 8.5.1 Nesting of *én is*

The lexical unit *én is* is the second most frequent two-gram in the KorSzak Child Language Corpus, with 144 occurrences in children's speech. *Én is* is a short phrase used as a response by the speaker to indicate that they agree with what the other party has said or done. There are also several occurrences of this meaning in the corpus. It also forms part of more extended lexical units in a similar function as an expression of agreement.

9. Table Nesting of *én is* in the KorSzak Child Language Corpus

| Lexical unit                              | Frequency |
|---|-----------|
| <b>én is</b> így vagyok ezzel/vele        | 8         |
| • ezzel <b>én is</b> így vagyok           | (2)       |
| • ezzel <b>én is</b> nagyon így vagyok    | (1)       |
| <b>én is</b> szoktam                      | 6         |
| ezzel <b>én is</b> egyetérték             | 3         |
| • ezzel <b>én is</b> teljesen egyetérték  | (1)       |
| <b>én is</b> nagyon szeretek + Inf.       | 2         |
| ezzel <b>én is</b> egyet tudok érteni     | 1         |
| <b>én is</b> nagyon szeretem              | 1         |
| meg <b>én is</b> ezen a véleményen vagyok | 1         |

In terms of its position in utterances, in the lexical units listed in the table above, it can be seen that *én is* generally retains its initial position or allows a maximum of one word to precede it. Therefore, regarding its nesting, *én* appears most often in the variants of the lexical unit *én is így vagyok ezzel/vele*, whose usage patterns will be discussed in the following subsection.

### 8.5.2 Nesting of *így vagyok* and *úgy vagyok*

Both two-grams comprise a pronoun (*így / úgy*) and the first person conjugated form of the verb *van*, meaning *to be*. The phrase *így vagyok* occurs 10 times and *úgy vagyok* 9 times in the KorSzak Child Language Corpus.

Looking at their collocations, we can see that there are similar words (*én* and *vele*) with which both phrases occur frequently, but at the same time, we can also notice that the collocations of *így vagyok* were more varied in the corpus. However, there is not much difference in the frequency ratios between the two lexical units. The unit *így vagyok* often appears with the personal pronoun *én*, while *úgy vagyok* most often appears with the conjunction *hogy* in the corpus.

**10. Table** Az így vagyok és az úgy vagyok kollokációi a KorSzak Child Language Corpus

| R. | így vagyok   |                         |                                 |         | úgy vagyok  |                         |                                 |         |
|----|--------------|-------------------------|---------------------------------|---------|-------------|-------------------------|---------------------------------|---------|
|    | Word         | Co-occurrences (-5 /+5) | Total occurrences in the corpus | LogDice | Word        | Co-occurrences (-5 /+5) | Total occurrences in the corpus | LogDice |
| 1  | <b>én</b>    | 10                      | 744                             | 8.76    | <b>hogy</b> | 9                       | 1,791                           | 7.36    |
| 2  | <b>is</b>    | 9                       | 1,218                           | 7.91    | <b>én</b>   | 6                       | 744                             | 8.03    |
| 3  | <b>ezzel</b> | 6                       | 36                              | 12.06   | <b>vele</b> | 4                       | 74                              | 10.62   |
| 4  | <b>vele</b>  | 4                       | 74                              | 10.61   |             |                         |                                 |         |
| 5  | <b>és</b>    | 3                       | 1,687                           | 5.86    |             |                         |                                 |         |

While the nesting of *így vagyok* is usually in the variants of *én is így vagyok vele/ezzel*, *úgy vagyok* is most often found in the lexical unit *úgy vagyok (vele)*. The difference between the meanings of the two longer lexical units is that the former implies that the child speaker agrees or identifies with what the other person is saying. In contrast, the speaker uses the latter to express their own point of view.

In connection with the location of pauses during utterances, we can see that while in the case of *én is így vagyok vele/ezzel*, the examination in Praat found that 33% of the time, children in the corpus paused precisely before and after the unit, in the case of *én úgy vagyok (vele)* 55% of the occurrences of the lexical unit, the pause occurs before and after the four items or after a more extended sequence.

**61. Table** Location of pauses during the production of *én úgy vagyok (vele)*

| Doc. | N-gram             | The pause is before/after the lexical unit |    | Comments  |
|------|--------------------|--|----|---|
|      |                    | Yes  | No |   |
| #2   | úgy vagyok         | x  |    | The pause can be measured with sufficient accuracy.   |
| #8   | én úgy vagyok vele | x  |    | Multi-element unit, pattern of the pause:<br><pause> én úgy vagyok vele, hogy rá<br><pause> |



| Doc. | N-gram              | The pause is before/after the lexical unit |    | Comments  |
|------|---------------------|--|----|---|
|      |                     | Yes  | No |   |
| #40  | én úgy vagyok vele  | x  |    | Multi-element unit, pattern of the pause:<br>< pause > mert én úgy vagyok vele,<br>hogy én<pause>                                 |
| #49  | én úgy vagyok       |  | x  | The pattern of the pause is different:<br><pause> én úgy vagyok a játékoknál,<br>hogy most újabban elkezdtem<br>podcastet <pause> |
| #49  | úgy vagyok          | x  |    | Multi-element unit, pattern of the pause:<br><pause> aztán én úgy vagyok vele<br><pause>  |
| #51  | én úgy vagyok vele  | x  |    | The pause can be measured with sufficient accuracy,   |
| #54  | én úgy vagyok vele  | x  |    | The pause can be measured with sufficient accuracy,   |
| #66  | úgy vagyok          |  | x  | The pattern of the pause is different:<br><pause> és <pause> most úgy<br><pause> vagyok, hogy így <pause>                         |
| #68  | én úgy vagyok ezzel | x  |    | The pause can be measured with sufficient accuracy,   |

The units *így vagyok* and *úgy vagyok* are nested in more extended lexical units, each time forming a sentence or a clause, with the vast majority of occurrences in the position of sentence or statement head. Concerning their other usage patterns, it should be noted that, although their low frequency does not allow us to draw any significant conclusions, it seems that the variants of the unit *én is így vagyok ezzel/vele* when they take the role of a clause, are most often followed by the conjunctions *hogyha* and *mint*. In contrast, the conjunction *hogy* is found after every occurrence of the phrase *én úgy vagyok (vele)* in the corpus.

12. Table Occurrences of conjunctions after the units *én is így vagyok ezzel / vele* and *én úgy vagyok (vele)*

| én is így vagyok ezzel / vele |             |                   | én úgy vagyok (vele) |             |                   |
|-------------------------------|-------------|-------------------|----------------------|-------------|-------------------|
| Conjunction                   | Occurrences | Total occurrences | Conjunction          | Occurrences | Total occurrences |
|                               |             |                   |                      |             |                   |

|               |   | %  |             |   | %   |
|---------------|---|----|-------------|---|-----|
| <b>hogyha</b> | 2 | 20 | <b>hogy</b> | 9 | 100 |
| <b>mint</b>   | 2 | 20 |             |   |     |
| <b>de</b>     | 1 | 10 |             |   |     |
| <b>és</b>     | 1 | 10 |             |   |     |
| <b>hogy</b>   | 1 | 10 |             |   |     |

### 8.5.3 Nesting of *mit tudom én*

The lexical unit has 29 occurrences in the KorSzak Child Language Corpus. Regarding collocations, it is most often used with the conjunction *vagy* but appears near *most* on several occasions.

| R. | Word         | Co-occurrences (-5/+5) | Total occurrences in the corpus | LogDice |
|----|--------------|------------------------|---------------------------------|---------|
| 1  | <b>vagy</b>  | 8                      | 516                             | 8.91    |
| 2  | <b>most</b>  | 4                      | 295                             | 8.66    |
| 3  | <b>egy</b>   | 6                      | 764                             | 7.95    |
| 4  | <b>ilyen</b> | 3                      | 541                             | 7.43    |
| 5  | <b>de</b>    | 4                      | 769                             | 7.36    |
| 6  | <b>akkor</b> | 3                      | 591                             | 7.31    |
| 7  | <b>hogy</b>  | 8                      | 1,791                           | 7.17    |
| 8  | <b>az</b>    | 7                      | 1,566                           | 7.17    |
| 9  | <b>nem</b>   | 5                      | 1,412                           | 6.83    |
| 10 | <b>meg</b>   | 3                      | 863                             | 6.78    |

In addition to appearing independently, the term *mit tudom én* is usually part of a four-gram

in the corpus. The examples are the following:

- *hogy mit tudom én* (5 occurrences),
- *vagy mit tudom én* (4 occurrences),
- *meg mit tudom én* (1 occurrence),
- *és mit tudom én* (1 occurrence).

Examining the location of the pauses with Praat, we see that in 80% of the occurrences of *hogy mit tudom én*, in 75% of the occurrences of *vagy mit tudom én*, and both *meg mit tudom én* and *és mit tudom én*, the pauses are either immediately before and after the lexical unit or after a more extended unit.

**62. Table** Location of pauses during the production of *hogy mit tudom én*

| Doc. | N-gram (szó)             | The pause is before/after the lexical unit |    | Comments   |
|------|--------------------------|--|----|--|
|      |                          | Yes  | No |  |
| #2   | <i>hogy mit tudom én</i> |  | x  | The pattern of the pause is different: <pause> <i>hogy</i> <pause> <i>mit tudom én</i> |
| #2   | <i>hogy mit tudom én</i> | x  |    | The pause can be measured with sufficient accuracy,                                    |
| #39  | <i>hogy mit tudom én</i> | x  |    | The pause can be measured with sufficient accuracy,                                    |
| #42  | <i>hogy mit tudom én</i> | x  |    | The pause can be measured with sufficient accuracy,                                    |
| #53  | <i>hogy mit tudom én</i> | x  |    | The pause can be measured with sufficient accuracy,                                    |

**63. Table** Location of pauses during the production of *vagy mit tudom én*

| Doc. | N-gram                   | The pause is before/after the lexical unit |    | Comments   |
|------|--------------------------|--|----|--|
|      |                          | Yes  | No |  |
| #0   | <i>vagy mit tudom én</i> | x  |    | The pause can be measured with sufficient accuracy,                                    |
| #2   | <i>vagy mit tudom én</i> |  | x  | The pattern of the pause is different: <pause> <i>vagy</i> <pause> <i>mit tudom én</i> |
| #9   | <i>vagy mit tudom én</i> | x  |    | Multi-element unit, pattern of the pause:  |

|     |                   |   |  |  |
|-----|-------------------|---|--|--|
|     |                   |   |  | <pause> vagy mit tudom én milyen<br><pause>  |
| #59 | vagy mit tudom én | x |  | Multi-element unit, the pattern of the pause:<br><pause> vagy mit tudom én mire<br><pause> |

**64. Table** Location of pauses during the production of *és mit tudom én* and *meg mit tudom én*

| Doc. | N-gram           | The pause is before/after the lexical unit |    | Comments  |
|------|------------------|--|----|---|
|      |                  | Yes  | No |   |
| #2   | és mit tudom én  | x  |    | The pause can be measured with sufficient accuracy, |
| #58  | meg mit tudom én | x  |    | The pause can be measured with sufficient accuracy, |

## 8.6 What is the role in the text of the two- and multi-element lexical units in the KorSzak Child Language Corpus?

65. Table Summary of the categories of lexical units according to their role in the text (based on Biber, 2006, pp. 139–146)

|   | A                                | B               |                    | C                            |                              |       | D                        |                |                           |                            | E  |
|---|----------------------------------|-----------------|--------------------|------------------------------|------------------------------|-------|--------------------------|----------------|---------------------------|----------------------------|--|
|   | the most frequent lexical units  | Stance bundles  |                    | Discourse organising bundles |                              |       | Referential bundles      |                |                           |                            | All the variations of the selected lexical units |
|   |                                  | epistemic terms | attitude, modality | topic introduction           | clarification of the subject | focus | indication of inaccuracy | identification | indicating time and place | multi-functional reference |  |
| 1 | <b>2-gram</b>                    | 1               | 10                 | 5                            | 87                           | 12    | 0                        | 0              | 8                         | 0                          | 448  |
| 2 | <b>3-gram</b>                    | 9               | 27                 | 6                            | 4                            | 6     | 8                        | 2              | 20                        | 3                          | 340  |
| 3 | <b>4-gram</b>                    | 1               | 19                 | 2                            | 0                            | 1     | 4                        | 1              | 0                         | 1                          | 28   |
| 4 | <b>Total number of varieties</b> | 11              | 56                 | 13                           | 91                           | 19    | 12                       | 3              | 28                        | 4                          | 816  |
| 5 |                                  | 67              |                    | 123                          |                              |       | 47                       |                |                           |                            |  |
| 6 | <b>Frequency</b>                 | 323             | 443                | 203                          | 637                          | 146   | 126                      | 12             | 284                       | 30                         | 12.748   |

## 9. Application of the research findings

The results of the research presented in the doctoral thesis – and the KorSzak Child Language Corpus itself – are theoretically relevant and have many practical applications in the language teaching and curriculum development field. The authors of the teaching materials can use the corpus in their work in the following areas:

1. When selecting useful vocabulary,
2. in establishing the proportions and order of the language forms taught, and in preparing a levelled lexical-grammar syllabus,
3. when examining the relationship between lexis, response variations and text types related to the topics,
4. when selecting real situations relevant to the learners,
5. using corpus data rather than their own intuition when producing texts that reflect natural language use.

The teaching material, *Magyar felfedező* (Kaşikci – Vermeki, 2021), will also use the research results in these areas. *Magyar felfedező* is a project-based, model-based (cf. Szita – Pelcz, 2017) and corpus-informed complex curriculum, which is designed for five language levels (A1-C1) according to the CEFR, primarily for children aged 12–13, and, by extending the ideal time frame, can be used effectively for children aged 11–14. The authors consider the target group of the teaching material to be primarily heritage language speakers. However, they also believe that since the chapters are at several levels, it is also suitable for use in Hungarian as a foreign language groups. Currently, one sample chapter and one more for each of the five language levels have been prepared, making a total of ten chapters.

## 10. Conclusion

The corpus-based research presented in this thesis aimed to observe and analyse the natural language use of native Hungarian-speaking children in the KorSzak Child Language Corpus and to use the results to produce corpus-informed teaching material. The lexical units useful for

language learning were mapped during the research, and their usage patterns were identified.

The research was based on the KorSzak Child Language Corpus, explicitly created for corpus linguistic research that could be used to produce teaching materials for children. Therefore the process of corpus building was an essential part of the thesis. However, it is important to note that corpus building was only a tool for producing teaching materials for children.

It is also important to point out that in teaching Hungarian as a foreign language and other languages, the creation of corpus-informed teaching materials based on the study of children's language use is not yet typical. Pedagogical corpora for children, such as the KorSzak Child Language Corpus, which in this thesis is understood as a set of authentic texts for language teaching, have yet to be available. Although there are corpora specifically designed for children (such as SACODEYL and CorpusMate), teachers have either no access to worksheets or only limited a limited number of worksheets created for them. Complex teaching materials, created by didacticising the texts, have yet to be developed for them.

The significance of the corpus linguistics research presented in this thesis is that, besides dealing with the less researched area of the Hungarian language and Hungarian as a foreign language education, it contributes to the expansion of corpus linguistics research in the field of children's language and pedagogical corpora, as well as in the study of agglutinating languages. It draws attention to the importance of corpus-informed teaching materials for children, including 11–15-year-olds, and shows how authors can use corpora in their work.

Notably, the research has produced a valuable and visible result. This teaching material can help children learn Hungarian and their teachers. It can be used with two methodological approaches that may bring innovation to teaching Hungarian as a foreign language to children. The first of these is the model-based approach (cf. Szita – Pelcz, 2017), which is already widely used in teaching Hungarian as a foreign language but is a novelty in children's language teaching. The second is the project-based approach to teaching (cf. Capraro, 2013; Pieratt, 2019), which, although one would think that it is already in the curricula of widely taught languages, the number of language books for children using this methodology is extremely limited internationally.

To conclude, the dynamic nature of the KorSzak Child Language Corpus means it is constantly being expanded with new additions and developed in several areas. Before the corpus construction started, these plans had already been made so that the corpus was built and

equipped with the metadata that allows it. We have already started to extend the age range of the children in the corpus (16–17 years old). We have already started to make trial recordings. We also want to extend the collection of recordings to other regions of Hungary. Furthermore, we plan to create several sub-corpora in the future, of which recordings with bilingual children have already started (see Schmidt, 2022), which could open up new perspectives in the study of children's language



## Mellékletek

### Melléklet A: A Korszak Gyermeknyelvi Korpuszban lévő kételemű lexikai egységek

| Lexikai egység  | Találatok száma | Változatok gyakorisága (%) |
|-----------------|-----------------|----------------------------|
| <b>NEM TUD</b>  | <b>292</b>      |                            |
| nem tudom       | 227             | 77,74                      |
| nem tudok       | 17              | 5,82                       |
| nem tudjuk      | 7               | 2,40                       |
| nem tudod       | 7               | 2,40                       |
| nem tudja       | 5               | 1,71                       |
| nem tud         | 5               | 1,71                       |
| nem tudtam      | 4               | 1,37                       |
| nem tudunk      | 4               | 1,37                       |
| nem tudnék      | 3               | 1,03                       |
| nem tudsz       | 2               | 0,68                       |
| nem tudta       | 2               | 0,68                       |
| nem tudtunk     | 2               | 0,68                       |
| nem tudnám      | 1               | 0,34                       |
| nem tudják      | 1               | 0,34                       |
| nem tudnának    | 1               | 0,34                       |
| nem tudhatjuk   | 1               | 0,34                       |
| <b>ÉN IS</b>    | <b>188</b>      |                            |
| én is           | 144             | 76,60                      |
| nekem is        | 37              | 19,68                      |
| engem is        | 4               | 2,13                       |
| nálam is        | 2               | 1,06                       |
| velem is        | 1               | 0,53                       |
| <b>ÉS AKKOR</b> | <b>177</b>      |                            |
| és akkor        | 177             | 100                        |
| <b>VAN A</b>    | <b>170</b>      |                            |
| van a           | 53              | 31,18                      |
| volt a          | 45              | 26,47                      |
| vagy a          | 20              | 11,76                      |
| vannak a        | 14              | 8,24                       |
| lenne a         | 9               | 5,29                       |
| vagyok a        | 8               | 4,71                       |
| lehet a         | 3               | 1,76                       |
| volna a         | 2               | 1,18                       |

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| voltunk a      | 2               | 1,18                       |
| voltam a       | 2               | 1,18                       |
| legyenek a     | 1               | 0,59                       |
| lehetek a      | 1               | 0,59                       |
| legyen a       | 1               | 0,59                       |
| voltál a       | 1               | 0,59                       |
| voltatok a     | 1               | 0,59                       |
| voltak a       | 1               | 0,59                       |
| lennék a       | 1               | 0,59                       |
| lenni a        | 1               | 0,59                       |
| legyünk a      | 1               | 0,59                       |
| legyek a       | 1               | 0,59                       |
| <b>EZ A</b>    | <b>169</b>      |                            |
| ez a           | 85              | 50,30                      |
| ezt a          | 42              | 24,85                      |
| ebben a        | 7               | 4,14                       |
| ezeket a       | 7               | 4,14                       |
| ennek a        | 6               | 3,55                       |
| ezek a         | 6               | 3,55                       |
| ezzel a        | 6               | 3,55                       |
| ezekkel a      | 2               | 1,18                       |
| erről a        | 2               | 1,18                       |
| ezeken a       | 1               | 0,59                       |
| ennél a        | 1               | 0,59                       |
| ettől a        | 1               | 0,59                       |
| ezektől a      | 1               | 0,59                       |
| ezekre a       | 1               | 0,59                       |
| ezeknél a      | 1               | 0,59                       |
| <b>AZ A</b>    | <b>151</b>      |                            |
| az a           | 107             | 70,86                      |
| azt a          | 18              | 11,92                      |
| azok a         | 11              | 7,28                       |
| abban a        | 4               | 2,65                       |
| annak a        | 3               | 1,99                       |
| azokat a       | 3               | 1,99                       |
| abból a        | 2               | 1,32                       |
| arra a         | 1               | 0,66                       |
| azokon a       | 1               | 0,66                       |
| abba a         | 1               | 0,66                       |
| <b>VAN EGY</b> | <b>128</b>      |                            |
| van egy        | 68              | 53,13                      |

| Lexikai egység   | Találatok száma | Változatok gyakorisága (%) |
|------------------|-----------------|----------------------------|
| volt egy         | 34              | 26,56                      |
| vagy egy         | 6               | 4,69                       |
| lenne egy        | 5               | 3,91                       |
| lehet egy        | 5               | 3,91                       |
| legyen egy       | 2               | 1,56                       |
| voltatok egy     | 1               | 0,78                       |
| voltam egy       | 1               | 0,78                       |
| legyél egy       | 1               | 0,78                       |
| volna egy        | 1               | 0,78                       |
| voltak egy       | 1               | 0,78                       |
| lehetne egy      | 1               | 0,78                       |
| lennél egy       | 1               | 0,78                       |
| voltunk egy      | 1               | 0,78                       |
| <b>NEM IS</b>    | <b>126</b>      |                            |
| nem is           | 126             | 100                        |
| <b>AZ ILYEN</b>  | <b>108</b>      |                            |
| az ilyen         | 101             | 93,52                      |
| azok ilyen       | 3               | 2,78                       |
| az ilyennel      | 1               | 0,93                       |
| az ilyenek       | 1               | 0,93                       |
| az ilyeneket     | 1               | 0,93                       |
| az ilyet         | 1               | 0,93                       |
| <b>EGY ILYEN</b> | <b>108</b>      |                            |
| egy ilyen        | 106             | 98,15                      |
| egy ilyet        | 1               | 0,93                       |
| első ilyen       | 1               | 0,93                       |
| <b>VAN AZ</b>    | <b>106</b>      |                            |
| volt az          | 34              | 32,08                      |
| van az           | 16              | 15,09                      |
| vannak az        | 9               | 8,49                       |
| vagy az          | 8               | 7,55                       |
| legyen az        | 4               | 3,77                       |
| lenne az         | 4               | 3,77                       |
| lehet az         | 2               | 1,89                       |
| vagyok az        | 2               | 1,89                       |
| vannak azok      | 2               | 1,89                       |
| van annak        | 2               | 1,89                       |
| volt annak       | 2               | 1,89                       |
| lehetett az      | 1               | 0,94                       |
| voltak az        | 1               | 0,94                       |
| lehetnék az      | 1               | 0,94                       |

| Lexikai egység       | Találatok száma | Változatok gyakorisága (%) |
|----------------------|-----------------|----------------------------|
| lehet attól          | 1               | 0,94                       |
| vagy abból           | 1               | 0,94                       |
| vannak annak         | 1               | 0,94                       |
| volt azokban         | 1               | 0,94                       |
| van arra             | 1               | 0,94                       |
| voltunk az           | 1               | 0,94                       |
| <b>MEG A</b>         | <b>96</b>       |                            |
| meg a                | 96              | 100                        |
| <b>ÉS A</b>          | <b>96</b>       |                            |
| és a                 | 96              | 100                        |
| <b>HOGY NEM</b>      | <b>92</b>       |                            |
| hogy nem             | 92              | 100                        |
| <b>OTT VAN</b>       | <b>84</b>       |                            |
| ott van              | 32              | 38,10                      |
| ott volt             | 23              | 27,38                      |
| ott vannak           | 7               | 8,33                       |
| ott voltak           | 5               | 5,95                       |
| ott vagyok           | 4               | 4,76                       |
| ott voltunk          | 4               | 4,76                       |
| ott lehetett         | 2               | 2,38                       |
| ott legyen           | 2               | 2,38                       |
| ott voltak           | 2               | 2,38                       |
| ott voltál           | 1               | 1,19                       |
| ott lehet            | 1               | 1,19                       |
| ott voltam           | 1               | 1,19                       |
| <b>NAGYON SZERET</b> | <b>79</b>       |                            |
| nagyon szeretem      | 29              | 36,71                      |
| nagyon szeretek      | 24              | 30,38                      |
| nagyon szerettem     | 8               | 10,13                      |
| nagyon szeretik      | 4               | 5,06                       |
| nagyon szereti       | 3               | 3,80                       |
| nagyon szeretünk     | 3               | 3,80                       |
| nagyon szeretjük     | 2               | 2,53                       |
| nagyon szeret        | 2               | 2,53                       |
| nagyon szeretsz      | 1               | 1,27                       |
| nagyon szeretnek     | 1               | 1,27                       |
| nagyon szeretnél     | 1               | 1,27                       |
| nagyon szerette      | 1               | 1,27                       |
| <b>AZ IS</b>         | <b>78</b>       |                            |
| az is                | 43              | 55,13                      |
| azt is               | 14              | 17,95                      |

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| attól is       | 7               | 8,97                       |
| abból is       | 2               | 2,56                       |
| abban is       | 2               | 2,56                       |
| azzal is       | 2               | 2,56                       |
| azok is        | 2               | 2,56                       |
| ahhoz is       | 1               | 1,28                       |
| addig is       | 1               | 1,28                       |
| azoktól is     | 1               | 1,28                       |
| azokat is      | 1               | 1,28                       |
| azokra is      | 1               | 1,28                       |
| abba is        | 1               | 1,28                       |
| <b>ÉN AZ</b>   | <b>78</b>       |                            |
| nekem az       | 36              | 46,15                      |
| én azt         | 20              | 25,64                      |
| én az          | 11              | 14,10                      |
| én azokat      | 3               | 3,85                       |
| nekem azt      | 2               | 2,56                       |
| nekem attól    | 1               | 1,28                       |
| én azoktól     | 1               | 1,28                       |
| nekem abból    | 1               | 1,28                       |
| én attól       | 1               | 1,28                       |
| engem az       | 1               | 1,28                       |
| én arra        | 1               | 1,28                       |
| <b>ÉS AZ</b>   | <b>77</b>       |                            |
| és az          | 45              | 58,44                      |
| és azt         | 20              | 25,97                      |
| és attól       | 3               | 3,90                       |
| és ahhoz       | 3               | 3,90                       |
| és abból       | 2               | 2,60                       |
| és annak       | 1               | 1,30                       |
| és azokat      | 1               | 1,30                       |
| és abban       | 1               | 1,30                       |
| és azok        | 1               | 1,30                       |
| <b>IS VAN</b>  | <b>77</b>       |                            |
| is van         | 26              | 33,77                      |
| is volt        | 18              | 23,38                      |
| is vannak      | 12              | 15,58                      |
| is lehet       | 11              | 14,29                      |
| is voltak      | 3               | 3,90                       |
| is lenne       | 2               | 2,60                       |
| is voltunk     | 2               | 2,60                       |

| Lexikai egység   | Találatok száma | Változatok gyakorisága (%) |
|------------------|-----------------|----------------------------|
| is lehetne       | 1               | 1,30                       |
| is vagyok        | 1               | 1,30                       |
| is vagyunk       | 1               | 1,30                       |
| <b>JÓ VAN</b>    | <b>76</b>       |                            |
| jó volt          | 33              | 43,42                      |
| jó lenne         | 14              | 18,42                      |
| jól van          | 11              | 14,47                      |
| jó lehet         | 3               | 3,95                       |
| jó vagy          | 2               | 2,63                       |
| jó van           | 2               | 2,63                       |
| jóban vagyok     | 1               | 1,32                       |
| jobb lenne       | 1               | 1,32                       |
| jóban vagyunk    | 1               | 1,32                       |
| jól vagyok       | 1               | 1,32                       |
| jól vagy         | 1               | 1,32                       |
| jók voltak       | 1               | 1,32                       |
| jóban lennék     | 1               | 1,32                       |
| jobb volt        | 1               | 1,32                       |
| jóba lennék      | 1               | 1,32                       |
| jóban voltunk    | 1               | 1,32                       |
| jól lehet        | 1               | 1,32                       |
| <b>HÁT ÉN</b>    | <b>75</b>       |                            |
| hát én           | 49              | 65,33                      |
| hát nekem        | 25              | 33,33                      |
| hát nálam        | 1               | 1,33                       |
| <b>A KEDVENC</b> | <b>75</b>       |                            |
| a kedvenc        | 44              | 58,67                      |
| a kedvencem      | 15              | 20,00                      |
| a kedvenced      | 6               | 8,00                       |
| a kedvencetek    | 4               | 5,33                       |
| a kedvence       | 3               | 4,00                       |
| a kedvenceim     | 3               | 4,00                       |
| <b>EZ AZ</b>     | <b>72</b>       |                            |
| ez az            | 37              | 51,39                      |
| ezt az           | 9               | 12,50                      |
| ezek az          | 6               | 8,33                       |
| ennek az         | 5               | 6,94                       |
| ebben az         | 4               | 5,56                       |
| ezeket az        | 2               | 2,78                       |

| Lexikai egység  | Találatok száma | Változatok gyakorisága (%) |
|-----------------|-----------------|----------------------------|
| ebbe az         | 1               | 1,39                       |
| ezekkel az      | 1               | 1,39                       |
| ezt abba        | 1               | 1,39                       |
| ezeknél az      | 1               | 1,39                       |
| ezzel az        | 1               | 1,39                       |
| ez azt          | 1               | 1,39                       |
| erről az        | 1               | 1,39                       |
| ezekből az      | 1               | 1,39                       |
| ehhez az        | 1               | 1,39                       |
| <b>AZ EMBER</b> | <b>72</b>       |                            |
| az emberek      | 22              | 30,56                      |
| az ember        | 21              | 29,17                      |
| az embernek     | 6               | 8,33                       |
| az embereket    | 4               | 5,56                       |
| az embereknek   | 3               | 4,17                       |
| az emberekkel   | 2               | 2,78                       |
| az embertől     | 2               | 2,78                       |
| az embert       | 2               | 2,78                       |
| az embereknél   | 1               | 1,39                       |
| az emberről     | 1               | 1,39                       |
| az emberektől   | 1               | 1,39                       |
| az embernél     | 1               | 1,39                       |
| az emberért     | 1               | 1,39                       |
| az emberekre    | 1               | 1,39                       |
| az emberben     | 1               | 1,39                       |
| az emberre      | 1               | 1,39                       |
| az embereké     | 1               | 1,39                       |
| az emberrel     | 1               | 1,39                       |
| <b>HOGY A</b>   | <b>70</b>       |                            |
| hogy a          | 67              | 95,71                      |
| ahogy a         | 3               | 4,29                       |
| <b>IS TUD</b>   | <b>69</b>       |                            |
| is tudom        | 50              | 72,46                      |
| is tud          | 4               | 5,80                       |
| is tudod        | 4               | 5,80                       |
| is tudnak       | 2               | 2,90                       |
| is tudok        | 2               | 2,90                       |
| is tudsz        | 2               | 2,90                       |
| is tudnál       | 1               | 1,45                       |

| Lexikai egység   | Találatok száma | Változatok gyakorisága (%) |
|------------------|-----------------|----------------------------|
| is tudtuk        | 1               | 1,45                       |
| is tudnék        | 1               | 1,45                       |
| is tudhatjuk     | 1               | 1,45                       |
| is tudtok        | 1               | 1,45                       |
| <b>NEM VAN</b>   | <b>68</b>       |                            |
| nem volt         | 23              | 33,82                      |
| nem lehet        | 16              | 23,53                      |
| nem vagyok       | 7               | 10,29                      |
| nem lenne        | 6               | 8,82                       |
| nem lehetne      | 3               | 4,41                       |
| nem voltam       | 3               | 4,41                       |
| nem voltunk      | 2               | 2,94                       |
| nem voltak       | 2               | 2,94                       |
| nem vagyunk      | 2               | 2,94                       |
| nem lehetnek     | 1               | 1,47                       |
| nem vagy         | 1               | 1,47                       |
| nem lehetett     | 1               | 1,47                       |
| nem lennék       | 1               | 1,47                       |
| <b>NAGYON JÓ</b> | <b>65</b>       |                            |
| nagyon jó        | 49              | 75,38                      |
| nagyon jól       | 14              | 21,54                      |
| nagyon jóban     | 1               | 1,54                       |
| nagyon jók       | 1               | 1,54                       |
| <b>VAN OLYAN</b> | <b>65</b>       |                            |
| van olyan        | 28              | 43,08                      |
| vannak olyan     | 12              | 18,46                      |
| volt olyan       | 12              | 18,46                      |
| vannak olyanok   | 3               | 4,62                       |
| van-e olyan      | 3               | 4,62                       |
| vagyunk olyan    | 2               | 3,08                       |
| voltak olyan     | 1               | 1,54                       |
| lenne olyan      | 1               | 1,54                       |
| lehet olyan      | 1               | 1,54                       |
| vagyok olyan     | 1               | 1,54                       |
| volna olyan      | 1               | 1,54                       |
| <b>ÉN NEM</b>    | <b>64</b>       |                            |
| én nem           | 52              | 81,25                      |
| nekem nem        | 9               | 14,06                      |
| nálam nem        | 2               | 3,13                       |



| Lexikai egység  | Találatok száma | Változatok gyakorisága (%) |
|-----------------|-----------------|----------------------------|
| engem nem       | 1               | 1,56                       |
| <b>ÉS EZ</b>    | <b>62</b>       |                            |
| és ez           | 37              | 59,68                      |
| és ezt          | 7               | 11,29                      |
| és ezek         | 4               | 6,45                       |
| és erről        | 3               | 4,84                       |
| és ettől        | 2               | 3,23                       |
| és ebben        | 2               | 3,23                       |
| és ennek        | 1               | 1,61                       |
| és ezzel        | 1               | 1,61                       |
| és ezekre       | 1               | 1,61                       |
| és eddig        | 1               | 1,61                       |
| és ezeket       | 1               | 1,61                       |
| és ebbe         | 1               | 1,61                       |
| és ezekből      | 1               | 1,61                       |
| <b>AZ EGYIK</b> | <b>62</b>       |                            |
| az egyik        | 57              | 91,94                      |
| az egyiket      | 2               | 3,23                       |
| az egyiknek     | 1               | 1,61                       |
| az egyikőjük    | 1               | 1,61                       |
| az egyikbe      | 1               | 1,61                       |
| <b>MEG AZ</b>   | <b>61</b>       |                            |
| meg az          | 50              | 81,97                      |
| meg azt         | 9               | 14,75                      |
| meg attól       | 1               | 1,64                       |
| meg azok        | 1               | 1,64                       |
| <b>HOGY EZ</b>  | <b>58</b>       |                            |
| hogy ez         | 35              | 60,34                      |
| hogy ezt        | 15              | 25,86                      |
| hogy ezzel      | 4               | 6,90                       |
| hogy ehhez      | 1               | 1,72                       |
| ahogy ez        | 1               | 1,72                       |
| hogy ennek      | 1               | 1,72                       |
| hogy ebben      | 1               | 1,72                       |
| <b>ÉN A</b>     | <b>58</b>       |                            |
| nekem a         | 30              | 51,72                      |
| én a            | 26              | 44,83                      |
| engem a         | 2               | 3,45                       |
| <b>DE NEM</b>   | <b>57</b>       |                            |

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| de nem         | 57              | 100                        |
| <b>ÉS ÍGY</b>  | <b>57</b>       |                            |
| és így         | 57              | 100                        |
| <b>DE AZ</b>   | <b>50</b>       |                            |
| de az          | 27              | 54,00                      |
| de attól       | 11              | 22,00                      |
| de azt         | 7               | 14,00                      |
| de abban       | 2               | 4,00                       |
| de azon        | 1               | 2,00                       |
| de azok        | 1               | 2,00                       |
| de azokat      | 1               | 2,00                       |
| <b>VAN Ő</b>   | <b>50</b>       |                            |
| vannak benne   | 8               | 16,00                      |
| vagyok vele    | 8               | 16,00                      |
| van benne      | 4               | 8,00                       |
| voltak benne   | 3               | 6,00                       |
| volt benne     | 3               | 6,00                       |
| vagyok benne   | 2               | 4,00                       |
| van vele       | 2               | 4,00                       |
| van rajta      | 2               | 4,00                       |
| lehetne rajta  | 1               | 2,00                       |
| van rá         | 1               | 2,00                       |
| volt nála      | 1               | 2,00                       |
| van-e neki     | 1               | 2,00                       |
| vagyunk vele   | 1               | 2,00                       |
| van róla       | 1               | 2,00                       |
| legyen rajta   | 1               | 2,00                       |
| lehet róla     | 1               | 2,00                       |
| van belőle     | 1               | 2,00                       |
| lehet neki     | 1               | 2,00                       |
| lehet benne    | 1               | 2,00                       |
| voltam vele    | 1               | 2,00                       |
| <b>AZ EGY</b>  | <b>49</b>       |                            |
| az első        | 30              | 61,22                      |
| az egy         | 16              | 32,65                      |
| az elsők       | 1               | 2,04                       |
| azt egy        | 1               | 2,04                       |
| annak egy      | 1               | 2,04                       |

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| <b>A MÁSIK</b> | <b>49</b>       |                            |
| a másik        | 32              | 65,31                      |
| a másikat      | 7               | 14,29                      |
| a másokra      | 3               | 6,12                       |
| a másiknak     | 3               | 6,12                       |
| a másikba      | 2               | 4,08                       |
| a másikban     | 1               | 2,04                       |
| a másiké       | 1               | 2,04                       |
| <b>AZ AZ</b>   | <b>47</b>       |                            |
| az az          | 27              | 57,45                      |
| azt az         | 5               | 10,64                      |
| azok az        | 4               | 8,51                       |
| annak az       | 4               | 8,51                       |
| az azt         | 1               | 2,13                       |
| abban az       | 1               | 2,13                       |
| azokat az      | 1               | 2,13                       |
| abból azt      | 1               | 2,13                       |
| arra az        | 1               | 2,13                       |
| abból az       | 1               | 2,13                       |
| addig az       | 1               | 2,13                       |
| <b>AZ VAN</b>  | <b>47</b>       |                            |
| az volt        | 24              | 51,06                      |
| az lenne       | 7               | 14,89                      |
| az van         | 6               | 12,77                      |
| annak legyen   | 2               | 4,26                       |
| abban van      | 2               | 4,26                       |
| arra lehet     | 2               | 4,26                       |
| annak van      | 1               | 2,13                       |
| azzal van      | 1               | 2,13                       |
| az lehet       | 1               | 2,13                       |
| azt lehet      | 1               | 2,13                       |
| <b>ÉS NEM</b>  | <b>46</b>       |                            |
| és nem         | 46              | 100                        |
| <b>HOGY AZ</b> | <b>45</b>       |                            |
| hogy az        | 37              | 82,22                      |
| hogy azt       | 3               | 6,67                       |
| hogy attól     | 2               | 4,44                       |
| hogy abból     | 1               | 2,22                       |

| Lexikai egység     | Találatok száma | Változatok gyakorisága (%) |
|--------------------|-----------------|----------------------------|
| hogya azoknak      | 1               | 2,22                       |
| ahogy az           | 1               | 2,22                       |
| <b>ÉS SZERINT</b>  | <b>45</b>       |                            |
| és szerintem       | 29              | 64,44                      |
| és szerinted       | 9               | 20,00                      |
| és szerintetek     | 7               | 15,56                      |
| <b>IS NAGYON</b>   | <b>45</b>       |                            |
| is nagyon          | 45              | 100                        |
| <b>HOGY MI</b>     | <b>44</b>       |                            |
| hogya mit          | 24              | 54,55                      |
| hogya mi           | 17              | 38,64                      |
| hogya nálunk       | 1               | 2,27                       |
| hogya miben        | 1               | 2,27                       |
| hogya nekünk       | 1               | 2,27                       |
| <b>MI A</b>        | <b>40</b>       |                            |
| mi a               | 34              | 85,00                      |
| nálunk a           | 3               | 7,50                       |
| minket a           | 2               | 5,00                       |
| nekünk a           | 1               | 2,50                       |
| <b>ANNYIRA NEM</b> | <b>39</b>       |                            |
| annyira nem        | 39              | 100                        |
| <b>EZ NEM</b>      | <b>38</b>       |                            |
| ez nem             | 20              | 52,63                      |
| ezt nem            | 12              | 31,58                      |
| ezzel nem          | 2               | 5,26                       |
| eddig nem          | 2               | 5,26                       |
| ezeket nem         | 1               | 2,63                       |
| ezek nem           | 1               | 2,63                       |
| <b>ÚGY VAN</b>     | <b>38</b>       |                            |
| úgy van            | 18              | 47,37                      |
| úgy vagyok         | 9               | 23,68                      |
| úgy volt           | 8               | 21,05                      |
| úgy voltam         | 2               | 5,26                       |
| úgy lehetett       | 1               | 2,63                       |
| <b>AKKOR AZ</b>    | <b>37</b>       |                            |
| akkor az           | 24              | 64,86                      |
| akkor azt          | 6               | 16,22                      |

| Lexikai egység   | Találatok száma | Változatok gyakorisága (%) |
|------------------|-----------------|----------------------------|
| akkor abból      | 5               | 13,51                      |
| akkor arra       | 1               | 2,70                       |
| akkor annak      | 1               | 2,70                       |
| <b>ÍGY VAN</b>   | <b>36</b>       |                            |
| így van          | 12              | 33,33                      |
| így vagyok       | 10              | 27,78                      |
| így volt         | 7               | 19,44                      |
| így vannak       | 3               | 8,33                       |
| így vagyunk      | 2               | 5,56                       |
| így lehet        | 2               | 5,56                       |
| <b>AZ NAGYON</b> | <b>35</b>       |                            |
| az nagyon        | 26              | 74,29                      |
| azt nagyon       | 5               | 14,29                      |
| attól nagyon     | 1               | 2,86                       |
| azok nagyon      | 1               | 2,86                       |
| annak nagyon     | 1               | 2,86                       |
| azokon nagyon    | 1               | 2,86                       |
| <b>ÉS NAGYON</b> | <b>34</b>       |                            |
| és nagyon        | 34              | 100                        |
| <b>A JÓ</b>      |                 |                            |
| a legjobb        | 16              | 48,48                      |
| a jó             | 11              | 33,33                      |
| a legjobban      | 5               | 15,15                      |
| a jóból          | 1               | 3,03                       |
| <b>IS AZ</b>     | <b>32</b>       |                            |
| is az            | 23              | 71,88                      |
| is azt           | 4               | 12,50                      |
| is arra          | 3               | 9,38                       |
| is azokat        | 1               | 3,13                       |
| is azzal         | 1               | 3,13                       |
| <b>MERT NEM</b>  | <b>27</b>       |                            |
| mert nem         | 27              | 100                        |
| <b>HÁT A</b>     | <b>25</b>       |                            |
| hát a            | 25              | 100                        |
| <b>VAN MÁR</b>   | <b>24</b>       |                            |
| volt már         | 9               | 37,50                      |
| voltál már       | 4               | 16,67                      |

| Lexikai egység | Találatok száma | Változatok gyakorisága (%) |
|----------------|-----------------|----------------------------|
| voltam már     | 3               | 12,50                      |
| volna már      | 2               | 8,33                       |
| vagy már       | 2               | 8,33                       |
| van már        | 2               | 8,33                       |
| voltál-e már   | 1               | 4,17                       |
| voltunk már    | 1               | 4,17                       |

**Melléklet B: A Korszak Gyermeeknyelvi Korpuszban lévő háromelemű lexikai egységek**

| Lexikai egység       | Találatok száma | Változatok gyakorisága (%) |
|----------------------|-----------------|----------------------------|
| <b>NEM IS TUD</b>    | <b>54</b>       |                            |
| nem is tudom         | 49              | 90,74                      |
| nem is tudod         | 3               | 5,56                       |
| nem is tudtuk        | 1               | 1,85                       |
| nem is tudnék        | 1               | 1,85                       |
| <b>MI TUD ÉN</b>     | <b>29</b>       |                            |
| mit tudom én         | 29              | 100                        |
| <b>OTT VAN A</b>     | <b>24</b>       |                            |
| ott van a            | 12              | 50                         |
| ott volt a           | 6               | 25                         |
| ott vannak a         | 3               | 12,50                      |
| ott voltunk a        | 1               | 4,17                       |
| ott vagyok a         | 1               | 4,17                       |
| ott voltak a         | 1               | 4,17                       |
| <b>JÓ ÉREZ MAGA</b>  | <b>21</b>       |                            |
| jól érzem magam      | 6               | 28,57                      |
| jól éreztük magunkat | 4               | 19,04                      |
| jól érezzük magunkat | 3               | 14,28                      |
| jól érzi magát       | 2               | 9,52                       |

| Lexikai egység          | Találatok száma | Változatok gyakorisága (%) |
|-------------------------|-----------------|----------------------------|
| jól érzem magamat       | 2               | 9,52                       |
| jól érzik magukat       | 1               | 4,76                       |
| jobban érzem magam      | 1               | 4,76                       |
| jól érezték magukat     | 1               | 4,76                       |
| jól éreztem magamat     | 1               | 4,76                       |
| <b>VAN EGY ILYEN</b>    | <b>20</b>       |                            |
| van egy ilyen           | 11              | 55                         |
| volt egy ilyen          | 7               | 35                         |
| voltam egy ilyen        | 1               | 5                          |
| voltunk egy ilyen       | 1               | 5                          |
| <b>DE NEM TUD</b>       | <b>17</b>       |                            |
| de nem tudom            | 12              | 70,59                      |
| de nem tudnám           | 1               | 5,88                       |
| de nem tudunk           | 1               | 5,88                       |
| de nem tudjuk           | 1               | 5,88                       |
| de nem tudja            | 1               | 5,88                       |
| de nem tudták           | 1               | 5,88                       |
| <b>VAGY NEM TUD</b>     | <b>17</b>       |                            |
| vagy nem tudom          | 17              | 100                        |
| <b>ÉN NAGYON SZERET</b> | <b>16</b>       |                            |
| én nagyon szeretem      | 8               | 50                         |
| én nagyon szeretek      | 6               | 37,50                      |
| én nagyon szerettem     | 2               | 27,95                      |
| <b>HOGY NEM TUD</b>     | <b>15</b>       |                            |
| hogy nem tudom          | 8               | 53,33                      |
| hogy nem tudok          | 3               | 20                         |
| hogy nem tudjuk         | 2               | 13,33                      |
| hogy nem tudod          | 1               | 6,67                       |
| hogy nem tudta          | 1               | 6,67                       |
| <b>IS NAGYON SZERET</b> | <b>15</b>       |                            |

| Lexikai egység                  | Találatok száma | Változatok gyakorisága (%) |
|---------------------------------|-----------------|----------------------------|
| *( N_Pro / N_Pro:Acc / N_Acc) + |                 |                            |
| is nagyon szeretem              | 6               | 40                         |
| is nagyon szerettem             | 4               | 26,67                      |
| is nagyon szeretik              | 2               | 13,33                      |
| is nagyon szeretek              | 2               | 13,33                      |
| is nagyon szerette              | 1               | 6,67                       |
| <b>VAN AZ A</b>                 | <b>14</b>       |                            |
| volt az a                       | 6               | 42,85                      |
| volt annak a                    | 2               | 14,28                      |
| van az a                        | 2               | 14,28                      |
| vagy az a                       | 2               | 14,28                      |
| vannak azok a                   | 1               | 7,14                       |
| legyen az a                     | 1               | 7,14                       |
| <b>ÉS NEM TUD</b>               | <b>14</b>       |                            |
| és nem tudom                    | 8               | 57,14                      |
| és nem tudod                    | 2               | 14,29                      |
| és nem tudok                    | 2               | 14,29                      |
| és nem tud                      | 1               | 7,14                       |
| és nem tudja                    | 1               | 7,14                       |
| <b>HÁT NEM TUD</b>              | <b>14</b>       |                            |
| hát nem tudom                   | 12              | 85,71                      |
| hát nem tudok                   | 1               | 7,14                       |
| hát nem tudunk                  | 1               | 7,14                       |
| <b>MI A KEDVENC</b>             | <b>13</b>       |                            |
| mi a kedvenc                    | 12              | 92,31                      |
| mi a kedvenced                  | 1               | 7,69                       |
| <b>NAGYON JÓ VAN</b>            | <b>13</b>       |                            |
| nagyon jó volt                  | 8               | 61,54                      |
| nagyon jól van                  | 2               | 15,38                      |
| nagyon jó lenne                 | 1               | 7,69                       |



| Lexikai egység        | Találatok száma | Változatok gyakorisága (%) |
|-----------------------|-----------------|----------------------------|
| nagyon jóban vagyunk  | 1               | 7,69                       |
| nagyon jók voltak     | 1               | 7,69                       |
| <b>ÉN NEM TUD</b>     | <b>13</b>       |                            |
| én nem tudom          | 10              | 76,92                      |
| én nem tudok          | 2               | 15,38                      |
| nekem nem tudom       | 1               | 7,69                       |
| <b>VAN EZ A</b>       | <b>10</b>       |                            |
| van ez a              | 4               | 40                         |
| volt ez a             | 3               | 30                         |
| vagyok ennek a        | 1               | 1                          |
| vagyok ez a           | 1               | 1                          |
| volt ennek a          | 1               | 1                          |
| <b>VAGY HOGY MOND</b> | <b>13</b>       |                            |
| vagy hogy mondjam     | 13              | 100                        |
| <b>IS ÍGY VAN</b>     | <b>13</b>       |                            |
| is így vagyok         | 7               | 53,85                      |
| is így van            | 3               | 23,08                      |
| is így volt           | 2               | 15,38                      |
| is így vannak         | 1               | 7,69                       |
| <b>VAN AZ ILYEN</b>   | <b>10</b>       |                            |
| volt az ilyen         | 6               | 60                         |
| vannak az ilyen       | 4               | 40                         |
| <b>ÉS AKKOR ÍGY</b>   | <b>12</b>       |                            |
| és akkor így          | 12              | 100                        |
| <b>ÉN EZ A</b>        | <b>12</b>       |                            |
| nekem ez a            | 4               | 33,33                      |
| én ezt a              | 3               | 25                         |
| én ez a               | 1               | 8,33                       |
| nekem ezek a          | 1               | 8,33                       |
| nekem ezt a           | 1               | 8,33                       |

| Lexikai egység            | Találatok száma | Változatok gyakorisága (%) |
|---------------------------|-----------------|----------------------------|
| én ezzel a                | 1               | 8,33                       |
| én ezekkel a              | 1               | 8,33                       |
| <b>EZ ÉN IS</b>           | <b>12</b>       |                            |
| *ezzel én is + egyetértek | 7               | 58,33                      |
| *ezt én is + Pst: Sg1     | 4               | 33,33                      |
| *ez engem is + megfog     | 1               | 8,33                       |
| <b>NEM TUD MI</b>         | <b>12</b>       |                            |
| nem tudom miért           | 5               | 41,67                      |
| *nem tudok mit + Inf      | 2               | 16,67                      |
| nem tudom mit             | 2               | 16,67                      |
| *nem tudunk mit + Inf     | 1               | 8,33                       |
| nem tudom mi              | 1               | 8,33                       |
| *nem tudsz mit + Inf      | 1               | 8,33                       |
| <b>VAN EGY OLYAN</b>      | <b>11</b>       |                            |
| volt egy olyan            | 5               | 45,45                      |
| van egy olyan             | 5               | 45,45                      |
| lenne egy olyan           | 1               | 9,09                       |
| <b>ÉS AKKOR AZ</b>        | <b>11</b>       |                            |
| és akkor az               | 7               | 63,64                      |
| és akkor azt              | 1               | 9,09                       |
| és akkor annak            | 1               | 9,09                       |
| és akkor abból            | 1               | 9,09                       |
| és akkor arra             | 1               | 9,09                       |
| <b>ÉN AZ A</b>            | <b>11</b>       |                            |
| nekem az a                | 7               | 63,64                      |
| nekem abból a             | 1               | 9,09                       |
| engem az a                | 1               | 9,09                       |
| én azt a                  | 1               | 9,09                       |
| én azokat a               | 1               | 9,09                       |
| <b>ÉS AKKOR OTT</b>       | <b>11</b>       |                            |

| Lexikai egység          | Találatok száma | Változatok gyakorisága (%) |
|-------------------------|-----------------|----------------------------|
| és akkor ott            | 11              | 100                        |
| <b>VAN AZ EGY</b>       | <b>11</b>       |                            |
| volt az első            | 7               | 63,64                      |
| lenne az első           | 2               | 18,18                      |
| voltak az elsők         | 1               | 9,09                       |
| van annak egy           | 1               | 9,09                       |
| <b>ÉS NAGYON SZERET</b> | <b>10</b>       |                            |
| és nagyon szeretek      | 3               | 30                         |
| és nagyon szeretem      | 3               | 30                         |
| és nagyon szerettem     | 2               | 20                         |
| és nagyon szeretjük     | 1               | 10                         |
| és nagyon szeretnek     | 1               | 10                         |
| <b>AZ VAN A</b>         | <b>10</b>       |                            |
| az volt a               | 5               | 50                         |
| az lenne a              | 4               | 40                         |
| *azzal van a + baj      | 1               | 10                         |
| <b>HÁT ÉN IS</b>        | <b>10</b>       |                            |
| hát én is               | 6               | 60                         |
| hát nekem is            | 4               | 40                         |
| <b>EGY IDŐ UTÁN</b>     | <b>10</b>       |                            |
| egy idő után            | 10              | 100                        |
| <b>AZ A BAJ</b>         | <b>10</b>       |                            |
| az a baj                | 9               | 90                         |
| az a bajom              | 1               | 10                         |
| <b>AZ NAGYON JÓ</b>     | <b>9</b>        |                            |
| az nagyon jó            | 9               | 100                        |
| <b>VAGY NEM IS</b>      | <b>9</b>        |                            |
| vagy nem is             | 9               | 100                        |
| *vagy nem is + tudom    | 7               | 77,77                      |
| <b>MERT NEM TUD</b>     | <b>9</b>        |                            |

| Lexikai egység        | Találatok száma | Változatok gyakorisága (%) |
|-----------------------|-----------------|----------------------------|
| mert nem tudom        | 8               | 88,89                      |
| mert nem tudja        | 1               | 11,11                      |
| <b>HOGY EZ A</b>      | <b>9</b>        |                            |
| hogyan ez a           | 4               | 44,44                      |
| hogyan ezt a          | 3               | 33,33                      |
| ahogy ez a            | 1               | 11,11                      |
| hogyan ennek a        | 1               | 11,11                      |
| <b>IS VAN EGY</b>     | <b>9</b>        |                            |
| is van egy            | 5               | 55,56                      |
| is volt egy           | 3               | 33,33                      |
| is lehet egy          | 1               | 11,11                      |
| <b>NEM NAGYON VAN</b> | <b>9</b>        |                            |
| nem nagyon van        | 5               | 55,56                      |
| nem nagyon lehet      | 2               | 22,22                      |
| nem nagyon voltam     | 2               | 22,22                      |
| <b>ÉS ÉN IS</b>       | <b>9</b>        |                            |
| és én is              | 8               | 88,89                      |
| és nekem is           | 1               | 11,11                      |
| <b>ÉS ÉN EZ</b>       | <b>9</b>        |                            |
| és nekem ez           | 3               | 33,33                      |
| és én ezt             | 3               | 33,33                      |
| és engem ez           | 1               | 11,11                      |
| és én ezekre          | 1               | 11,11                      |
| és nekem ezek         | 1               | 11,11                      |
| <b>ÍGY NEM TUD</b>    | <b>9</b>        |                            |
| így nem tudom         | 5               | 55,56                      |
| így nem tudok         | 2               | 22,22                      |
| így nem tudtam        | 1               | 11,11                      |
| így nem tudnának      | 1               | 11,11                      |
| <b>EZ AZ EGÉSZ</b>    | <b>9</b>        |                            |

| Lexikai egység         | Találatok száma | Változatok gyakorisága (%) |
|------------------------|-----------------|----------------------------|
| ezt az egészet         | 4               | 44,44                      |
| ez az egész            | 2               | 22,22                      |
| ebbe az egészbe        | 1               | 11,11                      |
| ezt az egész           | 1               | 11,11                      |
| ebben az egészben      | 1               | 11,11                      |
| <b>OTT VAN EGY</b>     | <b>8</b>        |                            |
| ott van egy            | 4               | 50                         |
| ott volt egy           | 3               | 37,50                      |
| ott voltak egy         | 1               | 12,50                      |
| <b>ÉN IS SZERET</b>    | <b>8</b>        |                            |
| én is szeretnék        | 3               | 37,50                      |
| én is szeretek         | 3               | 37,50                      |
| én is szeretem         | 2               | 25                         |
| <b>HOGY IS MOND</b>    | <b>8</b>        |                            |
| hogy is mondjam        | 8               | 100                        |
| <b>ÉS EGYSZER CSAK</b> | <b>8</b>        |                            |
| és egyszer csak        | 8               | 100                        |
| <b>ÉN IS ÍGY</b>       | <b>8</b>        |                            |
| én is így              | 8               | 100                        |
| <b>IS OTT VAN</b>      | <b>8</b>        |                            |
| is ott volt            | 4               | 50                         |
| is ott van             | 2               | 25                         |
| is ott voltam          | 1               | 12,50                      |
| is ott legyen          | 1               | 12,50                      |
| <b>Ő VAN A</b>         | <b>8</b>        |                            |
| benne van a            | 4               | 50                         |
| ő volt a               | 3               | 37,50                      |
| benne lennék a         | 1               | 12,50                      |
| <b>NEM AZ MOND</b>     | <b>8</b>        |                            |
| nem azt mondom         | 8               | 100                        |

| Lexikai egység         | Találatok száma | Változatok gyakorisága (%) |
|------------------------|-----------------|----------------------------|
| <b>NEM VAN JÓ</b>      | <b>8</b>        |                            |
| nem lenne jó           | 3               | 37,50                      |
| nem volt jó            | 3               | 37,50                      |
| nem vagyok jóban       | 1               | 12,50                      |
| nem vagy jó            | 1               | 12,50                      |
| <b>TI MI A</b>         | <b>8</b>        |                            |
| nektek mi a            | 8               | 100                        |
| <b>KI A KEDVENC</b>    | <b>8</b>        |                            |
| ki a kedvenc           | 5               | 62,50                      |
| ki a kedvenced         | 2               | 25                         |
| kik a kedvenc          | 1               | 12,50                      |
| <b>ÉS Ő IS</b>         | <b>8</b>        |                            |
| és ő is                | 7               | 87,50                      |
| és rá is               | 1               | 12,50                      |
| <b>NAGYON SZERET A</b> | <b>8</b>        |                            |
| nagyon szeretem a      | 5               | 62,50                      |
| nagyon szereti a       | 1               | 12,50                      |
| nagyon szeretik a      | 1               | 12,50                      |
| nagyon szerette a      | 1               | 12,50                      |
| <b>ÉN IS NAGYON</b>    | <b>8</b>        |                            |
| én is nagyon           | 7               | 87,50                      |
| nekem is nagyon        | 1               | 12,50                      |
| <b>HÁT ÉN EZ</b>       | <b>8</b>        |                            |
| hát nekem ez           | 3               | 37,50                      |
| hát én ezt             | 2               | 25                         |
| hát én ezzel           | 1               | 12,50                      |
| hát nekem ezek         | 1               | 12,50                      |
| hát én ezeken          | 1               | 12,50                      |
| <b>HÁT ÉN NEM</b>      | <b>8</b>        |                            |
| hát én nem             | 5               | 62,50                      |

| Lexikai egység                   | Találatok száma | Változatok gyakorisága (%) |
|----------------------------------|-----------------|----------------------------|
| hát nekem nem                    | 3               | 37,50                      |
| <b>DE VAN OLYAN</b>              | <b>7</b>        |                            |
| de van olyan                     | 6               | 85,71                      |
| de vannak olyan                  | 1               | 14,29                      |
| <b>HOGY EZ ÍGY</b>               | <b>7</b>        |                            |
| hogy ez a                        | 5               | 71,43                      |
| hogy ezt a                       | 1               | 14,29                      |
| hogy ennek a                     | 1               | 14,29                      |
| <b>DE SZERINT EZ</b>             | <b>7</b>        |                            |
| de szerintem ez                  | 6               | 83,71                      |
| de szerintem ettől               | 1               | 14,29                      |
| <b>Ő IS NAGYON</b>               | <b>7</b>        |                            |
| ő is nagyon                      | 7               | 100                        |
| <b>AKKOR OTT VAN</b>             | <b>7</b>        |                            |
| akkor ott volt                   | 3               | 42,86                      |
| akkor ott van                    | 3               | 42,86                      |
| akkor ott vagyok                 | 1               | 14,29                      |
| <b>HOGY MI TUD</b>               | <b>7</b>        |                            |
| *hogy mit tudom + én             | 6               | 85,71                      |
| *hogy mit tudok + belőle kihozni | 1               | 14,29                      |
| <b>ÉN AZ EGYIK</b>               | <b>7</b>        |                            |
| nekem az egyik                   | 6               | 85,71                      |
| nekem az egyiknek                | 1               | 14,29                      |
| <b>IS JÓ VAN</b>                 | <b>7</b>        |                            |
| is jó volt                       | 3               | 42,86                      |
| is jó lenne                      | 1               | 14,29                      |
| is jó vagyok                     | 1               | 14,29                      |
| is jó lehet                      | 1               | 14,29                      |
| is jó lennék                     | 1               | 14,29                      |
| <b>HOGY VAN EGY</b>              | <b>7</b>        |                            |

| Lexikai egység                      | Találatok száma | Változatok gyakorisága (%) |
|-------------------------------------|-----------------|----------------------------|
| hogy volt egy                       | 4               | 57,14                      |
| hogy van egy                        | 2               | 42,86                      |
| <b>ÉN IS AZ</b>                     | <b>7</b>        |                            |
| nekem is az                         | 3               | 41,86                      |
| én is azt                           | 2               | 28,57                      |
| én is azokat                        | 1               | 14,29                      |
| én is az                            | 1               | 14,29                      |
| <b>ÉN IS VAN</b>                    | <b>7</b>        |                            |
| nekem is van                        | 4               | 57,14                      |
| nekem is lenne                      | 1               | 14,29                      |
| nekem is voltak                     | 1               | 14,29                      |
| nekem is volt                       | 1               | 14,29                      |
| <b>AZ JÓ VAN</b>                    | <b>7</b>        |                            |
| az jó volt                          | 4               | 57,14                      |
| az jó lenne                         | 2               | 28,57                      |
| az jó lehet                         | 1               | 14,29                      |
| <b>HOGY ÉN IS</b>                   | <b>7</b>        |                            |
| hogy én is                          | 7               | 100                        |
| <b>AZ A LÉNYEG</b>                  | <b>7</b>        |                            |
| az a lényeg                         | 6               | 85,71                      |
| az a lényege                        | 1               | 14,29                      |
| <b>AZ VAN AZ</b>                    | <b>7</b>        |                            |
| az volt az                          | 5               | 71,43                      |
| az lenne az                         | 2               | 28,57                      |
| *az volt / lenne az első + Poss:1Sg | 6               | 85,71                      |
| <b>A JÓ BARÁT</b>                   | <b>7</b>        |                            |
| a jó barátok                        | 3               | 42,86                      |
| a jó barátokat                      | 2               | 28,57                      |
| a legjobb barátom                   | 1               | 14,29                      |



| Lexikai egység                               | Találatok száma | Változatok gyakorisága (%) |
|--|-----------------|----------------------------|
| a legjobb barát                              | 1               | 14,29                      |
| <b>HOGY NEM VAN</b>                          | <b>6</b>        |                            |
| hogy nem volt                                | 2               | 33,33                      |
| hogy nem voltak                              | 1               | 16,67                      |
| hogy nem vagyunk                             | 1               | 16,67                      |
| hogy nem lehet                               | 1               | 16,67                      |
| hogy nem voltam                              | 1               | 16,67                      |
| <b>ÉS NAGYON JÓ</b>                          | <b>6</b>        |                            |
| és nagyon jó                                 | 4               | 66,67                      |
| és nagyon jól                                | 2               | 33,33                      |
| <b>ÉN ÚGY VAN</b>                            | <b>6</b>        |                            |
| *én úgy vagyok + ezzel / vele / a játékoknál | 6               | 100                        |
| <b>DE NEM IS</b>                             | <b>6</b>        |                            |
| de nem is                                    | 6               | 100                        |
| <b>ÉS NEM IS</b>                             | <b>6</b>        |                            |
| és nem is                                    | 6               | 100                        |
| <b>IS NAGYON JÓ</b>                          | <b>6</b>        |                            |
| is nagyon jó                                 | 5               | 83,33                      |
| is nagyon jól                                | 1               | 16,67                      |
| <b>ÉN EZ NEM</b>                             | <b>6</b>        |                            |
| én ezt nem                                   | 4               | 66,67                      |
| én ezeket nem                                | 1               | 16,67                      |
| én eddig nem                                 | 1               | 16,67                      |
| <b>ÉS OTT VAN</b>                            | <b>6</b>        |                            |
| és ott van                                   | 2               | 33,33                      |
| és ott volt                                  | 2               | 33,33                      |
| és ott vannak                                | 1               | 16,67                      |
| és ott lehetett                              | 1               | 16,67                      |
| <b>NAGYON JÓ ÉREZ</b>                        | <b>6</b>        |                            |
| *nagyon jól éreztük + magunkat               | 4               | 66,67                      |

| Lexikai egység                        | Találatok száma | Változatok gyakorisága (%) |
|---------------------------------------|-----------------|----------------------------|
| *nagyon jól érezzük + magunkat        | 1               | 16,67                      |
| *nagyon jól éreztem + magam / magamat | 1               | 16,67                      |
| <b>MI VAN A</b>                       | <b>6</b>        |                            |
| mi volt a                             | 2               | 33,33                      |
| mi legyünk a                          | 1               | 16,67                      |
| mi lenne a                            | 1               | 16,67                      |
| mi vagyunk a                          | 1               | 16,67                      |
| mi lehet a                            | 1               | 16,67                      |
| <b>VAN AZ AZ</b>                      | <b>6</b>        |                            |
| volt az az                            | 3               | 50                         |
| lennék annak az                       | 1               | 16,67                      |
| lehetnék az az                        | 1               | 16,67                      |
| vannak azok az                        | 1               | 16,67                      |
| <b>MEG NEM IS</b>                     | <b>6</b>        |                            |
| még nem is                            | 6               | 100                        |
| <b>ÉS AZ IS</b>                       | <b>6</b>        |                            |
| és az is                              | 4               | 66,67                      |
| és azt is                             | 2               | 33,33                      |
| <b>DE Ő IS</b>                        | <b>6</b>        |                            |
| de ő is                               | 4               | 66,67                      |
| de neki is                            | 2               | 33,33                      |
| <b>NEM VAN OLYAN</b>                  | <b>6</b>        |                            |
| nem volt olyan                        | 3               | 50                         |
| nem vagyunk olyan                     | 2               | 33,33                      |
| nem vagyok olyan                      | 1               | 16,67                      |
| <b>EZ EGY ILYEN</b>                   | <b>6</b>        |                            |
| ez egy ilyen                          | 5               | 83,33                      |
| ezek egy ilyen                        | 1               | 16,67                      |
| <b>TŐK JÓ VAN</b>                     | <b>6</b>        |                            |

| Lexikai egység             | Találatok száma | Változatok gyakorisága (%) |
|----------------------------|-----------------|----------------------------|
| tök jó volt                | 4               | 66,67                      |
| tök jó lenne               | 2               | 33,33                      |
| <b>ÉN IS SZOKIK</b>        | <b>6</b>        |                            |
| én is szoktam              | 6               | 100                        |
| <b>AZ IS NAGYON</b>        | <b>6</b>        |                            |
| azt is nagyon              | 4               | 66,67                      |
| az is nagyon               | 2               | 33,33                      |
| <b>ÉS AKKOR ÚGY</b>        | <b>6</b>        |                            |
| és akkor úgy               | 6               | 100                        |
| <b>HÁT IGAZÁBÓL ÉN</b>     | <b>6</b>        |                            |
| hát igazából én            | 3               | 50                         |
| hát igazából nekem         | 2               | 33,33                      |
| hát igazából nálam         | 1               | 16,67                      |
| <b>ÉS HOGY EZ</b>          | <b>5</b>        |                            |
| és hogy ez                 | 2               | 40                         |
| és hogy ezt                | 2               | 40                         |
| és hogy ezzel              | 1               | 20                         |
| <b>IS AZ ILYEN</b>         | <b>5</b>        |                            |
| is az ilyen                | 5               | 100                        |
| <b>ITT VAN AZ</b>          | <b>5</b>        |                            |
| itt van az                 | 4               | 80                         |
| itt vagyok az              | 1               | 20                         |
| <b>ÉS SZERINT NAGYON</b>   | <b>5</b>        |                            |
| *és szerintem nagyon + Adj | 5               | 100                        |
| <b>ÉS Ő VAN</b>            | <b>5</b>        |                            |
| és benne van               | 2               | 40                         |
| *és hozzá lehetett + Inf   | 1               | 20                         |
| és ő volt                  | 1               | 20                         |
| és benne volt              | 1               | 20                         |

| Lexikai egység                                | Találatok száma | Változatok gyakorisága (%) |
|---|-----------------|----------------------------|
| <b>MI GONDOL A</b>                            | <b>5</b>        |                            |
| mit gondolsz a                                | 3               | 60                         |
| mit gondolhat a                               | 1               | 20                         |
| mit gondolok a                                | 1               | 20                         |
| <b>TÖK JÓ DOLOG</b>                           | <b>5</b>        |                            |
| tök jó dolog                                  | 4               | 80                         |
| tök jó dolgok                                 | 1               | 20                         |
| <b>EZ ÍGY NEM</b>                             | <b>5</b>        |                            |
| ez így nem                                    | 3               | 60                         |
| ezt így nem                                   | 2               | 40                         |
| <b>OTT VAN AZ</b>                             | <b>5</b>        |                            |
| ott volt az                                   | 2               | 40                         |
| ott van az                                    | 2               | 40                         |
| ott lehetett az                               | 1               | 20                         |
| <b>ÉS SZERINT EZ</b>                          | <b>5</b>        |                            |
| és szerintem ez                               | 5               | 100                        |
| <b>ÉN AZ VAN</b>                              | <b>5</b>        |                            |
| nekem az lenne                                | 3               | 60                         |
| nekem az volt                                 | 2               | 40                         |
| *nekem az volt / lenne + az (első) + Poss:1Sg | 4               | 80                         |
| <b>ÚGY VAN Ő</b>                              | <b>5</b>        |                            |
| úgy vagyok vele                               | 4               | 80                         |
| úgy voltam vele                               | 1               | 20                         |
| <b>ITT IS VAN</b>                             | <b>5</b>        |                            |
| itt is van                                    | 4               | 80                         |
| itt is vannak                                 | 1               | 20                         |
| <b>DE EZ AZ</b>                               | <b>5</b>        |                            |
| de ez az                                      | 2               | 40                         |
| de ennek az                                   | 1               | 20                         |

| Lexikai egység                    | Találatok száma | Változatok gyakorisága (%) |
|-----------------------------------|-----------------|----------------------------|
| de ezekkel az                     | 1               | 20                         |
| de ezek az                        | 1               | 20                         |
| <b>AZ ILYEN NAGY</b>              | <b>5</b>        |                            |
| az ilyen nagy                     | 4               | 80                         |
| az ilyen nagyobb                  | 1               | 20                         |
| <b>ÍGY VAN Ő</b>                  | <b>5</b>        |                            |
| *N_Pro/Adv + is + így vagyok vele | 5               | 100                        |
| <b>EZ IS EGY</b>                  | <b>5</b>        |                            |
| ez is egy                         | 4               | 80                         |
| ezekekről is egy                  | 1               | 20                         |
| <b>AMIKOR KICSI VAN</b>           | <b>5</b>        |                            |
| amikor kicsi voltam               | 2               | 40                         |
| amikor kisebb voltam              | 2               | 40                         |
| mikor kisebb voltam               | 1               | 20                         |
| <b>VAGY MI TUD</b>                | <b>5</b>        |                            |
| vagy mit tudom                    | 5               | 100                        |
| <b>HOGY MI GONDOL</b>             | <b>5</b>        |                            |
| hogy mit gondolsz                 | 2               | 40                         |
| hogy mit gondolnak                | 1               | 20                         |
| hogy mit gondolok                 | 1               | 20                         |
| hogy mit gondol                   | 1               | 20                         |
| <b>MÉG AKKOR IS</b>               | <b>5</b>        |                            |
| még akkor is                      | 5               | 100                        |
| <b>Ő VAN EGY</b>                  | <b>5</b>        |                            |
| benne van egy                     | 3               | 60                         |
| neki volt egy                     | 1               | 20                         |
| benne legyél egy                  | 1               | 20                         |

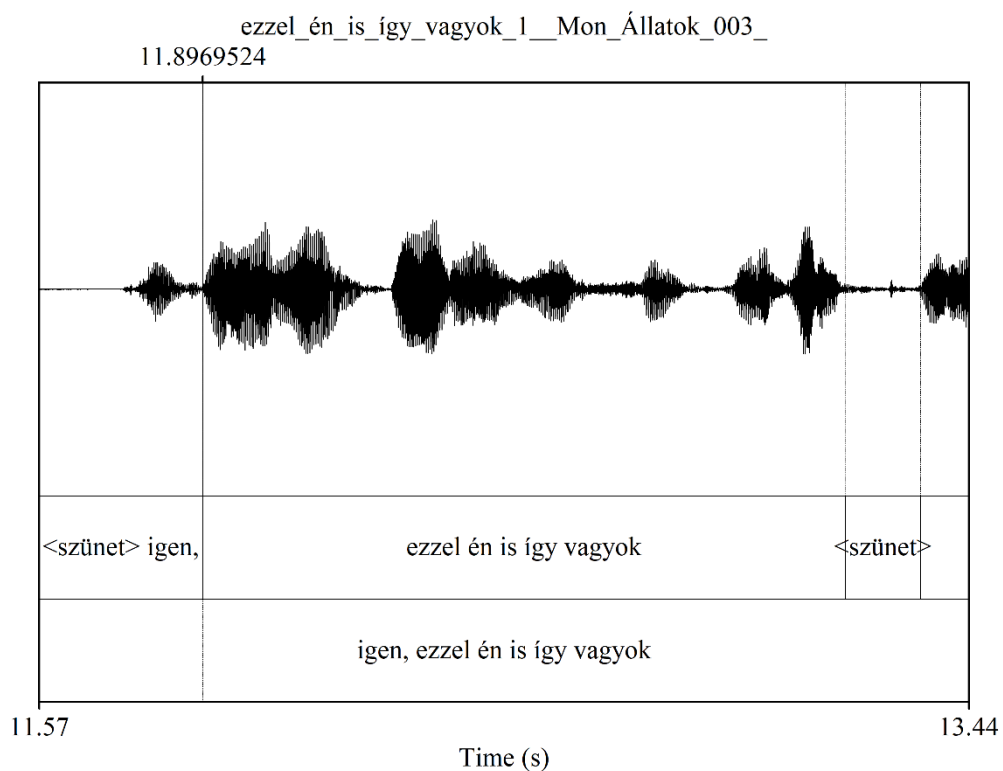
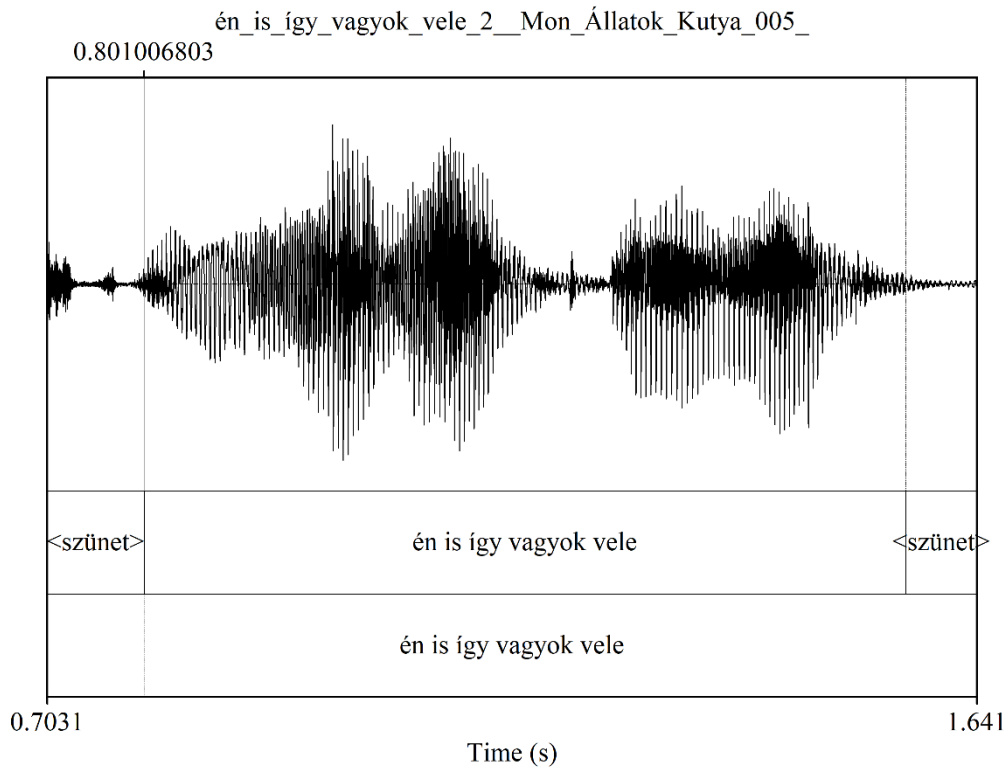
**Melléklet C: A Korszak Gyermeeknyelvi Korpuszban lévő négyelemű lexikai egységek**

| Lexikai egység              | Találatok száma | Változatok gyakorisága (%) |
|-----------------------------|-----------------|----------------------------|
| <b>VAGY NEM IS TUD</b>      | <b>7</b>        |                            |
| vagy nem is tudom           | 7               | 100                        |
| <b>ÉN IS ÍGY VAN</b>        | <b>7</b>        |                            |
| én is így vagyok            | 7               | 100                        |
| <b>AZ VAN AZ EGY</b>        | <b>6</b>        |                            |
| az volt az első             | 4               | 66                         |
| az lenne az első            | 2               | 33                         |
| <b>NAGYON JÓ ÉREZ MAGA</b>  | <b>6</b>        |                            |
| nagyon jól éreztük magunkat | 4               | 66                         |
| nagyon jól éreztem magamat  | 1               | 16                         |
| nagyon jól érezzük magunkat | 1               | 16                         |
| <b>HOGY MI TUD ÉN</b>       | <b>5</b>        |                            |
| hogy mit tudom én           | 5               | 100                        |
| <b>VAGY MI TUD ÉN</b>       | <b>4</b>        |                            |
| vagy mit tudom én           | 4               | 100                        |
| <b>ÉN IS NAGYON SZERET</b>  | <b>4</b>        |                            |
| én is nagyon szeretek       | 2               | 50                         |
| én is nagyon szerettem      | 1               | 25                         |
| én is nagyon szeretem       | 1               | 25                         |
| <b>ÉN ÚGY VAN Ő</b>         | <b>4</b>        |                            |
| én úgy vagyok vele          | 4               | 100                        |
| <b>AZ IS NAGYON SZERET</b>  | <b>4</b>        |                            |
| azt is nagyon szeretem      | 3               | 75                         |
| azt is nagyon szeretek      | 1               | 25                         |
| <b>AZ NAGYON JÓ VAN</b>     | <b>4</b>        |                            |
| az nagyon jó volt           | 4               | 100                        |
| <b>ÉN IS EGYET TUD</b>      | <b>4</b>        |                            |
| én is egyet tudok           | 4               | 100                        |

| Lexikai egység                       | Találatok száma | Változatok gyakorisága (%) |
|--------------------------------------|-----------------|----------------------------|
| *én is egyet tudok érteni            | 2               | 50                         |
| *én is egyet tudok ezzel érteni      | 1               | 25                         |
| *én is egyet tudok veletek érteni    | 1               | 25                         |
| <b>IS VAN EGY ILYEN</b>              | <b>4</b>        |                            |
| is van egy ilyen                     | 4               | 100                        |
| <b>ÉN AZ A KEDVENC</b>               | <b>4</b>        |                            |
| *nekem az a kedvenc + N (Poss: 1Sg)  | 3               | 75                         |
| nekem az a kedvencem                 | 1               | 25                         |
| <b>AZ VAN A KÉRDÉS</b>               | <b>3</b>        |                            |
| az lenne a kérdésem                  | 3               | 100                        |
| <b>DE HÁT NEM TUD</b>                | <b>3</b>        |                            |
| de hát nem tudom                     | 2               | 66                         |
| *de hát nem tudunk + mit csinálni    | 1               | 33                         |
| <b>TI MI A KEDVENC</b>               | <b>3</b>        |                            |
| *nektek mi a kedvenc + N (Poss: 2Pl) | 3               | 100                        |
| <b>DE Ő IS NAGYON</b>                | <b>3</b>        |                            |
| *de ő is nagyon + Adj                | 2               | 66                         |
| *de ő is nagyon + élvezte            | 1               | 33                         |
| <b>HOGY MI A KEDVENC</b>             | <b>3</b>        |                            |
| *hogy mi a kedvenc + N_Poss          | 3               | 100                        |
| <b>ÉN IS HASONLÓ A</b>               | <b>3</b>        |                            |
| nekem is hasonló a                   | 3               | 100                        |
| *nekem is hasonló a helyzet          | 2               | 66                         |
| *nekem is hasonló a véleményem       | 1               | 33                         |
| <b>ÉN NEM NAGYON VAN</b>             | <b>3</b>        |                            |
| nekem nem nagyon van                 | 3               | 100                        |

## Melléklet D: Szünetek helye (Praat vizualizációk)

### Szünetek helye én *is így vagyok (vele/ezzel)* kimondása közben (Praat vizualizáció)





én\_is\_így\_vagyok\_ezzel\_3\_\_Mon\_Kirándulás\_001\_  
18.748576

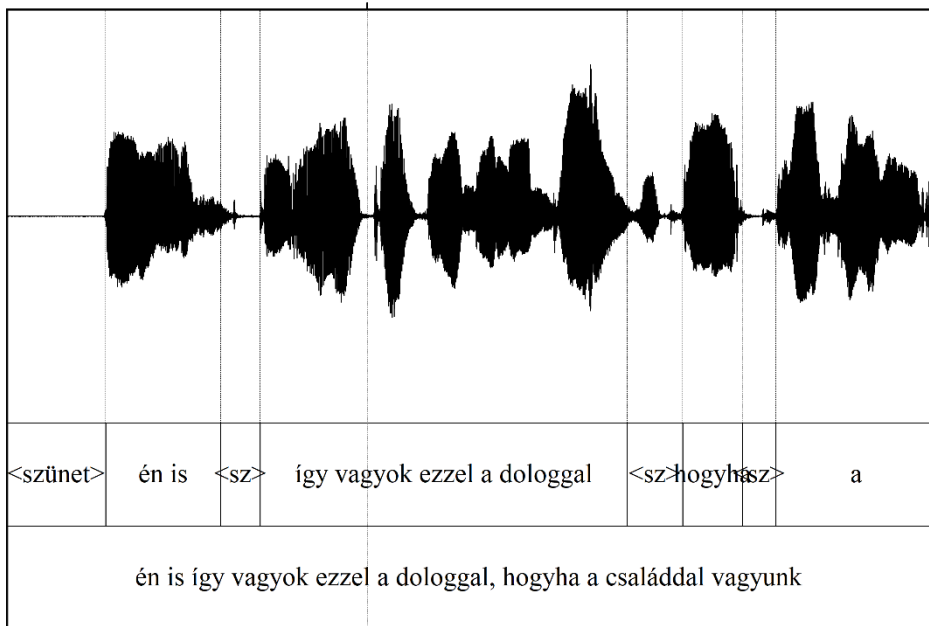


16.31

Time (s)

20.06

én\_is\_így\_vagyok\_ezzel\_4\_\_Mon\_Kirándulás\_001\_  
9.20953228



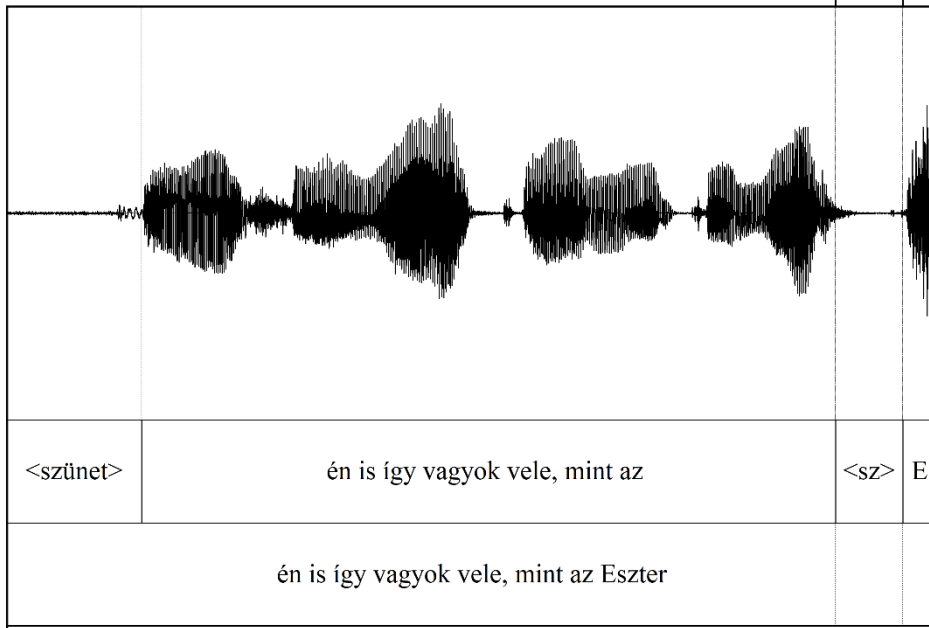
7.757

Time (s)

11.51

én\_is\_így\_vagyok\_vele\_5\_\_Mon\_Sorozatok\_001\_

18.4321826685261



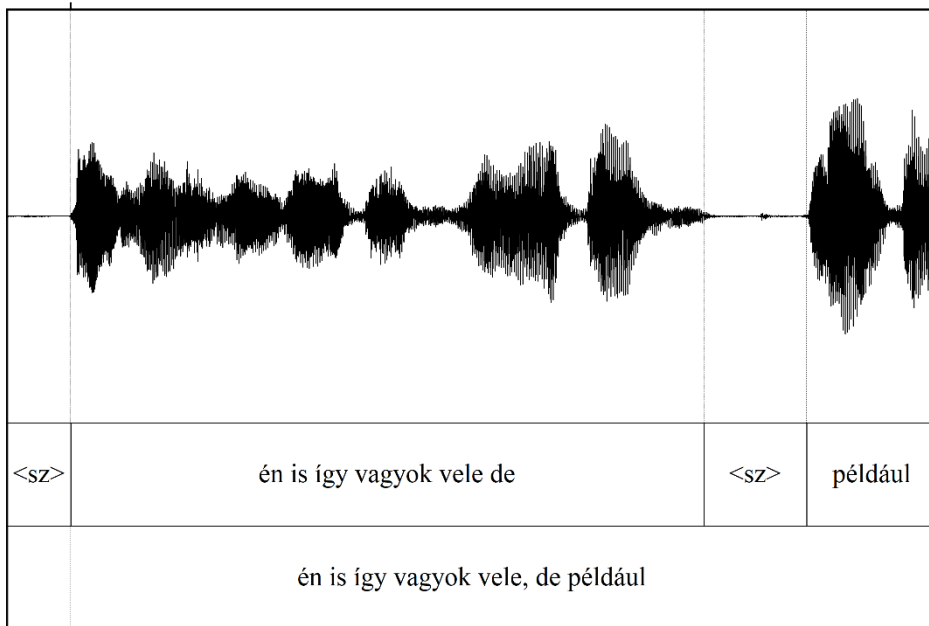
16.76

18.64

Time (s)

én\_is\_így\_vagyok\_vele\_7\_\_Mon\_Utazás\_001\_

8.85161905

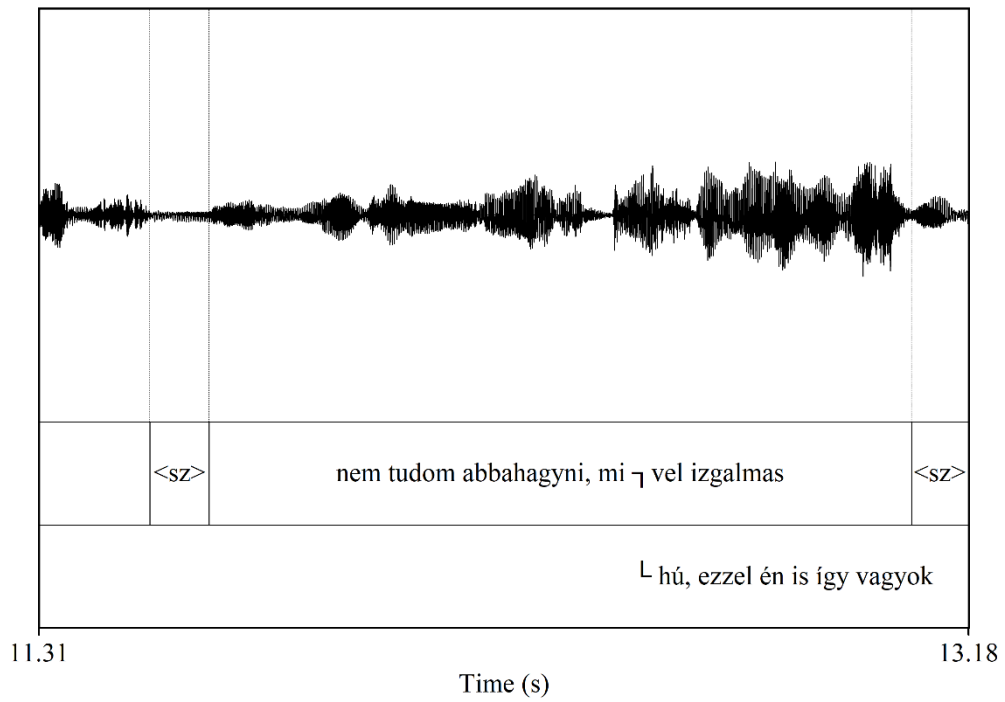


8.718

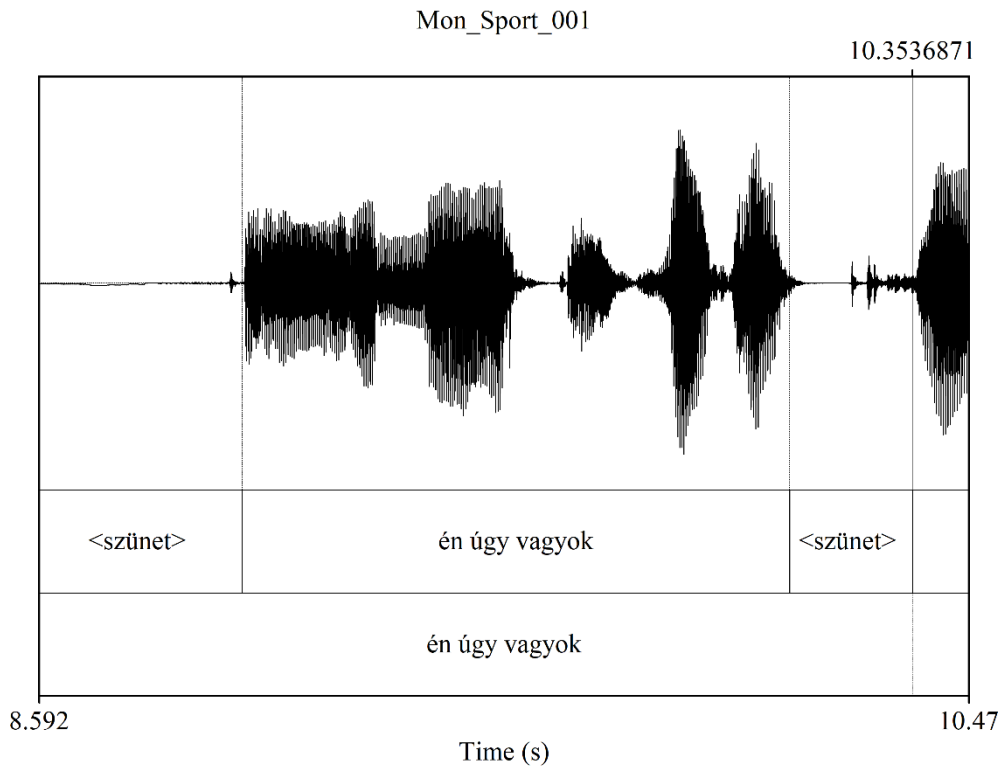
10.67

Time (s)

ezzel\_én\_is\_így\_vagyok\_6\_Mon\_Sorozatok\_002\_



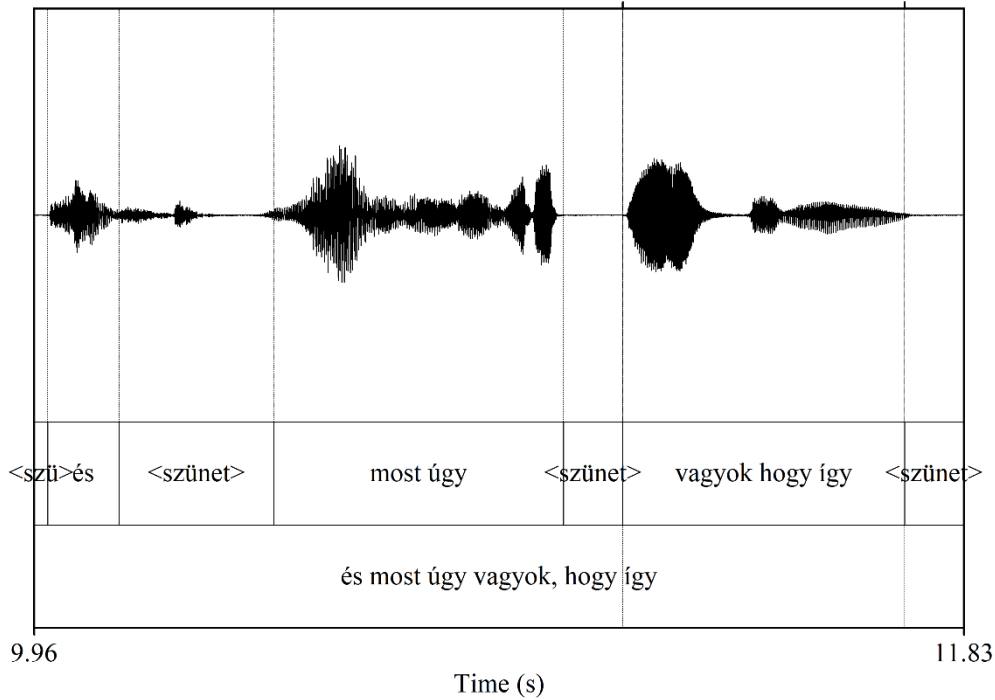
**Szünetek helye az (én) úgy vagyok (vele) kimondása közben (Praat vizualizáció)**



Mon\_Sorozatok\_002

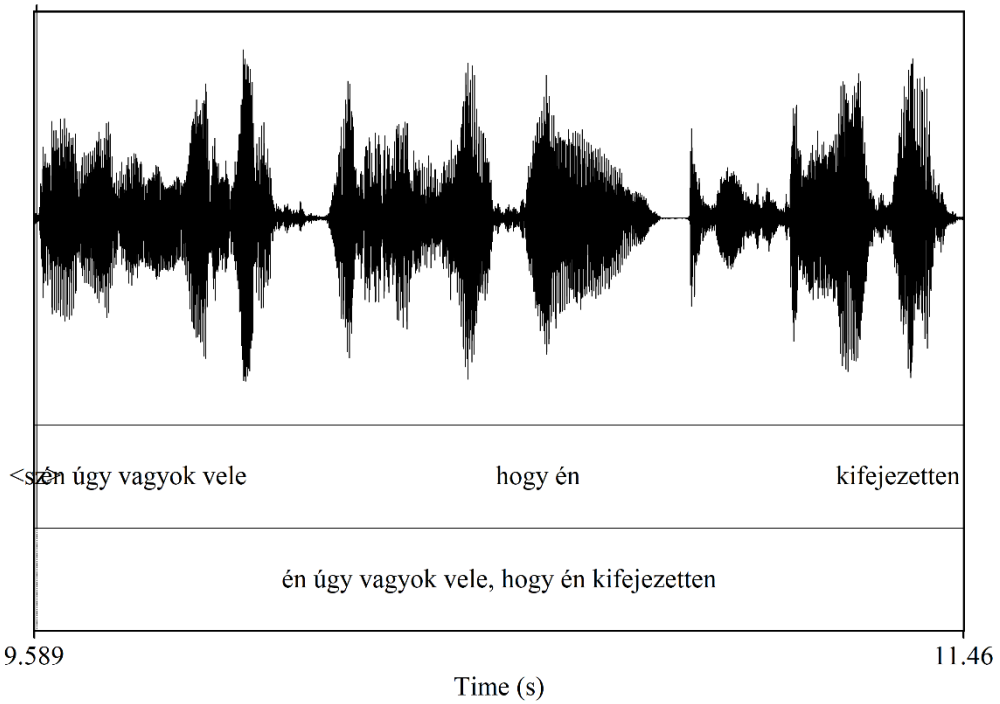
11.1471474

11.7151474

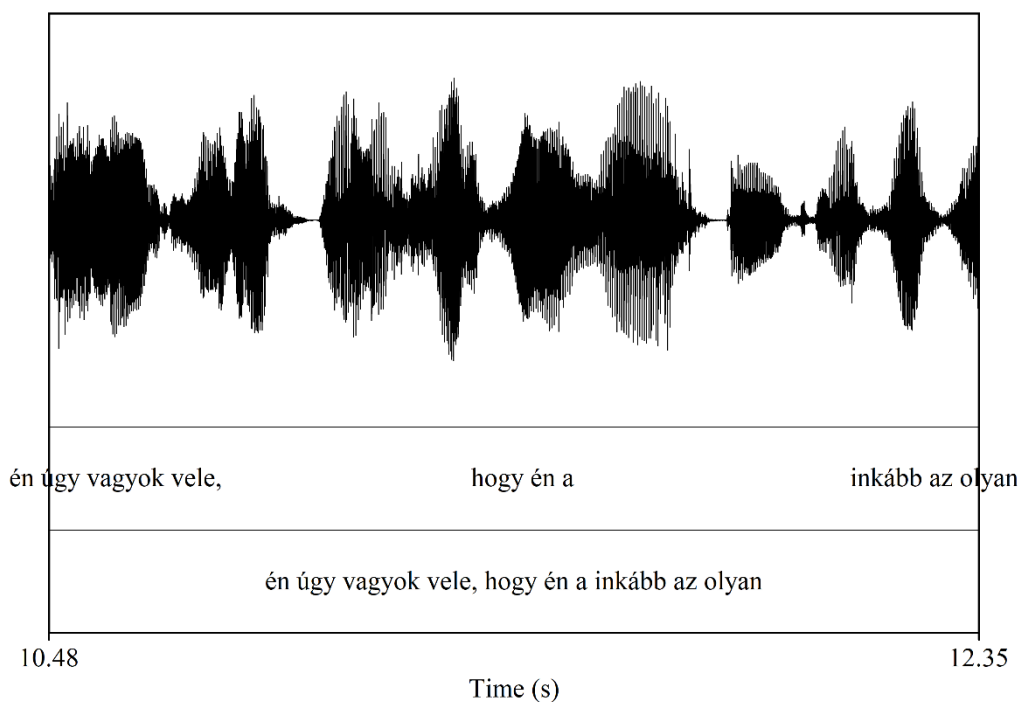


Mon\_Kirándulás\_001

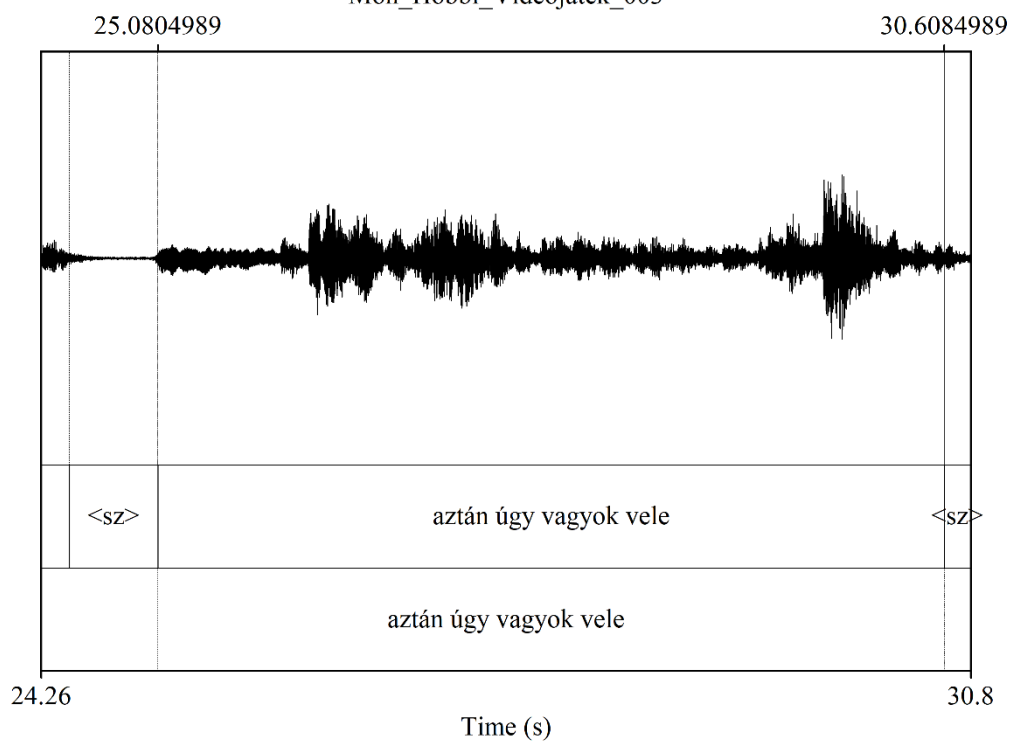
9.59464853



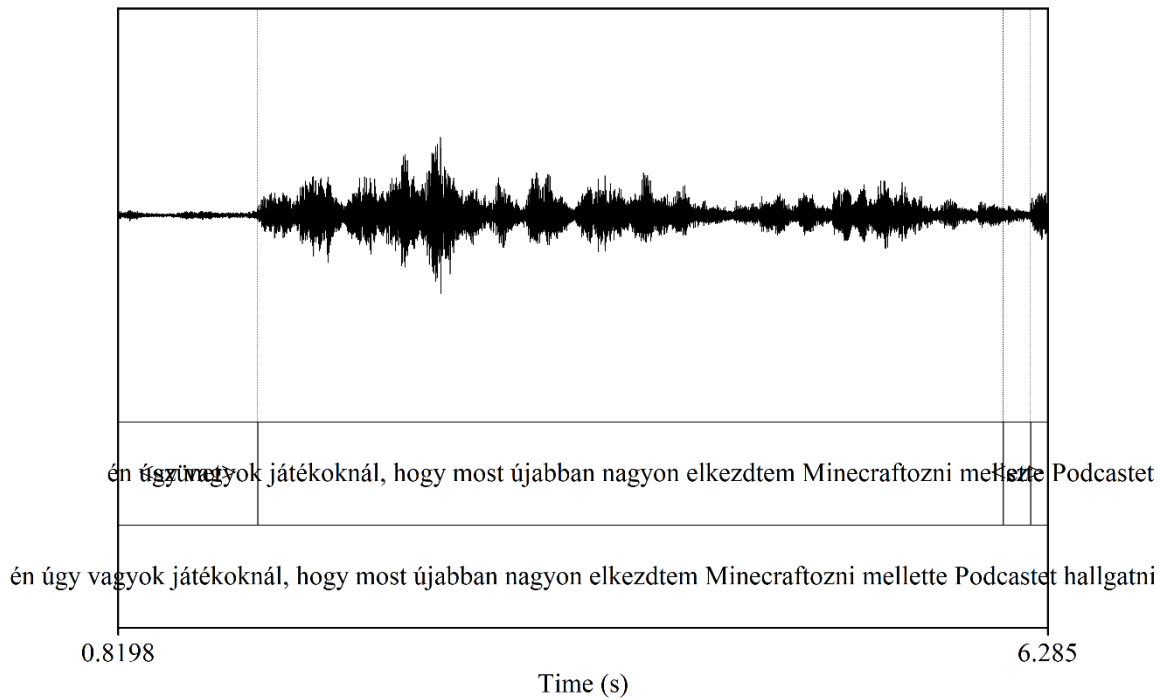
Mon\_Hp\_002



Mon\_Hobby\_Videójáték\_003



Mon\_Hobbi\_Videójáték\_003



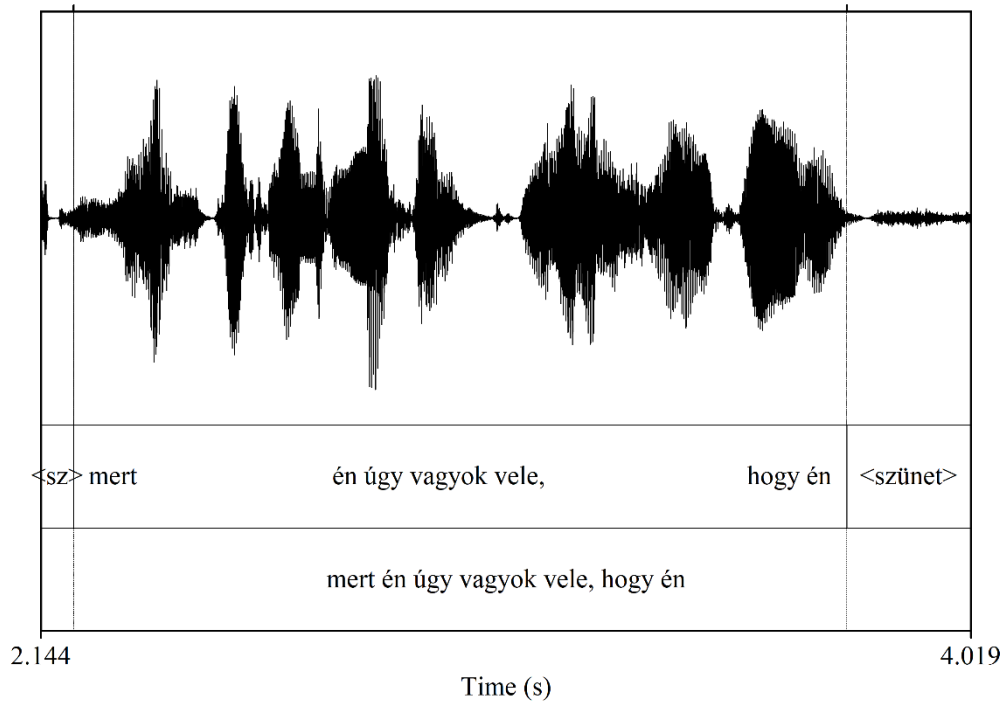
én úgy vagyok játékoknál, hogy most újabban nagyon elkezdtem Minecraftozni mellette Podcastet

én úgy vagyok játékoknál, hogy most újabban nagyon elkezdtem Minecraftozni mellette Podcastet hallgatni

2.20941043

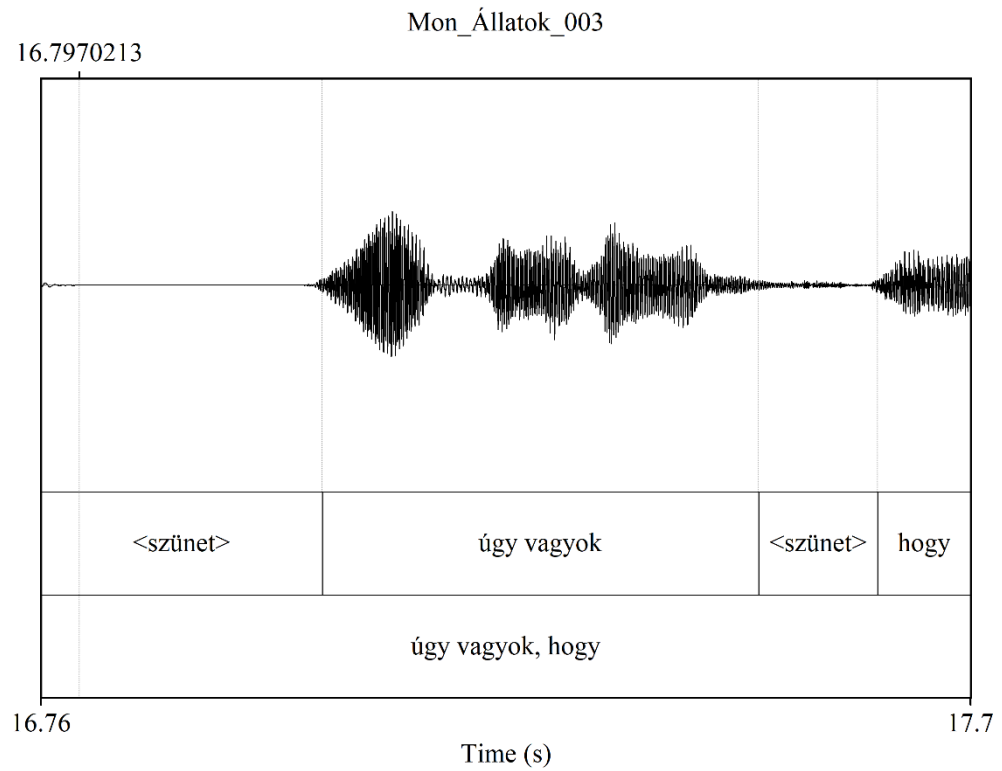
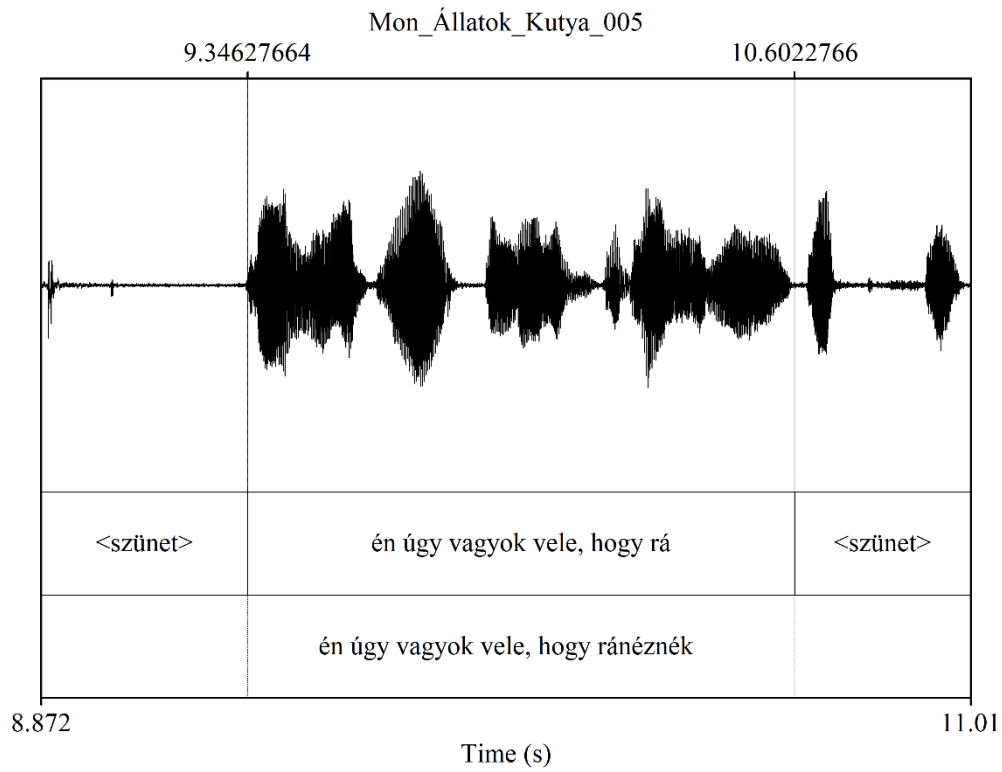
Mon\_Hobbi\_003

3.76938776

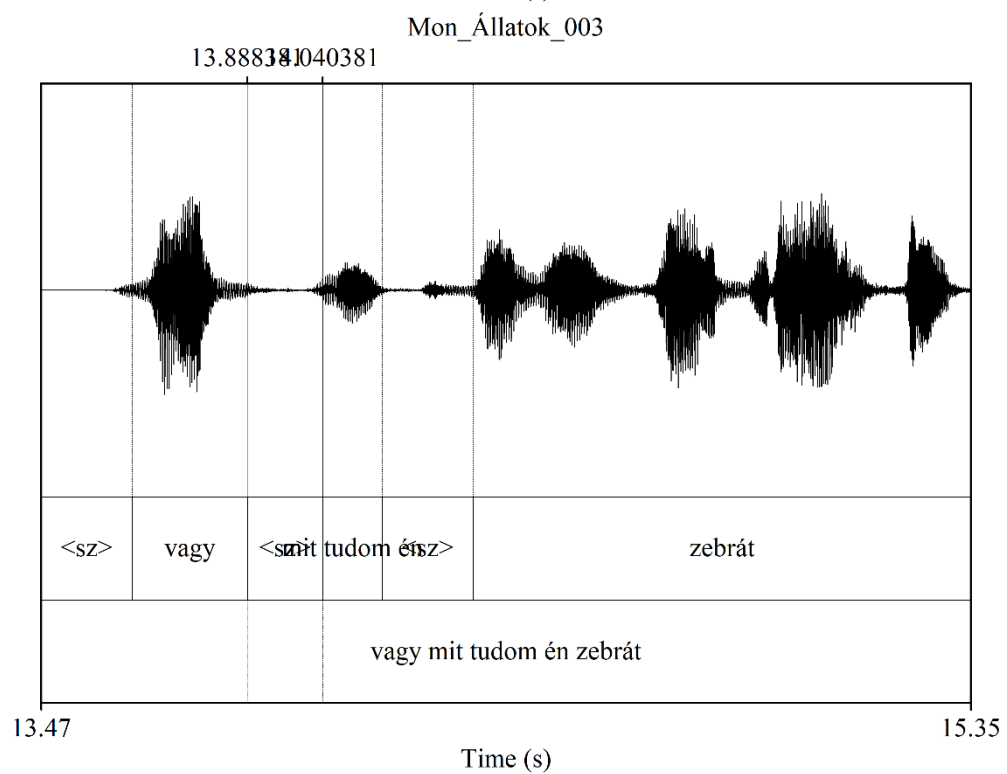
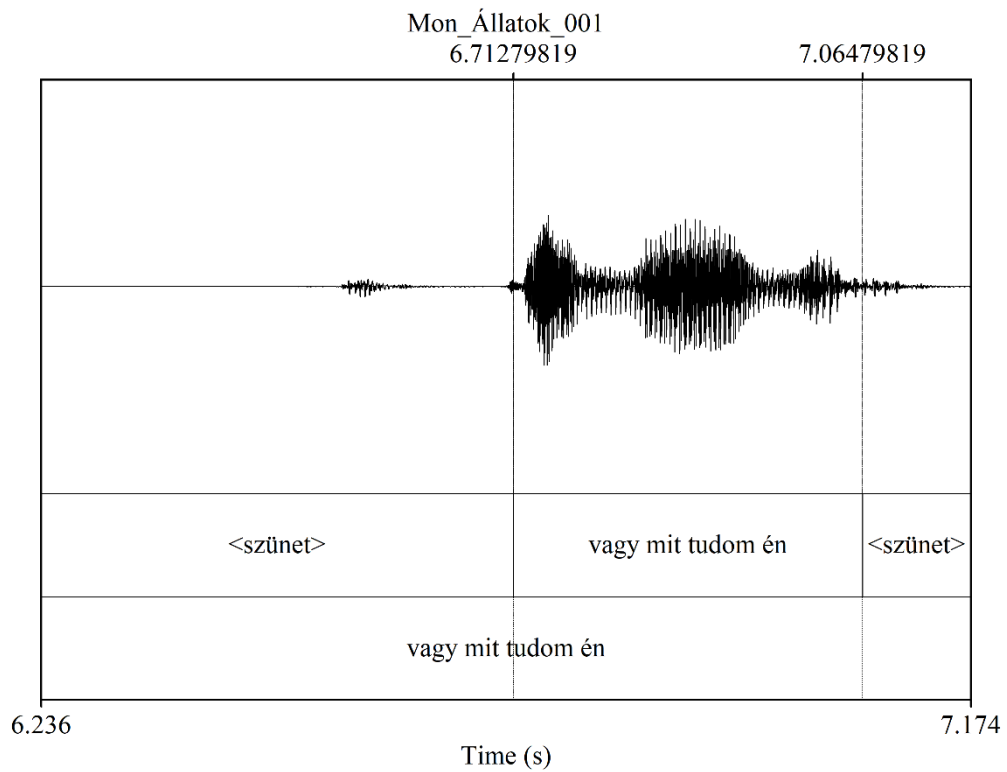


<sz> mert én úgy vagyok vele, hogy én <szünet>

mert én úgy vagyok vele, hogy én



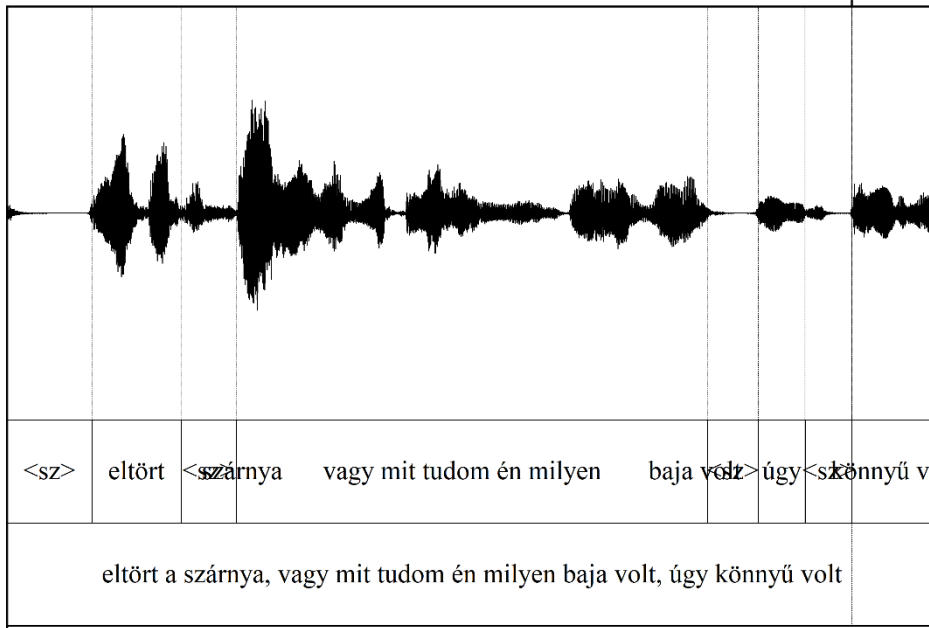
**Szünetek helye az *vagy mit tudom én* kimondása közben (Praat vizualizáció)**





Mon\_Állatok\_Kutya\_006

16.6652426



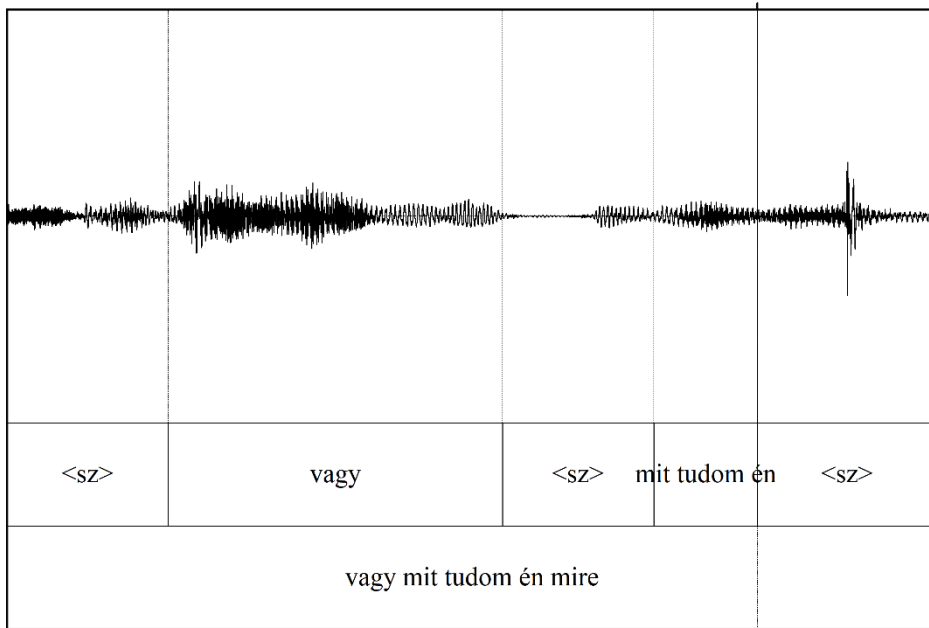
13.48

Time (s)

16.99

Mon\_Sajáttörténet\_003

9.94604082

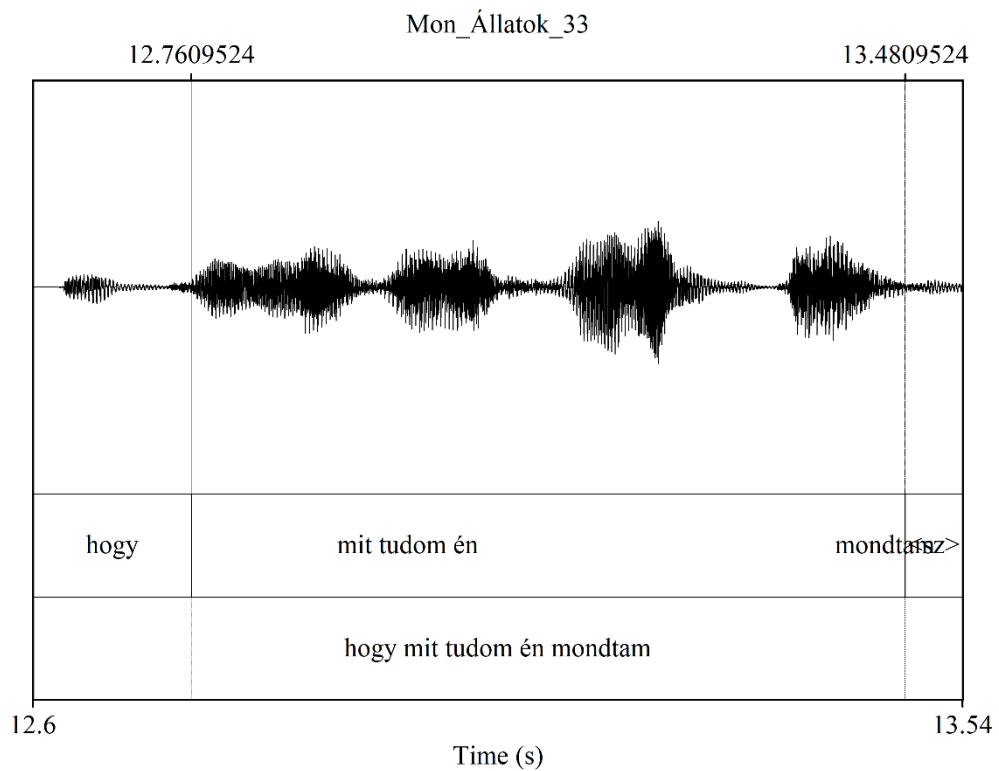
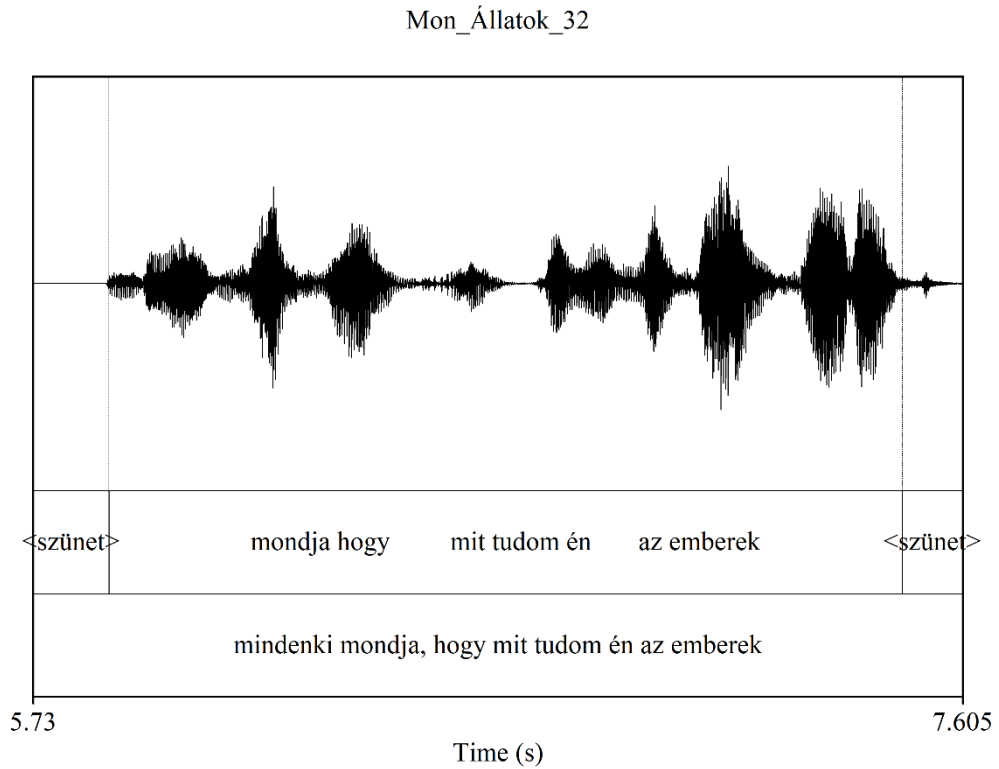


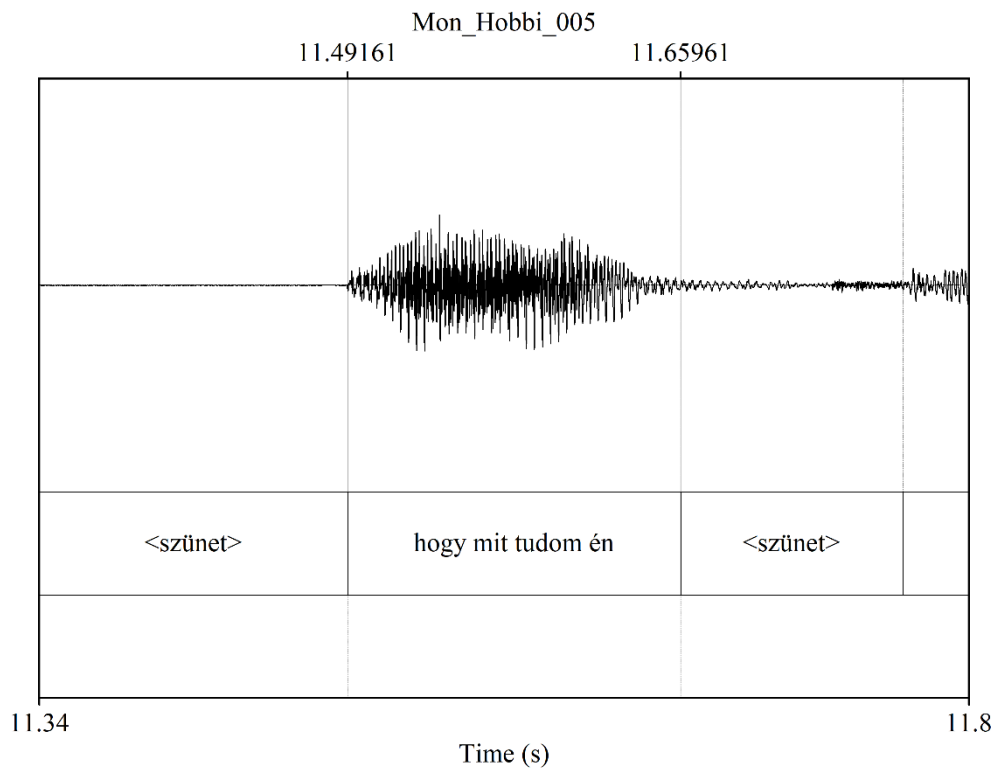
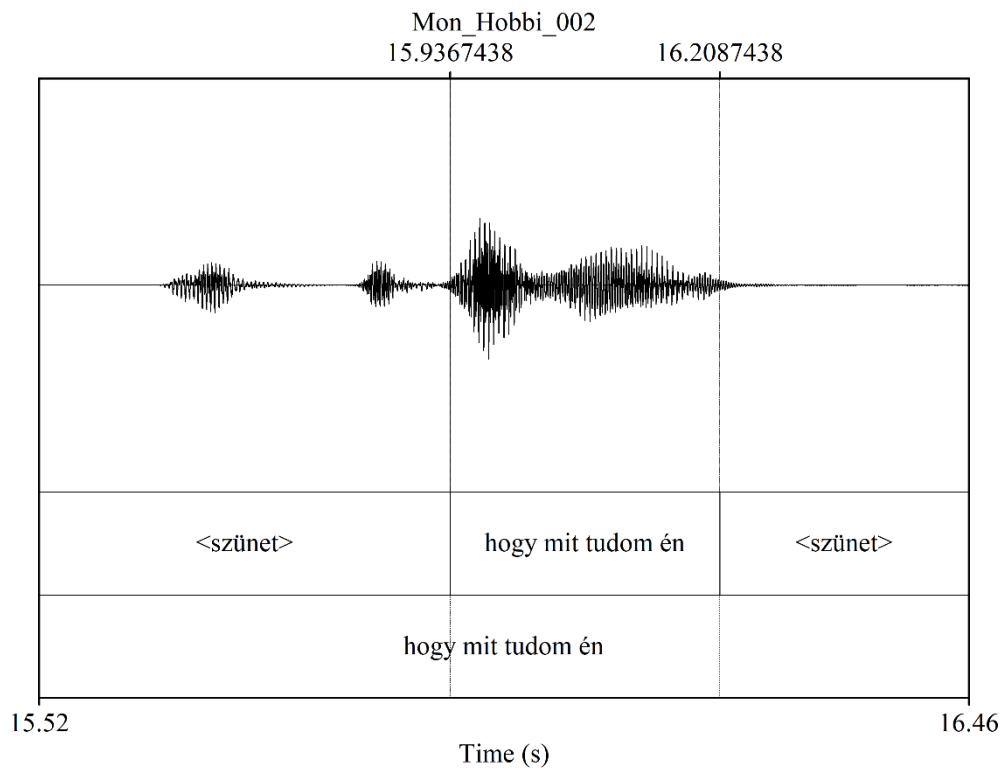
9.192

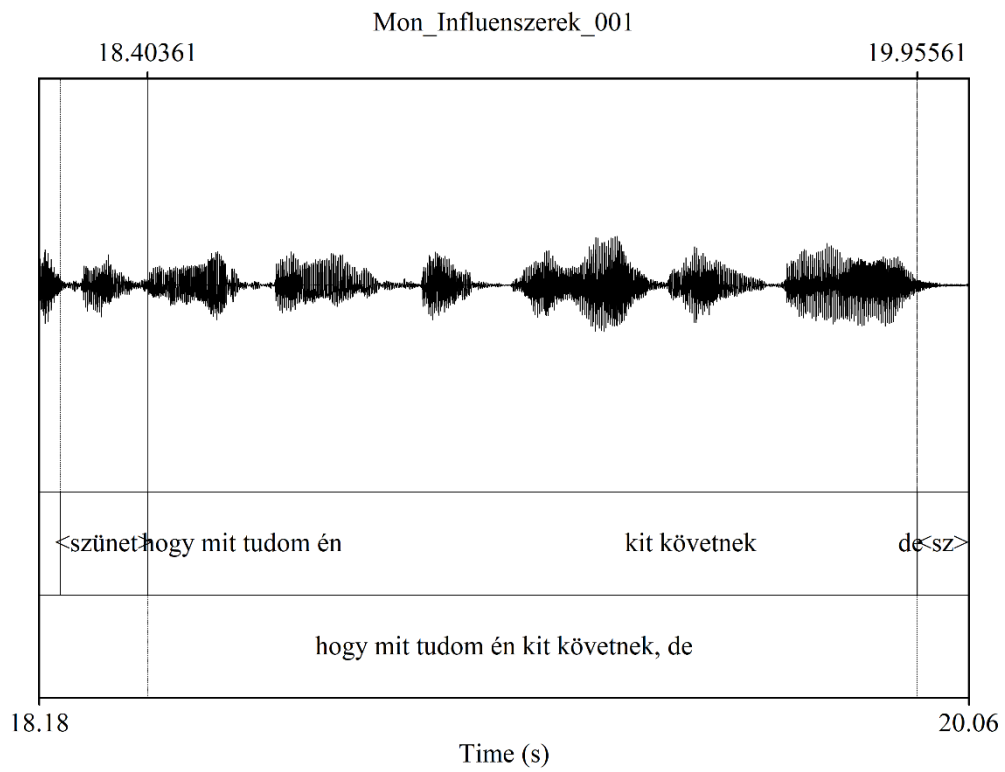
Time (s)

10.13

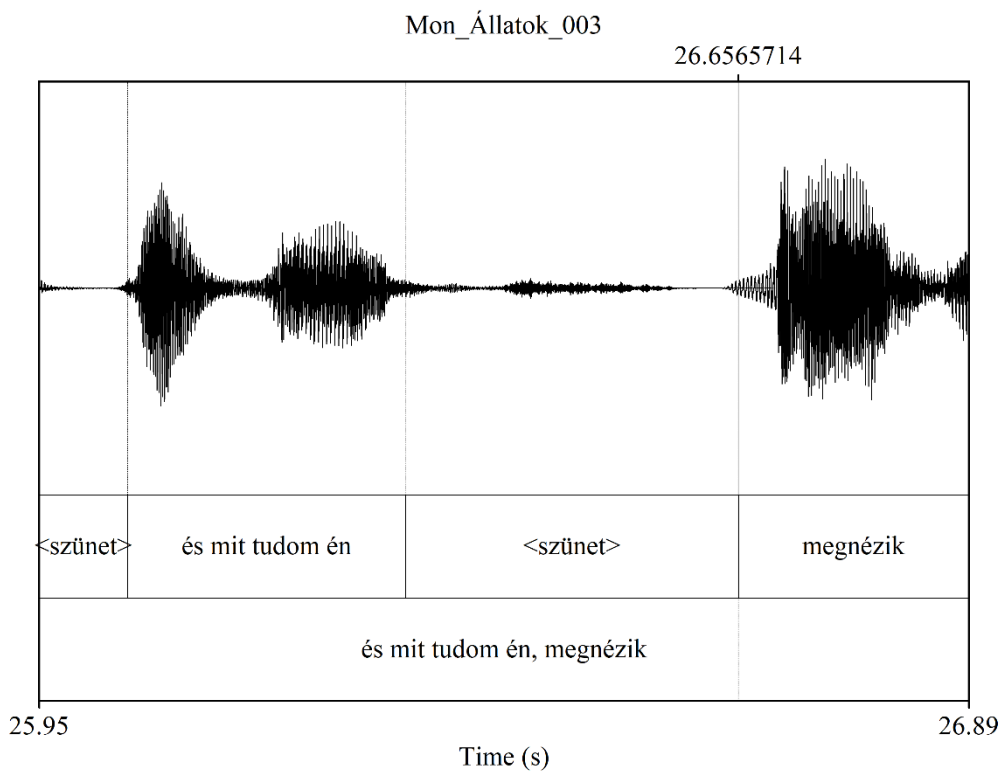
**Szünetek helye a *hogy mit tudom én* kimondása közben (Praat vizualizáció)**



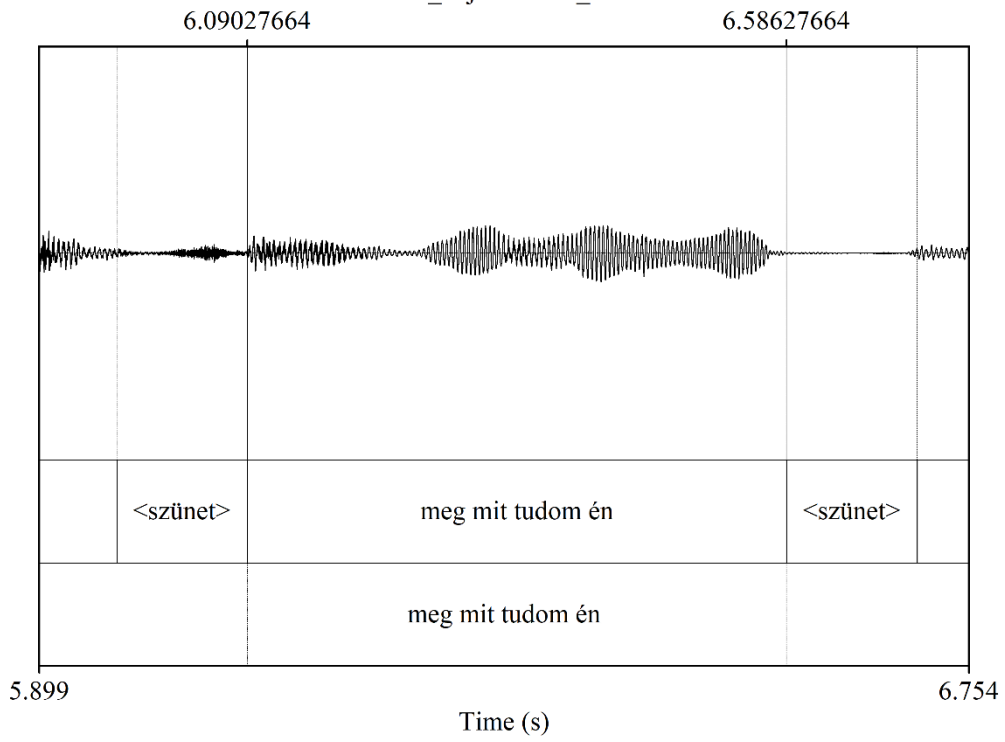




**Szünetek helye a *és/meg mit tudom én* kimondása közben (Praat vizualizáció)**



Mon\_Sajáttörténet\_002



## Melléklet E: Példa egy konkordanciasorra (VAN EGY ILYEN)



CQL [lemma\_lc=="van"] [lemma\_lc=="egy"] [lemma\_lc=="il... • 20  
279.54 per million tokens • 0.028%

Sort word ×



|    | Details | Left context                             | KWIC  | Right context                              |
|----|---------|--|---|--|
| 1  | doc#6   | s gyíkcokkáért. - Igen.</s><s>Neki is    | <b>volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | , már az utolsó napokban, meg minde        |
| 2  | doc#38  | például egész sokat hímeztem, meg        | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | , nem tudom, hogy jó vagy rossz szok       |
| 3  | doc#6   | Mindig mentek valahova, csak biztos      | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | . - Most gondolkodom.</s><s>Hát így        |
| 4  | doc#6   | . - Igen, arra emlékszem. - Nekünk is    | <b>volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | barátsgós játékunk, ami konkrétan a        |
| 5  | doc#39  | k szerettem volna.</s><s>Ó, ezzel is     | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | érdekes történet, hogy már le volt zsír    |
| 6  | doc#73  | /s><s>De igazából, hogy ha mondjuk       | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | érdekes város, meg jó történetek fűző      |
| 7  | doc#65  | ak természetesen ilyenek, de abban       | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | fantasy rész is, amit én viszont nagyo     |
| 8  | doc#1   | mlékeztek, mikor még osztályfőnökin      | <b>volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | feladat Zsófi nénivel, hogy három állat    |
| 9  | doc#1   | öli a kígyókat, de szó szerint.</s><s>   | <b>Volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | gyereknep, ahol hoztak állatokat, és k     |
| 10 | doc#39  | tudom.</s><s>Egyszer kiskoromban         | <b>voltam</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:1Sg Det_art_NDef Adj_Pro:Nom  | hobby, kreativitás táborban, nem tudor     |
| 11 | doc#53  | zekedései, amiket feltesznek.</s><s>     | <b>Van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | Instagram, ahol egy Celeb botrányok,       |
| 12 | doc#7   | ak. - Hát nekünk pont a szomszédban      | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | kutyus, úgy hívják, hogy Maci.</s><s>      |
| 13 | doc#54  | het sétálni.</s><s>Meg még mellette      | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | nagy rét is. - És ti miért szerettek kime  |
| 14 | doc#65  | ak. - Ezzel én is egyetértek, merthogy   | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | nagyon sötét horrorfilm, és egyszer cs     |
| 15 | doc#39  | : az emlékeim, emlékszem arra, hogy      | <b>volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | pajtahelység, ahol volt az ilyen szabac    |
| 16 | doc#59  | <s>Bocsátsd meg nekem!</s><s>És          | <b>volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | rozoga kunyhó, ott fent ilyen mindenfé     |
| 17 | doc#6   | öl.</s><s>Szóval, meg - Az kedves. -     | <b>Van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | speciális játékok, amikor így a Rozi ko    |
| 18 | doc#2   | úgy érted, mint te, vagy én. - Erről is  | <b>van</b> <b>egy</b> <b>ilyen</b><br>V:Prs:NDef:3Sg Det_art_NDef Adj_Pro:Nom     | szívszorító.</s><s>Gréti egy kutya fel     |
| 19 | doc#8   | íttük megint sétálni a futtatóba, és ott | <b>volt</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:3Sg Det_art_NDef Adj_Pro:Nom    | vizslakeverék, amivel fogták ketten, és    |
| 20 | doc#73  | oltunk, meg most meg ne kövesd, de       | <b>voltunk</b> <b>egy</b> <b>ilyen</b><br>V:Pst:NDef:1Pl Det_art_NDef Adj_Pro:Nom | völgyben, ami ott lent volt a lent folyt a |

SORTED. JUMP TO... ↻

## Melléklet F: A szülői hozzájáruló nyilatkozat

### HOZZÁJÁRULÓ NYILATKOZAT

személyes adatok kezeléséhez, továbbá  
képmás-, és hangfelvétel készítéséhez és felhasználásához

Az Európai Parlament és a Tanács (EU) 2016/679 rendelete (a továbbiakban: GDPR) 6. cikk (1) bekezdés a) pontja és a Polgári Törvénykönyvről szóló 2013. évi V. törvény 2:48. §-a alapján

alulírott, ..... (szülő, gondviselő)  
**hozzájárok és engedélyezem, hogy** gyermekem  
..... (gyermek neve) beszélgetéseiről készült  
felvételeket a Korpusznyelvészeti és Szakmódszertani Kutatócsoport (székhely: Pécs,  
Damjanich u. 30, 7624) beszéltnyelvi kutatásaihoz korlátlan ideig felhasználja, azok leiratait  
elérhetővé tegye későbbi kutatásaik számára.

**Megértetem**, hogy a gyermekek felvételei anonim módon kerülnek felhasználásra, sem a személyes adatok, sem maguk a felvételek nem kerülnek nyilvánosságra, csak a felvételek leiratai lesznek elérhetőek a kutatócsoport munkatársai számára.

Gyermekekről készülhetnek **videófelvelelek**:      IGEN                      NEM

Gyermekekről készülhetnek **hangfelvételek**:      IGEN                      NEM

**Tudomásul veszem**, hogy a személyes adatok kezelésével kapcsolatos jogaimat (adatokhoz való hozzáférés, adatok helyesbítése, törlése, tiltakozás az adatok kezelése ellen) a GDPR 15.-21. cikke (Infotv.14-21. §), a jogorvoslatra vonatkozó szabályokat (adatvédelmi hatósághoz, bírósághoz fordulás joga jogsérelem esetén) a GDPR 77. és 79. cikk (Infotv.22-23. §) tartalmazza.

Kelt:....., 202.....

.....  
szülő/gondviselő

.....  
szülő/gondviselő

## Melléklet G: Példa a Magyar felfedezőhöz készített kifejezésbankokból

### 4.fejezet A1



NÉZD EZT A SOK ÁLLATOT!

#### 4.FEJEZET – A1

#### Szószeret és hasznos kifejezések

##### Rövidítések:

fn. = főnév,  
mn.= melléknév,  
i. =ige,

#### SZAVAK

állatkert (fn.)  
elefánt (fn.)  
élelem (fn.)  
élőhely (fn.)  
hal (fn.)  
háziállat (fn.)  
illik valamihez / valakihez (i.)  
kenguru (fn.)  
ló (fn.)  
madár (fn.)  
magasság (fn.)  
majom (fn.)  
medve (fn.)  
oroszlán (fn.)  
őz (fn.)  
pingvin(fn.)  
róka (fn.)  
sün (fn.)  
testsúly (fn.)  
tigris (fn.)  
vadállat (fn.)  
zebra (fn.)  
zsiráf (fn.)

#### HASZNOS KIFEJEZÉSEK

A kutya a kedvenc állatom.

A súlya akár 60 kilogramm is lehet.

A vadon élő állatokat vagy a háziállatokat szereted inkább?

Ausztráliában él.

Melyik állatot szereted?

Mi a kedvenc állatod?

Mindegyik ugyanolyan fontos / aranyos /...

Nagyon szeretem az állatokat.

Nekem is hasonló a véleményem.

Növényeket / húst eszik.

Szereted az állatokat?

Van háziállatod?





NÉZD EZT A SOK ÁLLATOT!

## 4.FEJEZET – A1

### Kérdések és válaszok

#### 1. Melyik állat **hol** él?

- A hal a folyó**ban** él.
- A hal a víz**ben** él.
- A kenguru Ausztráliá**ban** él.
- A ló a mező**n** él.
- A ló Magyarorszá**gon** él.
- A madár az erdő**ben** él.
- A majom Ázsiá**ban** él.
- A medve az erdő**ben** él.
- A medve Európá**ban** él.
- A pingvin a Déli-sark**on** él.
- A tigris Ázsiá**ban** él.
- A zebra Afriká**ban** él.
- Az elefánt Afriká**ban** él.
- Az oroszlán Afriká**ban** él.

#### 2. **Milyen állatok vannak az állatkertben?**

- Elefánt**ok**.
- Hal**ak**.
- Kenguru**ok**.
- Madar**ak**.
- Maj**ok**.
- Medv**ék**.
- Oroszlán**ok**.
- Pandá**ok**.
- Pingvin**ek**.
- Tigris**ek**.
- Zebrá**ok**.
- Zsiráf**ok**.

#### 3. **Melyik országban van a város?**

- Ránua Finnorszá**gban** van.
- Berlin Németorszá**gban** van.
- Barcelona Spanyolorszá**gban** van.
- Torontó Kanadá**ban** van.
- Madrid Spanyolorszá**gban** van.
- Budapest Magyarorszá**gon** van.

#### 4. **Melyik állatot szereted?**

- Nagyon szeretem a hal**akat**.
- Nagyon szeretem a kenguru**kat**.
- Nagyon szeretem a kutyá**kat**.
- Nagyon szeretem a macská**kat**.
- Nagyon szeretem a madar**akat**.
- Nagyon szeretem a medv**éket**.
- Nagyon szeretem a róká**kat**.
- Nagyon szeretem az őz**eket**.

## 4.fejezet A2



NÉZD EZT A SOK ÁLLATOT!

### 4. fejezet /A2

#### Szószeret és hasznos kifejezések

#### SZAVAK

bagoly (fn.)

bálna (fn.)

bolodos (mn.)

büdös (mn.)

disznó (fn.)

drága (mn.)

erős (mn.)

farkas (fn.)

gepárd (fn.)

gyors (mn.)

halk (mn.)

hangos (mn.)

hideg (mn.)

hiperaktív (mn.)

hosszú (mn.)

jó (mn.)

kakas (fn.)

kevés (mn.)

kicsi (mn.)

könnyű (mn.)

lassú (mn.)

magas (mn.)

magányos (mn.)

meleg (mn.)

mély (mn.)

mókus (fn.)

nagy (mn.)

nehéz (mn.)

nyúl (fn.)

olcsó (mn.)

öreg (mn.)

papagáj (fn.)

pillangó (fn.)

róka (fn.)

széles (mn.)

szép (mn.)

teknős (fn.)

tyúk (fn.)

ugrálós (mn.)

vad (mn.)

vak (mn.)

veszélyeztetett (mn.)

#### Rövidítések:

fn. = főnév,  
mn.= melléknév,  
i. =ige,



NÉZD EZT A SOK ÁLLATOT!

#### 4. fejezet /A2

### Szószedet és hasznos kifejezések

#### HASZNOS KIFEJEZÉSEK

De jó neked!

Nektek mi a kedvenc kutyafajtátok?

Nagyon szeretném, (hogy)...

Kérlek szépen!

Légy szíves!

A kedvemért!

Biztos, hogy...

Hidd el, hogy ...

Ráadásul...

Ezenkívül,...

Megígérem, hogy...

Ha ..., akkor...

Biztos lehetsz benne, hogy...



NÉZD EZT A SOK ÁLLATOT!

#### 4.FEJEZET – A1

#### Kérdések és válaszok

##### 1. Mit esznek az állatok?

- A bagoly húsevő (állat).
- A bálna húsevő (állat).
- A disznó mindenevő (állat).
- A farkas húsevő (állat).
- A gepárd húsevő (állat).
- A kakas növényevő (állat).
- A mókus növényevő (állat).
- A nyúl növényevő (állat).
- A papagáj növényevő (állat).
- A pillangó növényevő (állat).
- A róka húsevő (állat).
- A tyúk növényevő (állat).

##### 2. Melyik állat hol él?

- A bagoly az erdőben él.
- A bálna az óceánban él.
- A disznó az udvaron él.
- A farkas az erdőben él.
- A gepárd a szavannán él.
- A kakas az udvaron él.
- A mókus az erdőben él.
- A nyúl a mezőn él.
- A papagáj az erdőben él.
- A róka a mezőn él.
- A tyúk az udvaron él.
- A pillangó az erdőben él.

##### 3. Melyik állat mit tud?

- A bagoly tud repülni.
- A bálna tud úszni.
- A farkas tud futni.
- A mókus tud ugrani.
- A nyúl tud ugrani.
- A pillangó tud repülni.
- A róka tud futni.

## 4.fejezet B1



NÉZD EZT A SOK ÁLLATOT!

### 4. fejezet /B1

#### Szószedet és hasznos kifejezések

##### SZAVAK

akvárium (fn.)  
állandó (mn.)  
állathordozó (fn.)  
ápol valakit / valamit (i.)  
barátságos (mn.)  
csendes (mn.)  
élettér (fn.)  
emlős (fn.)  
fajta (fn.)  
fekhely (fn.)  
feltűnő (mn.)  
fésű (fn.)  
gondoz valakit / valamit (i.)  
görény (fn.)  
gyapjú (fn.)  
hangos (mn.)  
használják valamire (i.)  
hiperaktív (mn.)  
hőmérséklet (fn.)  
hüllő (fn.)  
igényel valamit (i.)  
igénytelen (mn.)  
ikra (fn.)  
illik valamihez / valakihez (i.)

intelligens (mn.)  
játék (fn.)  
játékos (itt: mn.)  
kalitka (fn.)  
kaparófa (fn.)  
kényes (mn.)  
kételtű (fn.)  
ketrec (fn.)  
kígyó (fn.)  
kíváncsi (mn.)  
kopolyú (fn.)  
közönséges cserepesteknős (fn.)  
kutyatáp (fn.)  
labda (fn.)  
láma (fn.)  
macskaalom (fn.)  
macskabútor (fn.)  
madáreledel (fn.)  
magányos (mn.)  
mozgékony (mn.)  
nevelés (fn.)  
nyakörv, (fn.)  
nyugodt (mn.)  
óshaza (fn.)  
pete (fn.)  
pikkelyes (mn.)

##### Rövidítések:

fn. = főnév,  
mn.= melléknév,  
i. =ige,



NÉZD EZT A SOK ÁLLATOT!

sajátosság (fn.)  
származás (fn.)  
szelíd (mn.)  
szőrös (mn.)  
tál (fn.)  
táplál valakit / valamit (i.)  
táplálkozás (fn.)  
társas (itt: mn.)  
tartás (fn.)  
teherhordás (fn.)  
tengerimalac (fn.)  
tenyésztés (fn.)  
tollas (mn.)  
törpemalac (fn.)  
uszony (fn.)  
vad (mn.)  
változó (mn.)  
védett (mn.)  
viselkedés (fn.)

## HASZNOS KIFEJEZÉSEK

A kedvemért!  
Annyira / Olyan jó lenne.  
Bárcsak kapnék egy lámát!  
Biztos, hogy...  
Dél-Amerikában teherhordásra használják.  
Egy próbát megér.  
Én inkább nem szeretnék lámát.  
Ezenkívül...  
Gondold át!  
Ha lenne elég pénzem, vennék egy...  
Ha..., akkor...  
Hidd el, hogy...  
Kaphatnék egy lámát?  
Kérlek szépen!  
Légy szíves!  
Megígérem, hogy...  
Milyen felszerelésre lenne szükséged, ha lenne egy...  
Mind minden tevéfélnék, nekik is hosszú nyakuk van.  
Nagyon szeretném, (hogy)...  
Neked is tetszene / jót tenne.  
Nekünk állatkertünk van.  
Nem igényel különleges étrendet.  
Ráadásul...  
Te is biztosan szívesen...  
Úgy örülnék neki.



NÉZD EZT A SOK ÁLLATOT!

## PROJEKTHEZ TARTOZÓ HASZNOS KIFEJEZÉSEK

állatkerti séta  
az állatok csodálatos világa  
az év minden napján  
csodálatos környezetben  
élővilág  
fedezd fel  
fókashow  
látványtetések  
leglátogatottabb állatkert  
Magyarországon honos  
nagy élmény  
találkozz  
természetes élőhelyek  
vadon élő állatok  
várja látogatóit

## KÉRDÉSEK ÉS VÁLASZOK

### 1. **Mire** lenne szükséged, ha lenne egy háziállatod?

- Szükségem lenne egy akvárium**ra**.
- Szükségem lenne egy állathordozó**ra**.
- Szükségem lenne egy fekhely**re**.
- Szükségem lenne egy fésű**re**.
- Szükségem lenne egy kalitká**ra**.
- Szükségem lenne egy ketrec**re**.
- Szükségem lenne egy nyakörv**re**.
- Szükségem lenne egy tál**ra**.

### 2. **Mivel** etetnéd a háziállatodat?

- Gyümölc**ssel** etetném.
- Hú**ssal** etetném.
- Kutyatápp**al** etetném.
- Madáreledel**lel** etetném.
- Magok**kal** etetném.
- Széná**val** etetném.
- Zöldség**gel** etetném.

### 3. **Miben** tartanád a háziállatodat?

- Akvárium**ban** tartanám.
- Ketrec**ben** tartanám.
- Kalitká**ban** tartanám.
- A kert**ben** tartanám.
- A ház**ban** tartanám.
- Kennel**ben** tartanám.



NÉZD EZT A SOK ÁLLATOT!

4. **Mit vennél egy háziállatnak?**

- Ketre**ce**t vennék.
- Játé**ko**t vennék
- Állathordozó**t** vennék.
- Állateledel**t** vennék.
- Labd**á**t vennék.
- Nyakör**ve**t vennék.
- Fésű**t** vennék.

5. **Melyik állat mit igényel?**

- A csincsilla **különleges étrendet** igényel.
- A kutya szőre **sok ápolást** igényel.
- A láma **társaságot** igényel.
- A tengerimalac **friss gyümölcsöt és zöldséget** igényel.

6. **Mit csinálnál egy háziállattal?**

- Játssz**anék vele**.
- Beszélget**nék vele**.
- Sétál**nék vele**.
- Foglalkoz**nék vele**.
- De! Tanít**anám**.
- De! Vásárol**nék neki**.

7. **Mit csinálnál egy háziállattal?**

- Játssz**anék vele**.
- Beszélget**nék vele**.
- Sétál**nék vele**.
- Foglalkoz**nék vele**.
- De! Tanít**anám**.
- De! Vásárol**nék neki**.



## 4. fejezet B2



NÉZD EZT A SOK ÁLLATOT!

### 4. fejezet /B2

#### Szószeret és hasznos kifejezések

##### SZAVAK

árva (mn.)  
bajba jut (i.)  
belégzés (fn.)  
bolyhos (mn.)  
bundás (mn.)  
egészséges (mn.)  
egyensúlyoz valamit (i.)  
élettér (fn.)  
élőhely (fn.)  
elpusztít valamit / valakit (i.)  
elpusztul (i.)  
erdőtirtás (fn.)  
fajtábla (fn.)  
felépül (i.)  
fogságban tart,  
fóka (fn.)  
füst (fn.)  
gondoz valakit / valamit (i.)  
hegyi gorilla (fn.)  
hulladék (fn.)  
igényel valamit (i.)  
igényes (mn.)  
illik valamihez / valakihez (i.)  
jegesmedve (fn.)  
káros (mn.)

kézjel (fn.)  
kihalás (fn.)  
kipufogógáz (fn.)  
légszennyezés (fn.)  
megbetegszik (i.)  
megmenekül (valahonnan) (i.)  
megsérül (valahol / valamilye) (i.)  
menhely (fn.)  
mérgező (mn.)  
műanyag (fn.)  
nevelés (fn.)  
olaj (fn.)  
óriáspanda (fn.)  
önkéntes (itt: fn.)  
rehabilitációs központ (fn.)  
sérült (mn.)  
szabadon enged valakit / valamit (i.)  
szelíd (mn.)  
szennyez valamit (i.)  
szibériai tigris (fn.)  
táplálék (fn.)  
táplálkozás (fn.)  
tenyésztés (fn.)  
vad (mn.)  
vadász (fn.)  
vadászat (fn.)  
védett (mn.)

##### Rövidítések:

fn. = főnév,  
mn. = melléknév,  
i. = ige,



NÉZD EZT A SOK ÁLLATOT!

veszélyes valami számára (mn.)

veszélyeztetett (mn.)

vezényszó (fn.)

vízszennyezés (fn.)

### HASZNOS KIFEJEZÉSEK

(Nekem) az a véleményem, hogy...

Annyi ház, ahány szokás.

Annyi... van, mint égen a csillag.

Annyit ért hozzá, mint tyúk az ábécéhez.

Árván marad

Az igaz, hogy..., de...

Biztos vagyok benne, hogy...

Ebben igazad van.

Én ezt másképp látom.

Én is ezen a véleményen vagyok.

Én is úgy gondolom, hogy...

Én úgy gondolom, (hogy)...

Igen, de nem gondolod, hogy...

Jelentősen lecsökken a fókák száma.

Meg vagyok győződve arról, hogy...

Nem jutnak annyi táplálékhoz, mint amennyire szükségük lenne.

Sajnos ez nincs mindig így.

Sokszor előfordul, hogy...

Szerintem ebben nincs igazad.

Szerintem inkább...

Szerintem is fontos a / az...

Szerintem...

Teljesen egyetértek veled.

### PROJEKTHEZ TARTOZÓ HASZNOS KIFEJEZÉSEK

a leglátogatottabb állatkert

állatkerti séta

az állatok csodálatos világa

az év minden napján

csodálatos környezetben

éővilág

fedezd fel

látványtetések

Magyarországon honos

mentőmunka

nagy élmény

találkozz

természetes élőhelyek

vadon élő állatok

várja látogatóit

veszélyeztetett faj

## KÉRDÉSEK ÉS VÁLASZOK

### 1. Mi történik velem /veled /vele? ...

**jut**

- Bajba **jut**.
- Eszembe **jut** egy jó ötlet.
- Eszedbe **jut** egy érdekes történet.
- Eszébe **jut** egy izgalmas játék.

### 1. Kap valamit... **jut**

- Táplálékhoz **jut**.
- Pénzhez **jut**.

### 2. Szüksége van valamire... **szorul**

- A fóka segítségre **szorul**.
- A pingvin ápolásra **szorul**.
- A pingvin gondozásra **szorul**.
- A menhely támogatásra **szorul**.
- Az állatorvos autója javításra **szorul**.

### 3. Valaki rákényszerít valakit arra, hogy ne menjen el... **tart**

- Fogságban **tart**.
- Menhelyen **tart**.
- Kórházban **tart**.

### 5. Annyi... amennyi/ahány/mint

- A sikeres vadász **annyi** pénzt kapott, **ahány** fókászót össze tudott gyűjteni.
- Az itt gondozott vad fókák csak **annyi** időt töltenek a rehabilitációs központban, **amennyi** feltétlenül szükséges, utána szabadon engedik őket.
- A menhelyre azonban nem kerül be **annyi** fóka, **ahány** bejelentés érkezik.
- Gyakori, hogy a halászat miatt a felnőtt fókák nem jutnak **annyi** táplálékhoz, **mint** amennyire szükségük lenne.

## 4.fejezet C1



NÉZD EZT A SOK ÁLLATOT!

### 4. fejezet /C1

#### Szószedet és hasznos kifejezések

##### SZAVAK

áldozat (fn.)  
beolvad valahova (i.)  
biodiverzitás (fn.)  
császárharkály (fn.)  
császárpingvin (fn.)  
csimpánz (fn.)  
Dávid-szarvas (fn.)  
dodó (fn.)  
elengedhetetlen,  
élőhely (fn.)  
elrejt valamit valahova (i.)  
emberiség (fn.)  
észrevesz valamit (i.)  
fajtiszta (mn.)  
fehérfejű rétisas (fn.)  
fekete lágyhéjúteknős (fn.)  
fenyegetett (mn.)  
illik valamihez / valakihez (i.)  
jegesmedve (fn.)  
jog (fn.)  
kaméleon (fn.)  
karibi barátfóka (fn.)  
kék bálna (fn.)  
keverék (mn.)  
kihal (fn.)

kondor keselyű (fn.)  
mandrill (fn.)  
óriás panda (fn.)  
oroszlán (fn.)  
ragadozó (fn.)  
sáska (fn.)  
sebezhető (mn.)  
válík valamivé (i.)  
védekezik (valami ellen) (i.)  
veszélyeztetett (mn.)  
zebra (fn.)

##### Rövidítések:

fn. = főnév,  
mn.= melléknév,  
i. =ige,

##### HASZNOS KIFEJEZÉSEK

beolvad a környezetébe  
eszébe jut  
felhívja a figyelmét valamire  
hasonlóvá válík  
lehetővé tesz  
megszólalásig hasonlít  
nehézséget okoz

## INTERJÚ

(Nekem) az a véleményem, hogy...  
Arról érdeklődnék, hogy...  
Azon gondolkodtam, hogy...  
Azt gondolom, hogy...  
Azt hiszem, hogy...  
Biztos vagyok benne, hogy...  
Én úgy gondolom, hogy...  
Ez egy nehéz kérdés...  
Ez jó kérdés...  
Hadd gondolkodjam...  
Hát attól függ, hogy...  
Hát...  
Hogy érted /éri, hogy...  
Még sosem gondolkodtam ezen, de...  
Meg vagyok győződve arról, hogy...  
Megtudnád / Megtudná mondani, hogy...  
Mi a véleményed / véleménye a ... -ról/-  
ről  
Mit gondolsz / gondol... -ról / -ről  
Szeretném megtudni, hogy...  
Szerintem...  
Válasznál / Válaszolna nekem néhány  
kérdésre?

## PROJEKTHEZ TARTOZÓ HASZNOS KIFEJEZÉSEK

állatkertekben ritkán látható  
állatkerti séta  
az állatok csodálatos világát  
az év minden napján  
egész napos programot nyújt  
egzotikus fajok  
fajtamentő munka  
fedezd fel  
felejthetetlen élmény  
fokozottan védett állatok  
látványteretéseik  
Magyarországon honos  
találkozz  
természetes élőhelyek  
varázslatos környezetben  
várja látogatóit  
világraszóló  
világszenzáció

### 1. Engedély kérése

- Megengedi, hogy **leüljek**?
- Van lehetőség arra, hogy később **viSSzaadjam**?
- Megoldható, hogy azonnal **elinduljunk**?

### 7. Engedély adása

- Megengedem, hogy **kölcSönvegyétek**.
- Anyukám hagyja, hogy egyedül **csináljam**.
- A kutyám túri, hogy **megsimogassam**.
- Nincs ellene kifogása, hogy **átfessük** a szobát.

### 8. Akadályozás

- A rossz időjárás megakadályozza, hogy **viSSzaautazzunk**.
- Lehetetlenné teszi, hogy **eláruljam** az igazságot.
- Nem segít, hogy **megoldjam** a problémát.
- A szüleim nem adnak alkalmat arra, hogy **megmagyarázzuk** a történeteket.

### 9. Ellenzés

- Ellenzem, hogy új függönyöket **vegyünk**.
- Szó sem lehet róla, hogy **megváltoztassák** a szabályokat.

### 5. Tiltás

- Apa megtiltotta, hogy későn **érjek** haza.
- A menhely önkéntese tiltakozik az ellen, hogy **megváltoztassák** a szabályokat.

### 6. Nyelvtani kapcsolóelemek (a szövegkohézió eszközei)

- ezért
- így
- tehát
- egyszóval
- az előbbieK alapján látható, hogy
- mindezek után világos, hogy
  
- mert
- mivel
- amiatt, hogy
  
- figyelembe véve, hogy

## Melléklet H: Példa a Magyar felfedezőhöz készített Tanári (A1) kézikönyvből



NÉZD EZT A SOK ÁLLATOT!

### 4. FEJEZET 1.ÓRA

#### AZ ÓRA ÁTTEKINTÉSE

**Idő:** 45'-60'

**Cél:**

- egy állat bemutatása (kérdés-válasz)
- számok, színek és kérdőszók ismétlése, új szókincs megismerése

**Fókusz:** szókincs és hallott szövegértés

**Eszközök:** (K.A1.4.1) állatos kártyák (1 adag nagyban, páronként/ csoportonként 1 adag kicsiben), hanganyag, üres adatlapok

#### RÁHANGOLÓ FELADAT

**Cél:** a tanulók korábbi ismereteinek előhívása

- egyenként mutassuk fel a kártyákat,
- a tanulók próbálják kitalálni az állatok neveit
- hangoztassuk közösen az állatneveket (H.A1.4.1)
- csoportonként/ páronként adjunk egy adag szó- és képkártyát
- párosítsák a szó- és képkártyákat
- ellenőrizzük a megoldásokat

**Vegyes csoportok:** A többi nyelvi szint kártyáival együtt vegyes csoportokban is megoldhatjuk ezt a feladatot. Figyeljünk arra, hogy a páros / csoportos résznél az A1-es feladatlapnak megfelelő kártyákat adjuk oda.

#### SZÓKINCS

Az első három gyakorlat célja a régi és új szavak ismeretének elmélyítése, írásképük és hangzásuk gyakoroltatása.

**1.f.** Kössék össze az állatok képeit az elnevezésükkel. Ellenőrizzük le a megoldásokat.

**2.f.** Fejtsék meg a keresztrejtvényt a megadott segítséggel. Ellenőrizzük a megoldásokat.

**Megoldások:** 1.majom, 2.oroszlán, 3.tigris, 4.zebra, 5.pingvin

**3.f.** Keressék meg a hat elrejtett állatot a szókeresőben. Ellenőrizzük a megoldásokat.

**Megoldások:** medve, elefánt, pingvin, kenguru, zsiráf, ló

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| U | P | M | F | H | B | U | N | R |
| K | R | T | E | W | O | T | I | D |
| M | K | U | V | D | N | V | V | F |
| V | R | I | G | Á | V | L | G | L |
| K | R | I | F | N | Z | E | N | Ó |
| F | W | E | P | S | E | W | I | W |
| Y | L | J | F | L | Y | K | P | L |
| E | Z | S | I | R | Á | F | I | P |

## HALLOTT SZÖVEGÉRTÉS

### Hallgatás előtt:

4.f. A gyakorlat célja a megértés megkönnyítése, a kérdőszók és a hozzájuk kapcsolódó válaszok ismétlése. A tanulók egyénileg vagy párokban is megoldhatják a gyakorlatot. **(H.A1.4.2)**

### Megoldások:

- **Hány kg/m/cm: 70, 200, 50, 3000;**
- **Milyen színű: zöld, kék, fekete;**
- **Mit eszik: húst, növényt, halat, gyümölcsöt.**

5.f. A cél, hogy a tanulók megismerkedjenek a hallgatás során kitöltendő adatlapokkal, illetve, hogy az előző feladatban ismételt kérdéseket, válaszokat szóban is használják.

### Hallgatás közben:

6.f. Hallgassuk meg legalább kétszer a hanganyagot **(H.A1.4.3)**, a tanulók egyénileg töltsék ki az adatlapokat. Tanári kérdésekkel, kisebb csoport esetén egyéni beszámolóval ellenőrizzük a megoldásokat.

### Megoldások:

- **elefánt: Afrikában, szürke, 6000 kg, 3 m, növényeket;**
- **kenguru: Ausztráliában, barna, 90 kg, 70 cm, növényeket;**
- **pingvin: Déli-sarkon, fekete-fehér, 2 kg, 50 cm, halat;**
- **tigris: Ázsiában, sárga-fekete, 100 kg, 120 cm, húst.**

### Hallgatás után:

7.f. Egészítsék ki a szöveget az adatlapok segítségével. Ellenőrizzük a hanganyaggal.

### LEVEZETŐ FELADAT

**Cél:** az órán elhangzott új szavak ismétlése

- Írjunk fel egy mintapárbeszédet:

*Melyik állatot (nem) szereted? / (Nem) Szeretem az elefántot.*

- Álljanak fel a tanulók és kérdezzék egymást.
- Amikor mindenki mindenkivel beszélt, számoljanak be nekünk. Véletlenszerűen kérdezzük őket.

### KIS PROJEKT 1. – kártya készítése

A kis projektek fontos részei és előkészítői a fejezethez tartozó fő projektnek, ezért fontos, hogy mindenképpen kerítsünk rá sort. Ha az órán nem jut rá idő, adjuk fel házi feladatnak, kezdjük ezzel a következő óránkat.

**Eszközök:** kinyomtatott üres kártyák

#### A feladat:

- Készítsenek egy hasonló kártyát a kedvenc állatukról **(K.A1.4.2)**, amin minden fontos információ megtalálható.
- Amint elkészültek, barkochbázhatunk velük párokban vagy közösen.

**Cél:** a kérdések, válaszok gyakorlása és az órán tanult szavak ismétlése.

A kártyák később az állatkertünk állatairól készült adatlapok alapanyagai lesznek.



Melléklet I: Példa a Magyar felfedezőhöz készített kiegészített feladatlapokból (B1)



NÉZD EZT A SOK ÁLLATOT!



4. fejezet / B1.1

|   |   |   |
|---|---|---|
|    |    |     |
| <b>akvárium</b>   | <b>állathordozó</b>   | <b>fekhely</b>  |
|    |    |    |
| <b>fésű</b>   | <b>játék(ok)</b>  | <b>kalitka</b>  |
|  |  |  |
| <b>kaparófa</b>   | <b>ketrec</b>   | <b>kutyatáp</b>   |
|  |  |  |
| <b>labda</b>  | <b>macskaalom</b>   | <b>macskabútor</b>  |
|  |  |   |
| <b>madáreledel</b>  | <b>nyakörv</b>  | <b>tál</b>  |

Pécs

2023