

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 tézisei

Témavezető: Dr. Pelcz Katalin

Pécs

2023

1. 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 témáikat 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.

2. 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 nívóvumnak 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.

3. A kutatás célja

A doktori értekezésben bemutatott korpuszalapú kutatás során magyar anyanyelvű gyermekek természetes nyelvhasználatának megfigyelésére és vizsgálatára törekszem 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.

4. Elméleti keret

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 használat alapú nyelvelméleti modellek (vö. Tomasello, 2003, 2013, 2015; Bybee, 2008; Ortega et al., 2016; Ellis et al., 2016; Pelcz, 2023), amelyek főként a gyermekek nyelvelsajátítási folyamatával foglalkoznak, de lényeges megállapításaikat a nyelvtanulás során is érdemes figyelembe vennünk. A használat 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). Korpusznyelvészeti kutatások bebizonyították, hogy a kompetens nyelvhasználók beszédprodukciónak 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 fluensséghoz.

A negyedik alappillér a beszéd folyékonysághoz kapcsolódó elméletek. Ilyenek a Segalowitz-féle hármas modell, a legfontosabb beszédprodukciónak modellek (Levelt, 1999; Kormos, 2006), Tavakoli és Hunter piramis modelleje (Tavakoli, 2020), valamint az automatizálás fogalmának értelmezése.

Végül a kutatás ötödik alappillére a korpusznyelvészeti megközelítés (vö. Szirmai,

2005; Stefanowitsch, 2020; Hunston, 2022), 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.

5. Kutatási kérdések

A doktori értekezésben bemutatott korpuszalapú kutatás (vö. Szirmai, 2005; Stefanowitsch, 2020; Hunston, 2022) tehát a KorSzak Gyermekeyelvi 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).

Ennek a célnak az eléréséhez a következő kutatási kérdésekre keressük a választ:

1. Melyek a leggyakoribb egyelemű lexémák a KorSzak Gyermekeyelvi Korpuszban?
2. Milyen a leggyakoribb egyelemű lexémák szófaji (főnév, melléknév, ige) eloszlása a KorSzak Gyermekeyelvi Korpuszban?
3. Melyek a leggyakoribb két- vagy többelemű lexikai egységek a KorSzak Gyermekeyelvi Korpuszban?
4. Hogyan használják a KorSzak Gyermekeyelvi 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 többemű lexikai egységeknek?

A KorSzak Gyermeknyelvi Korpuszon végzett kutatás tehát a fenti hat kutatási kérdésre fókuszál. 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, 2019: 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. A kutatás több szekvenciából tevődött össze, amelyek alapvetően két meghatározott pontból indultak ki: az egyelemű lexémák és a két- vagy több elemű lexikai egységek vizsgálatából.

1. ábra A kutatás kvantitatív és kvalitatív vizsgálatok szekvenciái



6. A jelen kutatásban meghatározott feltételrendszer

6.1 Gyakoriság

Biber (2006) vizsgálatát, Dóla (2016, 2018) 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 (Vermekei, 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 (Vermekei, 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

6.2 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.

6.3 Funkció és jelentéstartalom

A gyakoriság alapján kiválasztott lexikai egységeket a Biber által meghatározott, szövegben betöltött szerepük szerinti kategóriákba kerülnek besorolásra.

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

Főkategória	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,	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,	<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>

Főkategória	Funkció	Alkategória	Biber példái	A KorSzak Gyermeknyelvi Korpusz példái
	szándékokat, jóslatokat, képességeket	előfeszítés, képesség, kötelezettség)		
Diskurzus-szervező kifejezések [discourse organising bundles]	a téma bevezetése, a témára való fókuszálás elérése, valamint a téma pontosítása és kidolgozása	témabevezetés	<i>want to talk about, going to talk about, if you look at</i>	<i>az lenne a kérdésem, akkor ott van a</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>

6.4 Feldolgozás és produkció

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 beszédprodukciónál 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 (Vermeki, 2023b, megjelenés alatt).

7. Eredmények

7.1 Melyek a leggyakoribb egyelemű lexémák a KorSzak Gyermeeknyelvi Korpuszban?

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 korpuszsal – amely a KorSzak Gyermeeknyelvi Korpuszban található, a kritériumok alapján kiválasztott lexikai egységekből épült fel – összevetve azt is észrevehettük, hogy ezek mindegyike a fókuszkorpusz többelemű 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 többelemű 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 többelemű 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.

7.2 Milyen a leggyakoribb egyelemű lexémák szófaji (főnév, melléknév, ige) eloszlása a KorSzak Gyermeknyelvi Korpuszban?

A KorSzak Gyermeknyelvi Korpuszban összesen 55.143 szó található, amelyből 17.180 főnév, 10.259 ige és 933 melléknév.

A főnevek tekintetében 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.

A melléknevek gyakorisági listáiban található a legnagyobb különbség a fókusz és a referenciakorpusz között. A melléknevek közül a KorSzak Gyermeknyelvi Korpusz leggyakrabban megjelenő lemmája a JÓ, amely a referenciakorpuszban a harmadik helyen szerepel. Az adatközlő gyermekek által használt melléknevek nagy része nem szerepel a huTenTen12 első száz lemmája között. Ilyenek az

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

Ezenkívül a huTenTen12 korpuszban szereplő elemek közül a fókuszkorpuszban lévők eltérő, legtöbbször sokkal alacsonyabb gyakorisági mutatóval (gyakoriság per egymillió szó) rendelkeznek. Ilyenek például az ÚJ vagy az UTOLSÓ lemmák, amelyek a szövegekben szignifikánsan ritkábban fordulnak elő. Az ÚJ lemmára a KorSzak Gyermeknyelvi Korpuszban csaknem 80%-kal kevesebb előfordulást találtam, amíg az UTOLSÓ esetében 60%-kal kevesebb előfordulást jegyeztem fel.

Az igék kategóriájában 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. 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. Ilyen a már az előző fejezetben említett MEGY (72%-kal magasabb), a CSINÁL (81%-kal magasabb), valamint az EMLÉKSZIK (78%-kal magasabb) igék.

7.3 Melyek a leggyakoribb két- vagy többelemű lexikai egységek a KorSzak Gyermeknyelvi Korpuszban?

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, amelyekből 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 veled, vagy mit tudom én, az volt az első, az nagyon jó volt, de ő is nagyon*

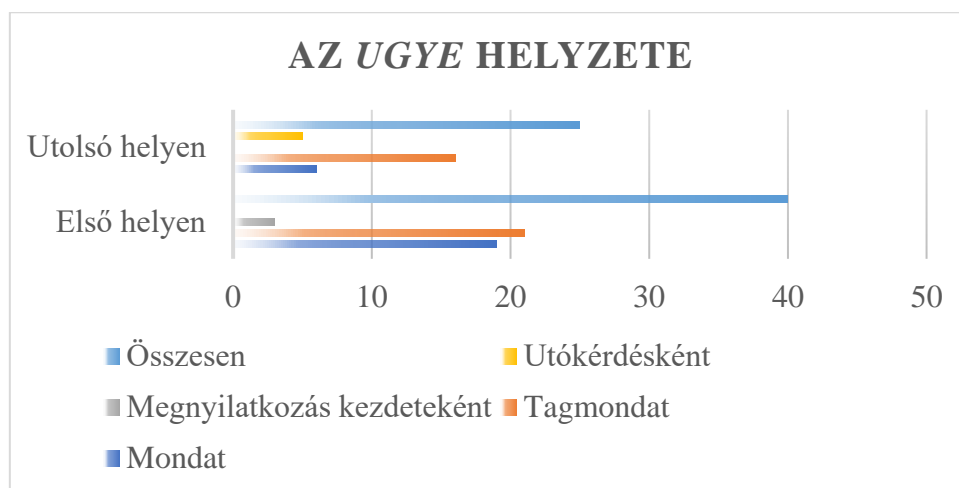
7.4 Hogyan használják a KorSzak Gyermeknyelvi Korpuszban szereplő gyermekek a kiválasztott lexikai egységeket? Milyen használati mintázatok figyelhetők meg a korpuszban?

7.4.1 Az ugye használati mintázatai

Az *ugye* a korpuszban négyféle funkcióval jelenik meg. Az első, amikor a gyermekek a szót időkitöltő elemként használják gondolkodás közben. A következő funkció a megerősítés,

amikor a beszélő biztos abban, hogy a hallgató egyetért vele, valamilyen közös háttérismeretet vagy megértést feltételez. A harmadik funkciója a megerősítés kérése, amely során a beszélő visszaigazolást kér arra vonatkozóan, hogy az általa említett kijelentés valóban megfelel a valóságnak. Az *ugye* 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 mondat- vagy tagmondatkezdő pozícióban 40-szer, mondat, illetve tagmondat végén összesen 25-ször szerepel. Ez utóbbi közül utókérdésként mindösszesen 5 alkalommal használják a gyermekek. 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. A többi kérdésben főleg időkitöltő szerepe van, ezenkívül egyszer jelenik meg megerősítés kérése funkcióban.

2. ábra Az UGYE pozíciója a megnyilatkozásokban



7.4.2 A NEM TUD használati mintázatai

A leggyakoribb kételemű lexikai egység a NEM TUD lemmákból származik, amely összesen 292-szer jelenik meg a KorSzak Gyermekegyelvi Korpuszban. 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. 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 található néhány példa erre. Ezenkívül 5 esetben megjelenik a *nem tudom* + főnévi igenév szerkezet is, amely a látható.

3. ábra Találatok a *nem tudom* + kérdőszó mintázatra a KorSzak Gyermekegyelvi Korpuszban

Left context	KWIC	Right context ↑
és akkor hű, de menő, hogy ott állt, jött az állat a	nem tudom N.Nom V.Prs.Def.1Sg	hány négyzetméteres kennelebe, és akkor de menő, hogy levadászta, amikor n
vesen gitározónék, és akkor még ott van a másik,	nem tudom Adv V.Prs.Def.1Sg	hány elfoglaltság, vagy éppen hegedűt kell gyakorolni, zongorát kell gyakoroln
s>Ő biztos, hogy ott van, mert nem férne el az a	nem tudom N.Nom V.Prs.Def.1Sg	hány tonna régiség, ami még a szülei nagyszülei volt, és azokat úgy őrizget
m.</s><s>így őszintén megmondom, hogy most	nem tudom Adv V.Prs.Def.1Sg	hol van, de biztos, hogy megvan még valahol. - És Bubu hogyan teljesített?</s><s>
: YouTube-on, hogy bement valami kínai erdőbe,	nem tudom Adv V.Prs.Def.1Sg	hol , és ott azt hiszem talált egy halotat és kiröhögte.</s><s>Pontosabban nei
sen van megformálva.</s><s>Halálfaló.</s><s>	Nem tudom Adv V.Prs.Def.1Sg	milyen részben van először, de egy a Voldemort leghűségesebb szolgája, és e
szeretem a meglepetéseket, mert idegesít, hogy	nem tudom Adv V.Prs.Def.1Sg	mi az, de persze, amikor megkapom, örülök neki, és van, amikor jó érzés meg
semelyik. - Hát én inkább kutya.</s><s>Igazából	nem tudom Adv V.Prs.Def.1Sg	miért , de valahogy én a kutyákat talán jobban kedvelem.</s><s>Nem azt mor
neg a macskát, pedig egyébként nem is.</s><s>	Nem tudom Adv V.Prs.Def.1Sg	miért , de nem tudom.</s><s>Múltkor jöttem rá, hogy igazából annyira nem is
tt.</s><s>Hát valamiért a szüeim elleneztek, de	nem tudom Adv V.Prs.Def.1Sg	miért . - Nem tudom, én nem úgy talán azért félek háziállatot tartani, mert ha n
csótányok, illetve a poloskáktól, így undorodom,	nem tudom Adv V.Prs.Def.1Sg	miért így, nem azoktól is, mint a csótányoktól. - Én a kigyóktól félek, mert ha v:
</s><s>Tessék. - Nekem a Grindelwald.</s><s>	Nem tudom Adv V.Prs.Def.1Sg	miért , de a Grindelwald. - De miért? - Nekem azért tetszik a Grindelwald karal

4. ábra A *nem tudom* + Inf. mintázat megjelenése a KorSzak Gyermekegyelvi Korpuszban

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

Details

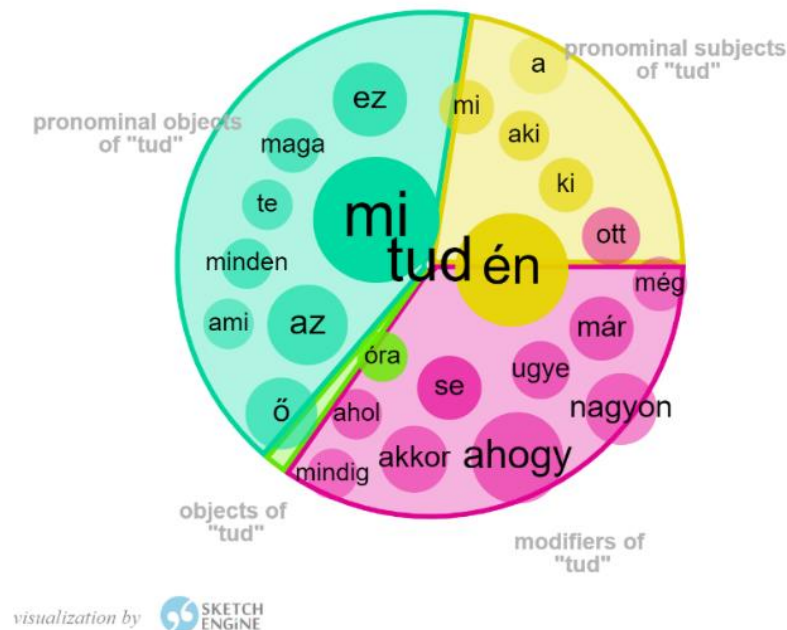
	Left context	KWIC	Right context ↑
1	doc#65 tom, én egy éve nézem körülbelül, és már hatszor végignéztém az egészet és	nem tudom Adv V.Prs.Def.1Sg	abbahagyni .</s><s>Nekem ez a kedvenc sorozatom. - Nekem nem nagyon v
2	doc#66 </s><s>Hát szerintem azért, mert úgy hogyha elkezdünk egy sorozatot, akkor	nem tudom Adv V.Prs.Def.1Sg	abbahagyni , 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	nem tudom Adv V.Prs.Def.1Sg	eldönteni , 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	nem tudom Adv V.Prs.Def.1Sg	elképzelni , 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	nem tudom Adv V.Prs.Def.1Sg	kikefélni , 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

A NEM TUD találatait érdemes összevetni az egyelemű TUD találati eredményeivel is, hogy további következtetéseket vonhassunk le a használatával kapcsolatban. 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

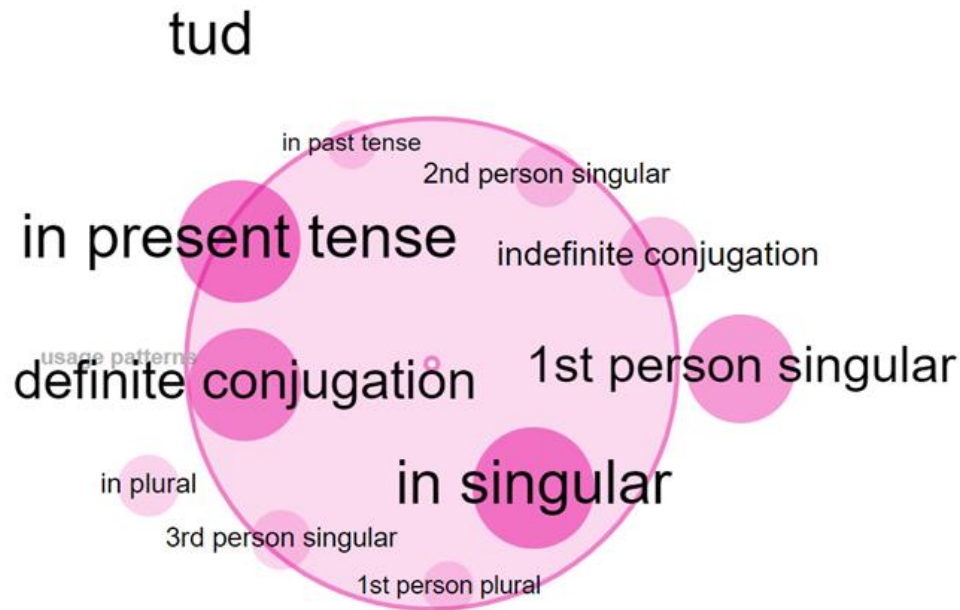
alkalommal *tudom* alakban jelenik meg, és ahogy fentebb láthattuk, ebből 227-szer, tehát a találatok 65%-ban *nem tudom* szerkezetben. Az alábbi ábra, amelyen a TUD kollokációit láthatjuk, szintén azt mutatja, hogy a TUD lemma legszorosabb kapcsolatban az alanyként megjelenő egyes szám első személyel (*én*), valamint a tárgyként megjelenő MI lemmával áll (Vermekei, 2023a).

5. ábra A TUD lemma kollokációi (Vermekei, 2023a: 120)



A TUD lemma kollokációt tekintve 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 (Vermekei, 2023a).

6. ábra A TUD lemma kollokációi (Vermekei, 2023a: 120)



A KorSzak Gyermeknyelvi Korpuszban a *nem tudom* lexikai egységnek négy funkciója jelenik meg. Amellett, hogy a *nem tudom* használata egy módja lehet a tudásunk hiányának kifejezésének, használhatjuk kérdések elkerülésére, időkitöltő elemként vagy arra, hogy tudassuk a beszélgetőtársunkkal, hogy hallottuk és értjük, amiről beszél.

7.4.3 A MI és TUD használati mintázatai

A Korszak Gyermeknyelvi 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 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.

7. ábra A *mit tudom én* háromelemű lexikai egység leggyakoribb megjelenéseinek konkordanciasora

Left context	KWIC ↑	Right context
nvalóan a haszonállatokat azokat munkára tartják, mit	tudom V.Pr:Def:1Sg	én , igás lovat szállításra használnak, és. - Vagy h
gy mennyi ideje van vele foglalkozni, mármint egy mit	tudom V.Pr:Def:1Sg	én , egy hörcsög nem feltétlenül igényli, hogy fogl
z állnak a ketrecen, és sosem látják a napot, vagy mit	tudom V.Pr:Def:1Sg	én , Magdi néni udvarán kaparászik a csirke, és b
most egy jobb gazdája is, ha nem hoztam el volna mit	tudom V.Pr:Def:1Sg	én , a menhelyről, vagy bármi, és igazából én talá
ít tartani.</s><s>Ilyen kis pici kutyát.</s><s>Ilyen, mit	tudom V.Pr:Def:1Sg	én pomerániai törpe spicc, az még úgy elmegey tal
állatok ezrei, vagy direkt nem szeretik az állatokat, mit	tudom V.Pr:Def:1Sg	én , most tudok egy ilyen példát, hogy nem mond
kor már elegend van belőle, akkor egyszerűen csak mit	tudom V.Pr:Def:1Sg	én , kiviszed az országútra, és otthagyt, de el se
az oroszlán levadássza a leggyengébb gnút, vagy mit	tudom V.Pr:Def:1Sg	én , zebrát akkor, nem arra törekszik, hogy az az :
<s>Igen. - Meg ez is, hogy mindenki mondja, hogy mit	tudom V.Pr:Def:1Sg	én , az emberek mennyivel intelligensebbek az áll
és tudod, ez a mennek a reptéren, és akkor jön a mit	tudom V.Pr:Def:1Sg	én , a vakvezető kutya, éppen ő is megy föl a gép

A 46 találatból 38-szor, vagyis az előfordulások 82%-ában szerepel ebben az egységben. 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: elsősorban egyes szám elsőben (40 előfordulás), többes szám elsőben (4 előfordulás) és végül egyes szám második személyben (2 előfordulás).

2. táblázat A MI + TUD együttes megjelenéseinek használati mintázatai a KorSzak Gyermeknyelvi Korpuszban

Lexikai egység	Találatok száma	Változatok gyakorisága (%)
MI + TUD	46	
MI + TUD + Def: 1Sg	38	82
mit tudom én	29	63
hogy mit tudom én	5	10,86
vagy mit tudom én	4	8,69
MI + TUD + NDef: 1Sg (+ Adv/Pro) + Inf	2	4,34
hogy mit tudok belőle kihozni	1	2,17
mit tudok még erre mondani	1	2,17

Lexikai egység	Találatok száma	Változatok gyakorisága (%)
MIT + TUD + NDef: 1Pl (+ Adv) + Inf	4	8,69
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
MIT + TUD + NDef: 1Sg + Inf	2	4,34
nem tudsz mit kezdeni	1	2,17
nem tudsz mit mondani	1	2,17

Az összefoglaló táblázatból 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, illetve időkitöltő funkcióban használják a korpuszban szereplő gyermekek.

7.4.4 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 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 a korpuszban.

3. táblázat Az *annyira nem* kételemű lexikai egység előfordulása a KorSzak Gyermeeknyelvi Korpuszban

Lexikai egység	Találatok száma	Változatok gyakorisága (%)
ANNYIRA NEM	39	
annyira nem	39	100

4. táblázat Az ANNYIRA és a NEM együttes előfordulásai a KorSzak Gyermeknyelvi Korpuszban

Sz.	Szó	Együttes előfordulások (-5 /+5)	Összes együttes előfordulás a korpuszban	LogDice
1.	nem	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 a KorSzak Gyermeknyelvi 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, amely esetén pont az ellenkezője történik annak, ami az első funkcionál. Ezzel a funkcióval a korpuszban lévő megjelenéseinek 59%-ában jelenik meg. Az *annyira nem* harmadik funkciója az összehasonlítás, amely összesen egy alkalommal jelenik meg a korpuszban.

7.4.5 Az ILYEN és az 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, 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.

5. táblázat Az ILYEN és az OLYAN lemma megjelenése a KorSzak Gyermeknyelvi Korpuszban

Sz.	Lemma	Gyak.	Gyakoriság / millió	DOKgy	Relatív DOKgy %
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 többemű lexikai egység részét képezi, amelyet a korpuszból elemzésre kiválasztottunk. Ezen kívül számos alkalommal jelenik meg melléknév és főnév előtt. Az ILYEN után álló melléknevek 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 *hobbi* (3 előfordulás).

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. 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.

6. 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
hobby	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		

7. táblázat Az ILYEN és az OLYAN mellett megjelenő leggyakoribb melléknevek, határozószók, névmások, kötőszók

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	
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 (Vermeki, 2023b, megjelenés alatt). Ugyanakkor a korpuszban szereplő gyermekek néhányszor pontatlanság indikálására is alkalmazzák a kifejezést. 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 (Vermeki, 2023b, megjelenés alatt).

(doc#07 Mon_Állatok_Kutya_004: A gyermekek kutyás emlékeikről beszélgetnek.)

(1) A: Ő eléggé hírhedt lett ott, és ő egy olyan kutya, 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** kutya, 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 Gyermeeknyelvi Korpusz, 2020)

(doc#53 Mon_Influenzerek_001: A gyermekek influenzerekről beszélgetnek. Az egyik beszélő elmesélt egy negatív történetet az egyikükről. Erre reagál a másik.)

(2) A: Ha **van egy olyan** influenzeser, akiről nem vagy jó véleménnyel, meg hülyeségeket posztol, akkor miért követed?

(KorSzak Gyermeeknyelvi Korpusz, 2020)

7.4.6 Az igazából használati mintázatai

Az *igazából* lemma összesen 170 alkalommal jelenik meg a korpuszban, amelyből 6 előfordulása a HÁT IGAZÁBÓL ÉN háromelemű lexikai egységben 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. 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 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.

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).

8. ábra Példák az *igazából* + *nem* használatára

Left context	KWIC	Right context
:ettő.	igazából nem	tudom miért, de valahogy én a kutyákat talán jobban kedvelem.
:szünkbe jut. - Igen. -És Zsófi néni is mondta, hogy ő is olyat írt fel, hogy utána	igazából nem	is annyira szereti, csak az jutott eszébe.
<Olyan, mintha látná a jövőt. - Hát már nem sokáig.	igazából nem	szeretnék ilyen lenni, meg mondjuk egy szűnyog is hiába él mondjuk ve
mondjuk egy ilyen tudományos, hogy mondjuk, tehát az ilyen rosszindulatú, ami	igazából nem	jó semmire, az mondjuk nem, de ami így nem árt az állatnak sem.
jes mértékben, tehát volt esélye arra, hogy túléljen. - Ez tök jó.	igazából nem	volt madarunk, csak egyszer, de anyuék azzal vigasztaltak, hogy elrept
g minden, már fű alig evett, meg alig ivott, az nagyon szörnyű volt.	igazából nem	is tudtuk végül, hogy mi lett a baja.
s>Úgy bent. - Hát nálam egy megint tapasztalatból mondott sztori.	igazából nem	teljesen szilveszterkor történt, de egyszer elkezdtek tűzijátékozni nálun

Amikor az *igazából* a háromelemű lexikai egység, vagyis a HÁT IGAZÁBÓL ÉN részét képezi, 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.

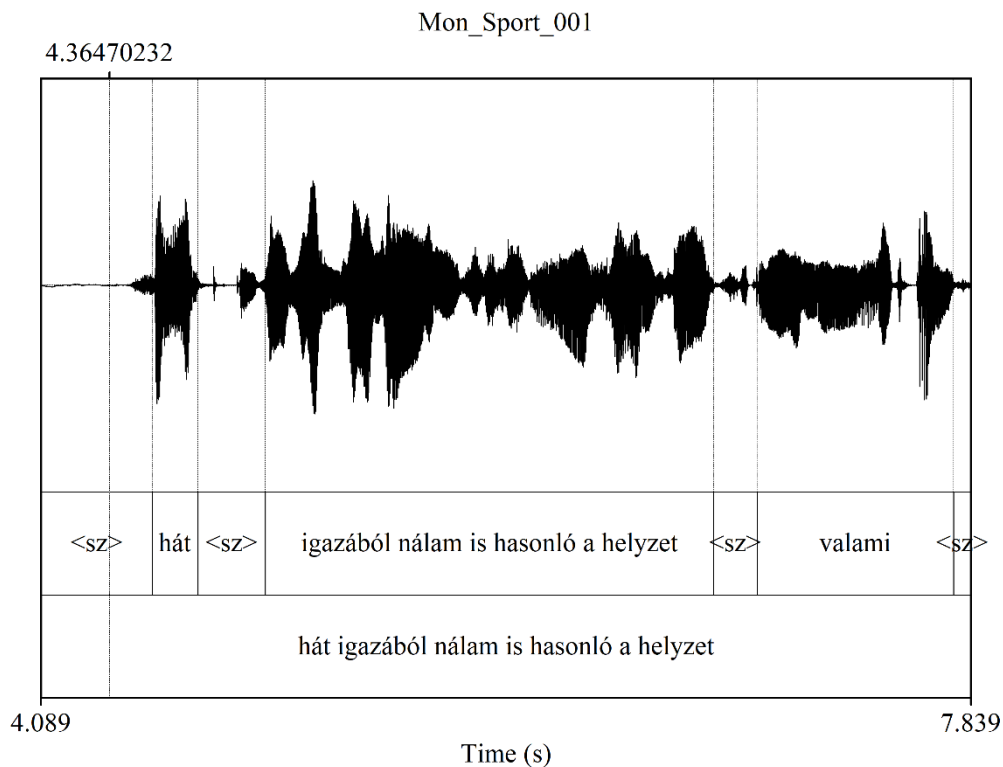
9. ábra A HÁT IGAZÁBÓL ÉN konkordanciasora

Left context	KWIC	Right context
; mert félek, hogy nem jól gondoskodik róla. -	Hát igazából én	pont attól félek, hogy nem tudok róla gondoskodni,
t, hogy szegényke még előbb meghalt volna. -	Hát igazából nekem	a hörcsögömmel az volt, hogy én így az elején mo
lis háziállat? - Szerintem teljesen ideális, mert	hát igazából én	úgy tudom, hogy a kutyák érzik, hogyha bármi bajc
jére. <Mesélj a kedvenc mozifilmedről. -	Hát igazából nekem	a kedvencem az talán a Star Wars, amit még kiset
advem se. <Én ezt a témát elhagytam. -	Hát igazából én	is egyszer kezdtem bele a Riverdale-be, de aztán :
oardoznak, és én ezt szívesen kipróbálnám. -	Hát igazából nálam	is hasonló a helyzet. <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 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*.

10. ábra Szünetek helye a *hát igazából nálam is hasonló a helyzet* kimondása közben (doc#68: Mon_Sport_001)



7.5 Hogyan ágyazódnak be a két- vagy többemű lexikai egységek hosszabb szekvenciákba? A beágyazódott lexikai egységek milyen használati mintázatot mutatnak?

7.5.1 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. Hasonló funkcióban, tehát az egyetértés kifejezésekként hosszabb lexikai egységek részét is képezi.

8. táblázat Az én is beágyazódásai hosszabb lexikai egységekbe a KorSzak GyermeKNyelvi Korpuszban

Lexikai egység	Gyak.
én is így vagyok ezzel/vele	8
• ezzel én is így vagyok	(2)
• ezzel én is nagyon így vagyok	(1)
én is szoktam	6
ezzel én is egyetérttek	3
• ezzel én is teljesen egyetérttek	(1)
én is nagyon szeretek + Inf.	2
ezzel én is egyet tudok érteni	1
én is nagyon szeretem	1
meg én is ezen a véleményen vagyok	1

A megnyilatkozásokban lévő pozícióját tekintve, a fenti táblázatban 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.

7.5.2 Az *így vagyok* és az *úgy vagyok* beágyazódása

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 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 *hogy* kötőszóval jelenik meg a korpuszban.

9. táblázat Az *így vagyok* és az *úgy vagyok* kollokációi a KorSzak Gyermeknyelvi Korpuszban

Sz.	így vagyok			úgy vagyok				
	Szó	Együttes előfordulások (-5 /+5)	Összes együttes előford.	LogDice	Szó	Együttes előfordulások (-5 /+5)	Összes együttes előford.	LogDice
1.	én	10	744	8,76	hogy	9	1791	7,36
2.	is	9	1218	7,91	én	6	744	8,03
3.	ezzel	6	36	12,06	vele	4	74	10,62
4.	vele	4	74	10,61				
5.	és	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, míg az utóbbi használatával a beszélő a saját álláspontját fejezi ki.

A megnyilatkozások közben tartott szünetek helyét megvizsgálva láthatjuk, hogy míg az *én is így vagyok ezzel/vele* esetében a 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.

10. táblázat Szünetek helye az *én úgy vagyok (vele)* kimondása közben

Dok.	N-gram (szó)	Szünet helye a lexikai egység 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:

Dok.	N-gram (szó)	Szünet helye a lexikai egység előtt / után van		Megjegyzés
		Igen	Nem	
				<sz> én úgy vagyok veled, hogy rá <sz>
#40	én úgy vagyok veled	x		Többelemű egység, a szünet mintája: <sz> mert én úgy vagyok veled, hogy én<sz>
#49	én úgy vagyok		x	A szünet mintája eltér: <sz> én úgy vagyok a játékoknál, hogy most újabban elkezdtem podcastet <sz>
#49	úgy vagyok	x		Többelemű egység, a szünet mintája: <sz> aztán én úgy vagyok veled <sz>
#51	én úgy vagyok veled	x		Jól mérhető a szünet.
#54	én úgy vagyok veled	x		Jól mérhető a szünet.
#66	úgy vagyok		x	A szünet mintája eltér: <sz> és <sz> most úgy <sz> vagyok, 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 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.

11. 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 %
hogya	2	20	hogy	9	100

én is így vagyok ezzel / vele			én úgy vagyok (vele)		
Kötőszó	Előfordulás	Az összes előfordulás %	Kötőszó	Előfordulás	Az összes előfordulás %
mint	2	20			
de	1	10			
és	1	10			
hogy	1	10			

7.5.3 A mit tudom én beágyazódása

A lexikai egység összesen 29 előfordulással szerepel a KorSzak Gyermekegyelvi 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.	vagy	8	516	8,91
2.	most	4	295	8,66
3.	egy	6	764	7,95
4.	ilyen	3	541	7,43
5.	de	4	769	7,36
6.	akkor	3	591	7,31
7.	hogy	8	1791	7,17
8.	az	7	1566	7,17
9.	nem	5	1412	6,83
10.	meg	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 *hogymit tudom én* előfordulásainak 80%-ában, a *vagy mit tudom én* előfordulásainak 75%-ában, illetve mind a *meg mit tudom én*, mind az *és mit tudom* esetében a szünetek tartására vagy közvetlenül a lexikai egység előtt és után kerül sor, vagy pedig egy hosszabb egység után.

12. táblázat Szünetek helye a *hogymit tudom én* kimondása közben

Dok.	N-gram (szó)	Szünet helye a lexikai egység előtt / után van		Megjegyzés
		Igen	Nem	
#2	hogymit tudom én		x	A szünet mintája eltér: <sz> vagy <sz> mit tudom én
#2	hogymit tudom én	x		Jól mérhető a szünet.
#39	hogymit tudom én	x		Jól mérhető a szünet.
#42	hogymit tudom én	x		Jól mérhető a szünet.
#53	hogymit tudom én	x		Jól mérhető a szünet.

13. táblázat Szünetek helye a *vagy mit tudom én* kimondása közben

Dok.	N-gram (szó)	Szünet helye a lexikai egység 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: <sz> vagy <sz> mit tudom én
#9	vagy mit tudom én	x		Többelemű egység, a szünet mintája: <sz> vagy mit tudom én milyen <sz>
#59	vagy mit tudom én	x		Többelemű egység, a szünet mintája: <sz> vagy mit tudom én mire <sz>

14. táblázat Szünetek helye a *mit tudom én* kifejezést tartalmazó egyéb lexikai egységekben kimondása közben

Dok.	N-gram (szó)	Szünet helye a lexikai egység 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.

7.6 Mi a szövegben betöltött szerepe a KorSzak Gyermeknyelvi Korpuszban lévő két- és többelemű lexikai egységeknek?

15. táblázat A lexikai egységek szövegben betöltött szerepük szerinti kategóriáinak összefoglaló táblázata (Biber, 2006: 139–146. alapján)

	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	11	56	13	91	19	12	3	28	4	816
5.	összesen	67		123			47				
6.	Gyakoriság	323	443	203	637	146	126	12	284	30	12.748

8 A kutatás eredményeinek felhasználása

A 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. A *Magyar felfedező* (Kasikci – Vermeki, 2021) 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 szerzők¹ 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 fejezet². A további fejezetek előkészületei pedig már zajlanak (Vermeki, 2021b).

A tananyagok szerzői a munkájuk során számos területen alkalmazhatják a korpuszokat. A következőkben azt mutatom be röviden, hogy ez hogyan történik a *Magyar felfedező* esetében.

8.1 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). 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

¹ A *Magyar felfedező* című tananyag szerzői Kasikci Katalin és Vermeki Boglárka

² magyarfelfedezo.hu

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.

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 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.*

A2-es nyelvi szinten újra visszatérnek ezek a leírást segítő lexikai egységek és ki is egészülnek újabb elemekkel. 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.

8.2 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 4. fejezethez készült részét láthatjuk az alábbi táblázatban röviden összefoglalva.

16. táblázat A Magyar felfedező 4. fejezetének lexikogrammatikai tanmenetének részlete

Nyelvi szint	Téma	Nyelvi forma
A1	Kedvenc állatok	Főnevek többes száma; létige: van/ vannak, nincs/ nincsenek; 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; 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)

8.3 A témához kapcsolódó lexika, válaszvariációk és szövegtípus kapcsolatának beépítése a tananyagba

A modellalapú nyelvoktatás módszere jól kapcsolódik a korpuszinformált tananyagoknak a felhasználásához, illetve a *Magyar felfedező* által is képviselt nyelvoktatási szemlélethez. A módszer arra a megfigyelésre épül, 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 többemű 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 többemű 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 fóka. 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. 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 beszéd-készség fejlesztését megcélzó feladatokban is hasonlóan alkalmazzuk ahhoz, ahogy ez a *MagyarOK* tankönyveiben történik. 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 ezt jól mutatja.

11. 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 bevegyétek-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>

12. ábra Vélemény kifejtése, információ kérése, válaszadás késleltetése (C1)

6. Képzeljétek el, hogy titeket kért fel az iskolaújság egy állatokról szóló interjúsorozat 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">• <i>Válaszolnál/Válaszolna néhány kérdésre?</i>• <i>Mit gondolsz/gondol ...ról/ről?</i>• <i>Szeretném megtudni, hogy ...</i>• <i>Arról érdeklődnék, hogy ...</i>• <i>Meg tudnád/tudná mondani, hogy ...?</i>• <i>Mi a véleményed/véleménye a/az ...ról/ről?</i>• <i>Azon gondolkodtam, hogy ...</i>	<ul style="list-style-type: none">• <i>Hát ...</i>• <i>Hadd gondolkozzak ...</i>• <i>Ez nehéz kérdés ...</i>• <i>Ez jó kérdés ...</i>• <i>Hát, attól függ, hogy ...</i>• <i>Hogy érted/érti, hogy ...</i>• <i>Még sosem gondolkodtam ezen, de ...</i>	<ul style="list-style-type: none">• <i>Szerintem ...</i>• <i>Én úgy gondolom, hogy ...</i>• <i>Biztos vagyok benne, hogy ...</i>• <i>Azt gondolom, hogy ...</i>• <i>(Nekem) az a véleményem, hogy ...</i>• <i>Azt hiszem, hogy ...</i>• <i>Meg vagyok győződve arról, hogy ...</i>

8.4 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. A szituációk kiválasztásánál is elsősorban a KorSzak Gyermekeyelvi 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.

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

A tananyag beszédképesség-fejlesztésre fókuszáló részeiben a KorSzak Gyermekeyelvi 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 többemű 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 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 kifejezéséeként. Például:

- Hát **nem tudom**, hogy a néptánc mennyire sport, de nekem a néptánc a hobbim (1.fejezet, A1).

A szövegekre visszatérve fent említésre került, hogy a tananyag beszéd-készség-fejlesztő részeibe a KorSzak Gyermekeknélvi 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, az alábbi példákön.

(doc#7 Mon_Állatok_Kutya_004: A gyermekek a kedvenc kutyafajtáikról beszélgetnek)

- (3) A: Szóval hogyha kezdhelném 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.**

13. 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 bichon, 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 következő ábrán.

14. 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 gyerekszempontból 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 jégtáblákon 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 Gyermekegyelvi Korpusz beszélt nyelvi szövegeire támaszkodunk, hiszen egy ilyen jellegű tananyag célja az is, hogy a magyarul tanuló gyermekeknek a tanulmányaik során használható tudományos szóincse, szóbeli és írásbeli kifejezőmódja is fejlődjön. Ennek a célnak az elérése érdekében használt forrásokról talán egy tanulmányban érdemes lenne majd beszámolni, most azonban erre nem térnék ki részletesebben.

9. Lezárás és kitekintés

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.

A kutatást a KorSzak Gyermeeknyelvi 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 Gyermeeknyelvi Korpuszhoz hasonló gyermekek számára készült pedagógiai korpuszok, amelyeket az értekezésben nyelvoktatásra készült autentikus szövegek halmazaként értelmezünk, 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 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 gyermeeknyelvi 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 Gyermekegyelvi 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 alkörpusz 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.

Referencia

- 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.
- Biber D. (2006). *University language: a corpus-based study of spoken and written registers*. J. Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- 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.
- CorpusMate (2023). Web: <https://corpusmate.baisa.cz/search>
- 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 idegennyelv-tanulásban–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.
- Ellis, N. C., O’Donnell, M. B., Romer, U. (2016). Constructions and Usage-based Approaches to Language Acquisition. *Language Learning*, 66, 23–44.
- Halliday M. (1978). *Language as Social Semiotic: The social interpretation of Language and*
- 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.
- Hunston, S. (2022). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum Associates Publishers.
- KorSzak Gyermekegyetemes Korpusz (2020-2022. október). ske.li/vfe
- 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.,

- 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.
- McCarten, J. (2010). Corpus-informed course book design. In: O’Keeffe, A., McCarthy, M. (Eds.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge. 413–427.
- 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.
- 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 (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.
- 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.
- Pieratt, J. (2019). *CraftED*, web: <https://craftedcurriculum.com/what-is-pbl/> (Utoljára megtekintve:2021.10.09.).
- SACODEYL (2004-2008). <http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp>
- Schmidt, I. (2022). Kétnyelvű gyermekek beszéde mint az élőnyelv egy változata. 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, 144-153.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Siyanova-Chanturia, A., Pellicer-Sanchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. New York, NY: Routledge.
- 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.).

- 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., Pelcz K. (2017). Modellalapú nyelvoktatás–Természetes nyelvhasználat a tanteremben és a tanterem 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 (Utoljára megtekintve: 2023.04.12.).
- Tavakoli P., Wright C. (2020). *Second language speech fluency: from research to practice*. Cambridge: Cambridge University Press.
- 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.
- 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.

Vermeki Boglárka az értekezés témakörében releváns publikációi és konferencia-előadásai

PUBLIKÁCIÓK

- 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 Gyermekegyetemi Korpuszon*. Nyelv és megismerés – Nyelvészeti kutatások határon innen és túl konferencia tanulmányai, Hungarológiai Közlemények. Újvidék, Szerbia
- 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*. САВРЕМЕНИ ТОКОВИ У ИЗУЧАВАЊУ ЈЕЗИКА, КЊИЖЕВНОСТИ И КУЛТУРИ tanulmánykötet, Belgrádi Egyetem, Szerbia.
- Vermeki B. (2022c, megjelenés alatt) Gyakori többszavas lexikai egységek gyermekekbeszédprodukción, Hungarológiai Évkönyv, 23.
- 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.
- Kašikci K., Vermeki B. (2021). Magyar felfedező. Nemzetstratégiai Kutatóintézet. (mintafejezetek egy gyerekeknek készülő magyar nyelvkönyvből) Web: <https://magyarfelfedezo.hu/>
- Baumann T.– Majoros J. Judit – Pelcz K. – Schmidt I. – Szita Sz. – Vermeki B. (2020). Bemutatkozik a Korpusznyelvészeti és Szakmódszertani Munkacsoport. Hungarológiai Évkönyv 21 (1-2), 32–41.

KONFERENCIA-ELŐADÁSOK

2023

Lexico-grammatical patterns of children's language use. 1st Symposium of Language Education and Research, University of Greenwich, London, UK

2022

A gyermekek spontán beszédének jellemzői - szógyakorisági és kulcsszó vizsgálatok a KorSzak Gyermeknyelvi Korpuszon. Nyelv és megismerés – Nyelvészeti kutatások határon innen és túl, Újvidéki Egyetem, Szerbia.

Lexical Priming in Children's Spoken Language. Cotemporary Philological Studies, Belgrádi Egyetem, Szerbia.

A lexikai jóslás jelentősége a nyelvoktatásban. Aktuális kihívások a szak/nyelvoktatásban – A módszertani megújulás lehetőségei, Corvinus Egyetem, Budapest

Formulaic language in child spontaneous speech productions, САРЕМЕНИ ТОКОВИ У ИЗУЧАВАЊУ ЈЕЗИКА, КЊИЖЕВНОСТИ И КУЛТУ, Belgrádi Egyetem, Szerbia.

Fixed and Semi-Fixed Lexical Chunks in a Children's Spoken Language Corpus. Annual Conference of the Canadian Association of Applied Linguistics. Western University and Université d'Ottawa, Canada (online).

2021

A gyermekek élőnyelvi nyelvhasználatában megfigyelhető lexikai egységek felhasználása a magyar mint idegen nyelv tanítása során Felelős nyelvészet – Alkalmazott Nyelvészeti Konferencia. ELTE, Budapest.

Élőnyelvi korpusz-informált tananyagok a MID oktatásában. 21. Élőnyelvi Konferencia, Beregszász.

Using a Child Spontaneous Spoken Language Corpus for Informing Hungarian as a Foreign Language Teaching Material. 3rd International Conference on Language Education and Research, Athens, Greece.

Magyar felfedező – egy készülő korpuszinformált tananyag gyerekeknek. MANYE, Veszprém.

Pécs

2023