

Vermeki Boglárka

Gyakori többszavas lexikai egységek gyermek beszédprodukcióiban

1. Bevezetés

Jelen tanulmány célja a *KorSzak Gyermeknyelvi Korpusz* gyakori többszavas lexikai egységeinek vizsgálata és beszédszándék szerinti kategorizálása. Korpusznyelvészeti kutatások igazolták, hogy az anyanyelvi beszélők beszédprodukciójuk során nagymértékben támaszkodnak a **formulaszerű nyelvhasználatra** (O'Keefe és mtsai. 2007), és korlátozottan élnek a grammatika által elméletileg biztosított korlátlan variációs lehetőségekkel (Pawley–Syder 1983). A spontán beszéd feldolgozásához kötődő kognitív igények miatt a szóbeli megnyilvánulások „korlátozott és ismétlődő repertoárral” rendelkeznek (Biber és mtsai. 1999: 1049), és formulaszerű, *előre összeépített (pre-constructed)* egységekre támaszkodnak (Wray 2002). A nyelv „olyan darabokból áll, amelyek egymással kombinálva folyamatos koherens szöveget hoznak létre” (Lewis 1997: 7, saját ford.).

Ezt bizonyítja például Altenberg (1998) kutatása, aki a *London-Lund* beszélt nyelvi angol korpusz kutatása révén arra a következtetésre jutott, hogy a rutinná vált, részben vagy teljesen előregyártott (*prefabricated*) struktúrák a nyelvi szerveződés mindegyik szintjére kihatnak, az egyes megnyilatkozásoktól a teljes diskurzusig. A többé-kevésbé konvencionizált kifejezések a teljesen lexikalizált egységek és a szabad konstrukciók határán helyezkednek el. Altenberg becslése szerint a korpuszban szereplő szavak több mint nyolcvan százaléka ilyen egység része.

Különféle, a lexikai szerveződés szintjén álló többelemű egységeket a beszélők kisgyermekkoruktól kezdve tanulnak és használnak. A tömböknek vagy daraboknak (*chunks*) (Lewis 1997), más kifejezéssel többszavas szekvenciáknak (*multiword sequence*) (Götz 2013) nevezett egységek beszéd közben nem igényelnek nagy kognitív erőfeszítést: a formulaszerű elemsoportokat először holisztikusan memorizálják a gyermekek, és csak később kezdik el elemezni őket (Tomasello 2003: 305–307).

2. A formulaszerű szekvenciák jellemzői

Alison Wray (2019: 267) meghatározásában formulaszerű szekvencia minden olyan többmorfémás egység (morféma- vagy szólánc), amelyet az ágens (például a nyelvtanuló) **egyetlen lexikai egységként azonosíthatónak vagy hasznosíthatónak érzékel**. A jelenség leírására számos kifejezés található a szakirodalomban; a magyar terminológia kapcsán lásd Dóla (pl. 2006) írásait. Wray (2002) könyvében több mint ötven olyan fogalmat gyűjt össze, amely meghatározza a formulaszerű nyelvhasználat különböző aspektusait. Az ötven kifejezés némelyike a jelentést (pl. *idiom, unit of meaning*) vagy a formát (pl. *sentence stem*), míg mások a rögzítettséget (pl. *multiword unit*) hangsúlyozzák (Hunston 2022: 102). A fogalmakban azonban közös, hogy szavak olyan csoportosulásaira utalnak, amelyek gyakran fordulnak elő együtt, egymás társaságában (vö. Firth, J. R. 1957, Sinclair 1991, Hoey 2005).

Az együttesen előforduló (*co-occurring*) elemláncokra – a mentális folyamatoktól függetlenül – az angol szakirodalomban főként a *lexikai tömb (lexical chunk)*, a *lexikai köteg (lexical bundle)* és a *többszavas kifejezés (multiword expression)* terminusokkal utalnak. Siyanova-Chanturia és Pellicer-Sanchez (2019: 2) szerint a legjobb a *formulaszerű nyelv* vagy

nyelvhasználat kifejezést használni, mert az egyszavas egységeket is magában foglal. Beletartoznak például a felkiáltások (pl. *Hurrá!*) vagy különféle beszédformulák (pl. *ja*). Ugyanakkor, ha például a többszavas kifejezések fogalmát használnánk, akkor a fent említettek kikerülnének a definícióból, pedig működésük hasonló a többszavas egységekéhez.

A formulaszerű szekvenciák fő jellemzője, hogy feltehetően holisztikusan tárolódnak és hívódnak elő a memóriából, és ezzel csökkentik a kognitív terhelést (Wray 2002). Emellett fontos karakterizáló tényező lehet még a *gyakoriság* (frequency), az *ismertség* (familiarity), a *kiszámíthatóság* (predictability), a *rögzültség* (fixedness) és a *pragmatikai funkció* (pragmatic function) is (Siyanova-Chanturia és Pellicer-Sanchez 2019: 3). A **gyakoriságot** tartják az egyik legmeghatározóbb tulajdonságnak, bár egyes kutatási eredmények alacsony korrelációt mutatnak a gyakori formulaszerű szekvenciák és a vizsgált csoport által megtanult szekvenciák között (vö. Garnier–Schmitt 2016; González és mtsai. 2015). Siyanováék szerint ennél sokkal meghatározóbb a beszélő közösség általi **ismertség**, konvenció, valamint a **kiszámíthatóság**. Ez utóbbira példaként az olvasás folyamatát említik: amikor az olvasó meglát egy ismerős szót, megjósolja a soron következő szavakat, és akár ugrik is a szövegben, ezáltal felgyorsítja az olvasási sebességét (Siyanova-Chanturia és Pellicer-Sanchez 2019: 4). **Rögzültségük** alapján ezek az előregyártott formulaszerű szekvenciák lehetnek teljesen vagy részben rögzültek, ami azt jelenti, hogy a bizonyos elemei variálhatóak. A következő példákban erre a két rögzültségi változatra láthatunk példákat (forrás: *KorSzak Gyermeknyelvi Korpusz* 2020):

- (1) öszintén szólva
- (2) jó emlékem van a FŐNÉV+-val/-vel kapcsolatban

Az (1)-es példa egy rögzült formulaszerű szekvencia, míg a (2)-es csak részben rögzült (l. üres hely). Nyelvtanulói szempontból fontos megjegyezni, hogy amellett, hogy természetesebbé tesz a beszédünket, az ilyen részben rögzült formulaszerű szekvenciák nagy kommunikációs értékkel bírnak, hiszen „bejártódtak” bizonyos jelentések kifejezésére, ugyanakkor rugalmasak is.

Egy másik kulcsfontosságú tulajdonsága a formulaszerű szekvenciáknak kommunikációs, illetve **pragmatikai funkciójuk**. Nézzük például az alábbi példákat (forrás *uo.*):

- (3) Ja, értem.
- (4) Én is egyetérték az XY-val/-vel.
- (5) a) Nekem az lenne az első kérdésem, hogy...
b) Nekem az első kérdésem az lenne, hogy...

Vannak olyan kifejezések, amelyeket például az interakciók során arra használunk, hogy fenntartsuk a kommunikációt (3), kifejezzük az egyetértésünket (4), vagy a kommunikációs céljainknak megfelelően szervezzük a diskurzusunkat (5a–b) (Siyanova-Chanturia és Pellicer-Sanchez 2019: 5). Emellett, ahogy korábban említettem, a formulaszerű nyelvhasználat hozzájárul a beszédprodukciónak hatékonyságához és természetességéhez.

3. A formulaszerű nyelvhasználat nyelvoktatásban betöltött szerepe

Ahogy korábban említettem, mind a gyermek, mind a felnőtt anyanyelvi beszélők jelentős számú többszavas lexikai egységet ismernek és használnak; a formulaszerű nyelvhasználat lényeges része a kommunikációnak. Ugyanakkor a nyelvtanulók számára is fontos és hasznos lehet a különféle formulaszerű szekvenciák tanulása és használata. A kutatások szerint hozzájárulnak a kommunikációs kompetenciák fejlődéséhez, a **beszédfolyamatosság** növeléséhez (Wray–Fitzpatrick 2008) és az adekvát nyelvhasználathoz (Kirk 2014: 105), illetve megfelelő használatuk elősegíti a beszéd anyanyelvihez hasonló **természetességét** és az üzenetfeldolgo-

zási terhelés csökkentését; használatuk révén könnyebbé és hatékonyabbá válik a kommunikáció (Segalowitz 2010: 126). Azok a nyelvtanulók, akik nem megfelelően és nem formulaszerűen használják a többszavas lexikai egységeket, nem jutnak hozzá a fenti előnyökhöz: a nagyobb **beszédfeldolgozási teher** lassíthatja beszédfolyamatosságukat. Akinek hiányos a kollokációs kompetenciája, hosszabb, terjengősebb kifejezéseket alkalmaz, és eközben többet hibázhat, továbbá az anyanyelvi beszélő számára gyakran nehezebben érthető módon formálhatja meg mondandóját (Hill 2000). Ezen a gondolaton alapul az ún. *lexikai megközelítés* ('lexical approach') is (Lewis 1993), noha a panelek tanítása korábbra nyúlik vissza: már az audiolingvális módszerrel megjelent a nyelvoktatásban (Tavakoli 2020: 32), sőt, Palmer még ennél is korábban, 1925-ben szorgalmazta, hogy az angol társalgási nyelv fejlesztésének alapvető irányelve legyen a gyakori, hasznos „szócsoportok” memorizálása (1925, 1999: 185).

A többszavas egységek a nyelvtanulók receptív készségeinek használata során is hasznosak lehetnek, hiszen a gyakran ismétlődő elemsorok felismerése lehetővé teszi a nyelvi input gyorsabb feldolgozását. Kimutatták, hogy a tanulók gyorsabban felismerik a szavakat, ha olyan más szavak társaságában hallják vagy olvassák őket, amelyekkel tipikusan meg szoktak jelenni (Sivanova-Chanturia és mtsai. 2011; Pellicer-Sánchez és mtsai. 2022). Összegezve, a formulaszerű szekvenciák a nyelvfeldolgozás, a nyelvi produkció és a megértés szempontjából is fontosak.

4. A korpuszról

A gyermekek beszédprodukciónak megvalósuló formulaszerű nyelvhasználatot a **KorSzak Gyermeknyelvi Korpuszban** vizsgáltam meg. Ez a korpusz a Korpusznyelvészeti és Szakmódszertani Munkacsoport (rövidítve KorSzak) által épített korpusz részét képezi. A munkacsoport 2020 februárjában jött létre azzal a céllal, hogy „hidat építsen a korpusznyelvészeti kutatások és a magyar mint idegen nyelv szakmódszertana között” (Baumann és mtsai. 2021: 32). Ennek a célnak az elérése érdekében a tagok almunkacsoportok létrehozásával egyszerre több korpusz építését kezdték meg (Baumann és mtsai. 2021: 32–33); jelenleg élőnyelvi, tanulói és gyermeknyelvi korpusz épül.

A KorSzak Gyermeknyelvi almunkacsoport célja olyan pedagógiai célú korpusz építése, 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 és mtsai. 2021: 33). A *KorSzak Gyermeknyelvi Korpusz* egy dinamikus pedagógiai célú korpusz, amely jelenleg – 2022 októberében – hetvennyolc felvételre épül, melyek huszonkilenc 11–15 éves gyermek (9 fiú és 20 lány) beszélgetéseiről készültek.

A korpusz dinamikus, hiszen folyamatosan növekszik, pedagógiai céllal készült, vagyis a szövegeket előre meghatározott témák jegyzik. Jelenlegi állapotában több mint 71 000 tokent tartalmaz; 68 párbeszédből és 10 monológból áll. A monológok a 2021-es évben először a járvány okozta karantén idején és okán keletkeztek, majd később a *személyleírás* témakör bevezetése miatt további felvételekkel gyarapodott a korpusz ezen része. A dialógusok során 2–5 adatközlő beszélget egymással egy-egy témáról; többnyire ez két vagy három gyermeket jelent. Az adatközlő gyermekek magyar vidéki iskolákban tanuló osztálytársak és barátok. A legtöbben egynyelvű magyarok, akik az általános iskola felső tagozatán angolt vagy németet tanulnak idegen nyelvként. A videó- és hangfelvételek során párokban vagy kiscsoportokban szabadon beszélgetnek egy-egy előre meghatározott témáról. A feladat időtartama nem volt előre megszabva, hogy a gyermekek ne érezzék feszélyezve magukat. Emiatt a videó- és hanganyagok között találhatóak 1–2 perces és ennél jóval hosszabb, akár 45 perces felvételek is. A hanganyagok átírása az Alrite nevű beszédfelismerő programmal történt. A szövegek lek-

torálás után felkerültek a Sketch Engine-re, egy korpusznyelvi eszközök tartalmazó online felületre, amely több mint 90 nyelven elérhető.

A korpuszépítés a témák meghatározásával kezdődött. Ehhez 11–14 éves gyerekek körében készítettem egy felmérést arról, hogy milyen témákról szeretnek beszélni, mik a kedvenc szabadidős tevékenységeik, kikkel töltik szívesen a szabadidejüket. A felmérés során a résztvevő 138 általános iskolásnak párban vagy kiscsoportban kellett egy gondolattérképet kitöltenie, majd később néhány kiegészítendő kérdésekre kellett válaszolniuk. Ezek után a témákat életkor és gyakoriság szerint szétválogattam, majd kiválasztottunk közülük 12-t: állatok/kutyák, kirándulások, Harry Potter, hobbik/kreatív hobbik, sport, influenzaszerek, tévésorozatok, találmányok, technikai újítások, utazás, saját történetek, ételek). Minden témához készítettünk egy-egy rövid kérdéseket tartalmazó gondolattérképet, hogy a gyermekeknek esetlegesen segítséget nyújtsunk a beszélgetéshez: például ha elakadnának a felvétel során, a kérdések új ötleteket adhatnak a folytatáshoz (Baumann és mtsai. 2021: 33-35). A beszélgetések rögzítése 2020 tavaszán kezdődött, és azóta is folyamatosan tart.

5. A jelen kutatás módszertanáról

A formulaszerűsége esélyes többszavas lexikai egységek azonosításához alkalmazhatók korpusznyelvészeti eszközök, de a pusztán gyakorisági alapon megtalált elemsorok nyelvi (l. grammatikai, szemantikai) és kognitív (l. tárolás, előhívás) egységstátusza nem garantált; ahhoz többféle eszközt kell használni. A legjobb eredmények korpuszeszközök használatával úgy érhetőek el, ha kongram (vagyis változó végű Ngram) keresést végzünk (vagy manuálisan kikeressük őket a konkordanciasorokból; Hunston 2022: 102) további kritériumok meghatározása mellett. Biber (2006) például egyetemi tananyagokban és tanórai beszédben keresett pusztán statisztikai alapon előálló ún. lexikai kötegeket ('lexical bundles'), és kutatásához csak az egymillió szavas korpuszban legalább negyvenszer előforduló négyszavas többeseket használta fel (133–134). Mivel a jelen kutatáshoz használt korpusz nem írott nyelvi, hanem spontánbeszéd-korpusz, és a magyar nyelv az izoláló angoltól eltérően agglutináló, más kritériumok meghatározása szükséges. Vizsgálatomban ezért – az első vizsgálatok alkalmával – a kétszavas többeseket is felsorolom a gyakorisági listákban, továbbá a korpusz mérete miatt az első öt megjelenést vizsgálom – még akkor is, ha azok között van is kisebb gyakorisági mutatóval rendelkező lexikai egység, mint a Biber (2006) által meghatározott (Vermeki 2022).

A többszavas egységeket a gyakoriságuk alapján gyűjtöm ki a Sketch Engine N-grams keresőeszközzel, amely token sorozatokból generál gyakorisági listákat. A leggyakoribb elemsorokon konkordancia-kereséseket is végzek, a magyar nyelv szórendi sajátosságai miatt ugyanis előfordulhatnak olyan szekvenciák, amelyeket az eszköz csak a második kereséssel talál meg, vagy külön csoportba sorol (Vermeki 2022). Íme egy példa erre a problémára (forrás: *Korszak Gyermeknyelvi Korpusz* 2020):

- (6) a) **ezzel** én is így vagyok
 b) én is így vagyok **ezzel**
 c) én is így vagyok **vele**

A (6) alatti példamondatok jelentése megegyezik, a gyerekek egyetértésüket fejezték ki egy adott témával kapcsolatban. A keresőeszköz az előfordulásokat három különálló N-gramként kezelte az első keresés során, az *ezzel* és a *vele* névmás helyzete miatt. Az *ezzel* megjelenhet az *én is így vagyok* előtt vagy után, míg a *vele* csak utána állhat. Ezért elengedhetetlen a keresési eredmények részletes vizsgálata és egy második kulcsszó-kontextusban (KWIC) keresés által létrehozott konkordanciasor elemzése (Vermeki 2022) (1. ábra).

Details	Left context	KWIC	Right context
1 doc#2 on embertelen, vagy hogy mondjam. - Igen, ezzel én is így vagyok , hogyha már egyszer valamin ki kell próbálni, akk			
2 doc#8 net kitalálni, mert nem ismered a személyiséget. - Én is így vagyok vele.</s><s>És milyen színű golden retriever lenn			
3 doc#53 gymond bárkivel, akivel jól érzem magam, igen. - Én is így vagyok ezzel, mint az Amira, hogy én is bárkivel kimegyek			
4 doc#53 yünk, akkor csak egyedül is szeretek sétálgatni. - Én is így vagyok ezzel a dologgal, hogyha családdal vagyunk, akkc			
5 doc#60 iket se nézem, én inkább az angol sorozatokat. - Én is így vagyok vele, mint az Eszter.</s><s>Igazából nem nagyon			
6 doc#61 m tudom abbahagyni, mivel izgalmas. - Hű ezzel én is így vagyok .</s><s>Én függővé válok és ez nagyon durva. - I			
7 doc#68 n természeti látnivalók inkább jobban vonzanak. - Én is így vagyok vele, de például, amikor mentünk Firenzébe, akko			

1. ábra: Konkordanciasor egy egyetértő kifejezés elemzésére (Vermeki 2022)

6. Az eredmények kategorizálásáról

A *KorSzak Gyermeknyelvi Korpuszban* történő gyakorisági vizsgálatok után a kapott eredményeket beszédszándékok szerint kategorizáltam (vö. Vermeki 2022). Jelen tanulmányban nem kívánom áttekinteni az ide kapcsolódó pragmatikai szakirodalmat, csak röviden ismertetem, mit értek beszédszándék alatt.

A kommunikáció során a beszélők nemcsak „beszélnek” (pl. kijelentéseket tesznek), hanem cselekszenek is: megnyilatkozásaikkal tetteket hajtanak végre (pl. ígérenek, bocsánatot kérnek, utasítanak) – ezeket klasszikusan beszédteknek nevezzük (Austin 1962), de a nyelvoktatásban elterjedtebb a *beszédszándék* szó használata (Yule 1996). Jelen kutatásomban a leggyakoribb szótöbbségeket a következő beszédszándékok alá soroltam: *egyetértés, egyet-nem-értés, véleménykérés, nyomatékosítás, pontosítás, kétség és bizonytalanság* kifejezése.

7. Kutatási eredmények

Jelen kutatás során összesen 1434 gyakori szótöbbséget azonosítottam be (az 5. fejezetben megjelölt módszerrel). A találatok hossza 2–6 szó volt; a magyar nyelv sajátosságai miatt a legtöbb találat 2 vagy 3 szó terjedelmű volt.

Sorszám	Szavak száma	Találatok száma
1	2	861
2	3	73
3	4	5
4	5	3
5	6	3

1. táblázat: Többszavas lexikai egységek találatainak száma (Vermeki 2022)

A legmagasabb arányban kétséget, bizonytalanságot kifejező lexikai egységeket találtam. A változatok száma kevés, mégis ezek fordultak elő a legnagyobb számban a korpuszban. A jelenség nemcsak a gyermekek beszédére jellemző, hanem magára a spontán beszédre is, hiszen az írott kommunikációval ellentétben a spontán beszéd „valós idejű, rögtönzött, a hallgató és a beszélő rövid távú memóriájának kapacitása befolyásolja” (Andó 2003: 98), emiatt sok időnyerésre, gondolat- és beszédtervezésre alkalmas elemet tartalmaz. Ezek jelentős része

olyan enyhítő kifejezés, amely a bizonytalansággal van kapcsolatban – mint például a gyakorisági mutatók alapján első helyen szereplő *nem is tudom*, illetve *nem tudom* (2. táblázat).

Sorszám	Lexikai egység	Találatok száma	Gyakoriság (találatok száma / egymillió token)
1	nem (is) tudom	275	3914.98
2	hogya (is) mondjam	31	441.32
3	mit tudom én	29	412.85
4	vagy nem (is) tudom	27	284.38
5	vagy hogyan mondjam	18	256.25

2. táblázat: Többszavas lexikai egységek a kétség, bizonytalanság kifejezésére

A korpusz anyagának gyűjtése során használt feladat jellegéből adódóan a második legnagyobb kategória az egyetértés és egyet-nem-értés lett. A beszélgetések videó- és hangfelvételei során ugyanis a gyermekeknek különböző előre kiválasztott témákban kellett elmondaniuk a véleményüket, majd egymás véleményére reagáltak is (3. táblázat). A táblázatban látható, hogy a korpuszban leggyakrabban az *ez/az igaz* és a *hát én/nekem is* kifejezések szerepelnek leggyakrabban az egyetértés kifejezésére.

Sorszám	Lexikai egység	Találatok száma	Gyakoriság (találatok száma / egymillió token)
1	ez / az igaz	23	327.43
2	hát én/nekem is	14	142.36
3	én is egyetértek	8	113.89
4	(ezzel) én is így vagyok ezzel / vele	7	99.65
5	igen, (az/ azok) abszolút	7	99.65
6	(én is) egyet tudok érteni	6	85.42
7	én is inkább	6	85.42
8	nekem is hasonló a véleményem / a helyzet / mint...-nél	4	42.71

3. táblázat: Az egyetértés kifejezésére használt többszavas lexikai egységek (ld. Vermekei 2022)

Az egyetértéshez kapcsolódó beszédzándék az egyet-nem-értés (4. táblázat). Ezzel kapcsolatban fontos megjegyezni, hogy az adatközlő gyermekek barátok és osztálytársak, következésképpen beszélgetéseik során általában egyetértenek egymással, mivel bizonyos témákban hasonlóan gondolkodnak.

Sorszám	Lexikai egység	Találatok száma	Gyakoriság (találatok száma / egymillió token)
1	(így) annyira nem	39	555.22
2	de szerintem ez/az	10	142.36
3	én / nekem nem nagyon	7	99.65
4	de szerintem nem	3	42.71
5	hát nekem nem	3	42.71

4. táblázat: Az egyet-nem-értés kifejezésére szolgáló többszavas lexikai egységek

Amennyiben valamiben nem értenek egyet, akkor is csak „puha” egyet-nem-értést kifejező stratégiákat alkalmaznak. A kutatás ezen korlátjával tisztában kell lennünk, ugyanakkor a körülmények és az adatközlők életkori sajátosságai más típusú felvételek készítését egyelőre nem teszik lehetővé (vö. Vermeki 2022). Mindezek mellett a 4. táblázatban látható, hogy a korpuszban az *annyira nem* (és az *így annyira nem*), valamint a *de szerintem* szótöbbsesek a leggyakrabban előforduló egyet-nem-értést kifejező formulák.

Az egyetértéshez és egyet-nem-értéshez mint véleménynyilvánításhoz kapcsolódik a következő beszédzándék: a vélemény kérése. Az adatközlő gyermekek ennek megvalósulásához nem sok féle szerkezetet alkalmaztak, de az alábbi szótöbbsesek többször is előfordultak (vö. Vermeki 2022). Látható, hogy az *és szerinted* a leggyakoribb szekvencia.

Sorszám	Lexikai egység	Találatok száma	Gyakoriság (találatok száma / egymillió token)
1	És szerinted...?	11	156.06
2	(Te) Mit gondolsz ...-ról/-ről	8	113.89
3	Neked mi a kedvenc...?	2	28.47
4	Neked mi a véleményed...?	2	28.47

5. táblázat: Vélemény kérése

Mondanivalójuk nyomtatékosítását sokféle szótöbbsessel fejezik ki a kutatásban résztvevő gyermekek; jellemzően részben rögzült elemsorokkal, melyeknek lexikailag kitöltött része nem alkot egy grammatikai frázist – pl. *az nem* (6. táblázat):

Sorszám	Lexikai egység	Találatok száma	Gyakoriság (találatok száma / egymillió token)
1	az nem + TÁRGY/ MELLÉKNÉV / BIRTOKOS (+BIRTOK)	41	583.69
2	(de még) akkor is	36	512.51
3	mennyire + MELLÉKNÉV	26	370.14
4	pont a(z) ...-ról/-ről (már) beszéltünk	5	71.18
5	és igenis	4	56.95
6	semmiképp se(m)	3	42.71
7	mindegyik ugyanolyan fontos / aranyos	2	28.47

6. táblázat: Nyomatékosításra használt többszavas lexikai egységek

Végül a *pontosítás* kategória is viszonylag nagy számú találatot hozott. A 7. táblázatban láthatók az itt használt leggyakoribb szekvenciák, élükön a *tehát hogy* szótöbbsessel.

Sorszám	Lexikai egység	Találatok száma	Gyakoriság (találatok száma / egymillió token)
1	tehát hogy	16	227.78
2	tehát nem	8	113.89
3	nem feltétlenül (hanem)	7	99.65
4	mármint hogy (nem)	5	71.18

7. táblázat Pontosítás kifejezésére használt többszavas lexikai egységek

8. Összegzés

A formulaszerű nyelvhasználatról számos kutatás kimutatta, milyen nagy arányban jelen van az anyanyelvi beszélők (gyermekek és felnőttek) beszédében, és mekkora szerepe van a nyelvfeldolgozásban és a kommunikációban. A formulaszerű szekvenciák megfelelő használatának a nyelvtanulásban is lehet előnye; ezek közül itt kettőt emelek ki:

- 1) Fejlesztheti a tanuló produktív készségeit (beszéd és írás), mert a holisztikusan tárolt egységek emlékezetből való felidézésével felszabaduló idő és energia más területeken használható. A nyelvtanulók beszéde folyékonyabbá válhat a formulaszerű szekvenciák használatával.
- 2) Fejlesztheti a tanuló receptív készségeit (olvasás és hallás), mivel a gyakran ismétlődő lexikai egységek felismerése lehetővé teszi a nyelvi input gyorsabb feldolgozását.

A jelen tanulmányban arra kerestem a választ, milyen formulaszerűnek látszó többszavas lexikai egységek fordulnak elő leggyakrabban a gyermekek spontán beszélgetéseiben. A *Kor-Szak Gyermeknyelvi Korpuszban* beazonosítottam a leggyakoribb szótöbbségeket, és megállapítottam, hogy ezek a következő beszédshándékokat valósítják meg: kétség és bizonytalanság, egyetértés és egyet-nem értés, nyomatékosítás, véleménykérés, valamint pontosítás kifejezése.

Irodalom

- Altenberg, Bengt 1998. On the phraseology of spoken English: the evidence of recurrent word-combinations. In: Cowie, A. P. (szerk.): *Phraseology: theory, analysis and application*. Oxford University Press, Oxford. 101–122.
- Andó Éva 2003. Beszélt nyelvi narratívumok szerkezeti összetevőinek és beszédtempójának összefüggése. In: Tóth Szergej (szerk.): *Nyelvek és kultúrák találkozása*. A XII. Magyar Alkalmazott Nyelvészeti Kongresszus kiadványai III. SZTE Juhász Gyula Tanárképző Főiskolai Kar, Szeged. 98–103.
- Austin, John L. 1962. *How to Do Things with Words*. Oxford University Press, Oxford.
- Baumann Tímea, Majoros Judit, Pelcz Katalin, Schmidt Ildikó, Szita Szilvia, Vermekei Boglárka 2020. Bemutatózik a Korpusznyelvészeti és Szakmódszertani Munkacsoport. *Hungarológiai Évkönyv* 21: 32–41.
- Biber, Douglas 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Benjamins, Amsterdam.
- Biber, Douglas – Johansson, Stig – Leech, Geoffrey – Conrad, Susan – Finegan, Edward 1999. *Longman Grammar of Spoken and Written English*. Longman, Harlow.
- Dóla Mónika 2006. Formulaszerű nyelvhasználat – az angol *formulaic language* terminus beillesztése a magyar nyelvtudományi terminológiába. In: Fóris Ágota – Pusztay János (szerk.): *Utak a terminológiához*. Berzsenyi Dániel Főiskola, Szombathely. 94–116.
- Götz, Sandra 2013. *Fluency in Native and Nonnative English Speech*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Hill, Jimmie 2000. Revising priorities: from grammatical failure to collocational success. In: Lewis, Michael – Conzett, Jane (szerk.): *Teaching Collocation: Further Developments in the Lexical Approach*. Language Teaching Publications, Hove. 47–67.
- Hunston, Susan 2022. *Corpora in Applied Linguistics*. Cambridge University Press, Cambridge.
- Kirk, Steven 2014. Addressing Spoken Fluency in the Classroom. In: Muller, Theron – Adamson, John – Brown Shigeo – Herder, Steven (szerk.): *Exploring EFL fluency in Asia*. Palgrave Macmillan, Basingstoke. 101–119.
- Lewis, Michael 1993. *The Lexical Approach. The state of ELT and a Way Forward*. Language Teaching Publications, Thomson/Heinle.
- Lewis, Michael 1997. *Implementing the Lexical Approach*. Language Teaching Publications, Hove.
- O’Keeffe Anne – McCarthy, Michael – Carter, Ronald 2007. *From corpus to classroom*. Cambridge University Press, Cambridge.
- Palmer, Harold E. 1925. Conversation: The fundamental guiding principle for the student of conversation, In: Smith, Richard C. 1999. (szerk.): *The Writings of Harold E. Palmer. An Overview*. Hon-no-Tomoshia, Tokio. <http://homepages.warwick.ac.uk/~elsdr/WritingsofH.E.Palmer.pdf> (2022.06.12.)

- Pawley, Andrew – Syder, Frances H. 1983. Two puzzles for linguistic theory: Native-like selection and native-like fluency. In: Richards, Jack C. – Schmidt, Richard W. (szerk.): *Language and Communication*. Longman, London. 191–226.
- Pellicer-Sánchez, Ana – Siyanova-Chanturia, Anna – Parente, Fabio 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.
- Segalowitz, Norman 2010. *Cognitive bases of second language fluency*. Routledge, New York.
- Siyanova-Chanturia, Anna – Pellicer-Sanchez, Ana 2019. *Understanding Formulaic Language. A Second Language Acquisition Perspective*. Routledge, New York.
- Siyanova-Chanturia, Anna – Conklin, Kathy – Schmitt, Norbert 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.
- Tomasello, Michael 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge MA.
- Vermeki Boglárka (2022). „Hát így annyira nem” – How children ask and express opinions in Hungarian. *Studia Finno-Ugrici* (megjelenés előtt).
- Wray, Alison 2002. *Formulaic Language and the Lexicon*. Cambridge University Press, Cambridge.
- Wray, Alison 2019. Concluding Question: Why don't Second Language Learners More Proactively Target Formulaic Sequences? In: Siyanova-Chanturia, Anna – Pellicer-Sanchez, Ana (szerk.): *Understanding Formulaic Language. A Second Language Acquisition Perspective*. Routledge, New York. 248–269.
- Wray, Alison – Fitzpatrick, Tess 2008. Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In: Meunier, Fanny – Granger, Sylviane (szerk.): *Phraseology in foreign language learning and teaching*. John Benjamins, Amsterdam. 123–148.
- Yule, George 1996. *Pragmatics*. Oxford University Press, Oxford.

Vermeki, Boglárka

Frequent multiword expressions in children's speech

The research presented in this study focuses on the frequency of the appearance of multi-word lexical units in children's spoken language. Several corpus-based studies have established that fluent language users use a great deal of more or less fixed prefabricated lexical units in most of their speech (cf. Altenberg 1998, O'Keefe és mtsai. 2007). Studies on first language acquisition show that children learn these formulaic elements holistically, without analysis (Tomasello 2003: 305–307). In the current research, the most common multi-word lexical units are identified in the KorSzak Children's Language Corpus, and then they are classified based on the speech acts they realize.