

## Szavak gyakoriságának szubjektív megítélése gyermekkorban

### 1. Bevezetés

A gyermeket körülvevő beszélőközeg által használt nyelvi jelek elraktározása az anyanyelv-elsajátítás során történik. Az agyunknak azt a részét, amelyben az adott nyelv különböző egységeire vonatkozó információkat (lexikális elemek, morfológiai és szintaktikai szabályok, fonológiai információk) tároljuk, mentális lexikonnak nevezzük (Gósy 2005). Ennek a rendszernek a működtetése teszi lehetővé a beszédalapú kommunikációt, hiszen ennek aktiválása mind a beszédprodukción, mind pedig a beszédfeldolgozás során nélkülözhetetlen. A mentális lexikon feltöltése az anyanyelv-elsajátítás kezdetén indul, és gyakorlatilag egész életünk során folytatódik. Ebből következik, hogy a mentális lexikonban tárolt információk, illetve a mentális lexikon felépítésének különböző aspektusai személyenként nagymértékben eltérnek (Gósy 2005). Többek között befolyásolja a gyermek közvetlen környezetéből érkező nyelvi input, a beszélő környezet nyelvfeljesztő hatása, a tágabb értelemben vett szociokulturális környezet, az intézményesített oktatás, majd később a szakma, a hobbi stb.

A mentális lexikon felépítésével és működésével számos magyar és nemzetközi kutatás foglalkozik (vö. Gósy 1998; Navracscics 2001; Aitchison 2003; Lengyel Zs. 2008; Kovács 2013; Bóna–Imre 2017), így számos meghatározása ismeretes. Hasonlították szótárhoz, madárkalitkához (Platón), kincses tárhoz (Ciceró), könyvtárhoz (Kant) (Gósy 2005), illetve metróhálózathoz (Aitchison 2003) is. A 21. század elejére tehető a matematikai hálózat kutatás kezdete, amely a komplex rendszerek újszerű megközelítését hozza. A hálózatelmélet egyik alapfeltevése, hogy ezek a komplex hálózatok (internet, sejtek stb.) hasonlítanak egymásra, és jellemzőik így matematikailag leírhatók. Bár a mentális lexikont számos szerző jellemzi valamiféle hálózatként, a nyelv (és azon belül a mentális lexikon) hálózatközpontú vizsgálata viszonylag új kutatási terület, amelyet nem csupán nyelvészek, hanem matematikusok és fizikusok is alakítanak (Kovács 2013). Az utóbbi évek kutatásai a magyar mint idegen nyelv tanításának szempontjából vizsgálták a mentális lexikon felépítését és elemeinek kapcsolódását (vö. Marthy 2010; Nádor 2013). Ezek a szóasszociációs kísérletek nem csupán a mentális lexikon szerkezetének nyelvspecifikus leírását teszik lehetővé, hanem jelentősen hozzájárulhatnak a nyelvtanítás sikerességéhez is (Somos 2015).

Felépítésével kapcsolatban számos, olykor egymásnak ellentmondó elképzelés létezik, abban azonban a legtöbb kutató egyetért, hogy a mentális lexikon alapvetően három területre bontható: az aktív, a passzív és az éppen aktivált szókincs (vagy rész). Ezeknek az egységeknek a határai, illetve tartalmuk élethelyzettől függően folyamatosan változnak (Gósy 2005), a bennük tárolt elemek pedig gyakran „vándorolnak”. Az aktív tartományhoz tartoznak azok a nyelvi és beszédjelek, amelyek az *egyén* nyelvhasználatában gyakran előfordulnak. A passzív részben található az az elemek, amelyek *ugyanezen* egyén nyelvhasználatában ritkábban aktivizálódnak. Az aktivált szókincs az egy adott beszédhelyzetben, a kommunikációs célnak megfelelően aktivált nyelvi és beszédjeleket jelenti, amelyek mind az aktív, mind a passzív részben tárolódhatnak.

A mentális lexikon tartalma, illetve elemeinek elrendeződése és egymáshoz való viszonya nem csupán egyének között, hanem egyazon személy különböző élet-helyzeteiben vagy életszakaszaiban is változik. A tárolt lexikális elemek száma természetesen folyamatosan növekszik az anyanyelv-elsajátítás során, ám az elemek mentális lexikonon belüli helye sem állandó. Ha a gyermek új szót hall (amely mentális lexikonának nem képezi még részét), először a kontextus alapján megpróbálja értelmezni, majd az új elem beépül a passzív szókincsbe (Gósy 2005). Az életkor előrehaladásával és a nyelvi környezet változásával (pl. intézményesített oktatás kezdete, új tantárgyak tanulása, szakkörökön és sportfoglalkozásokon való részvétel stb.) ugyanez a lexikális elem – a beszédprodukció és a beszédfeldolgozás során történő gyakori aktiválásnak köszönhetően – később átkerülhet az aktív szókincsbe. Az írott vagy hangzó szöveg megértésének sikerességét a lexikális hozzáférés stabilitása és sebessége befolyásolja. Ebből következik, hogy minél több olyan szó fordul elő egy adott szövegben, amely nem tagja a gyermek mentális lexikonának, vagy azt csak ritkán szükséges a mindennapi nyelvhasználat során aktiválnia, annál bizonytalanabb és nehezebb az értés (Gósy 2005).

Az anyanyelv-elsajátítás során a szókincs mennyiségi és minőségi változásokon is keresztül megy (Bóna 2017). Gósy és Kovács (2001) a mentális lexikont vizsgálta szóasszociációkon keresztül, az eredmények pontos képet adnak a pubertáskorú, 12-13 éves gyermekek által ismert mentális szavak és lexémák életkori sajátosságairól (kutatásuk alapjául Cser János 1939-es szóasszociációs kísérlete szolgált). Eredményeik alapján elmondható, hogy a 13 évesek szignifikánsan több mentális szót és lexémát aktiváltak, mint a 12 évesek. Az eredményeket egyrészt a mentális lexikon kitöltöttségével (az abban tárolt elemek kvantitatív különbségével) magyarázzák, másrészt pedig azzal, hogy az idősebb korosztálynál a lexikális hozzáférés gyorsabban és pontosabban működik (Gósy–Kovács 2001). A szófaji megoszlás tekintetében elmondható, hogy a 12-13 éveseknél egyaránt a főnevek dominálnak, ezt követik az igék, a melléknévek, az igenevek és a határozószók (Gósy–Kovács 2001).

A szókincs összetételének változásában kulcsfontosságú szerepe van az oktatási rendszernek, amely hatására a szókincs specifikusan (pl. tantárgyi kereteken mentén) bővül (vö. Lukács et al. 2014; Bunta et al. 2016). A szókincs átrendeződését mutatja be Horváth (2016) kísérlete, aki a *játszik* szó relatív gyakoriságát vizsgálta 7–9 éves gyermekek narratívájában. Azt találta, hogy míg a *játszik* előfordulása fokozatosan csökken, más szavaké (pl. *iskola*) növekszik. A szókincs változása természetesen nem fejeződik be 9 éves korban; 10-14, illetve 18-24 éves beszélők esetében is nagy különbségeket figyelhetünk meg (Lengyel Zs. 2011).

A mentális lexikon felépítését, illetve a lexikális hozzáférés (temporális) sajátosságait meghatározó tényezők vizsgálatában az életkor előrehaladása mellett az egyik legfontosabb kérdés, hogy a lexikális hozzáférés során milyen „formájú” elemeket érünk el, azaz kész szóalakokat vagy pedig különálló morfémákat aktiválunk. A holisztikus modell szerint minden szóalak önálló reprezentációval rendelkezik (Butterworth 1983), bizonyos nyelvek esetében azonban (pl. finn vagy magyar) valószínűtlennek tűnik a több száz lehetséges alak különálló szóként való tárolása. Pléh szerint (2000) – többek között – nincs elegendő időnk arra, hogy minden egyes

szóalakot legalább egyszer halljunk. A dekompozíciós modellek (vö. Stanners et al. 1979; Caramazza et al. 1985) a mentális lexikon kettős tagolódását feltételezik: a szótövek és a toldalékok külön egységként tárolódnak. E modellek előnye, hogy lehetséges magyarázatot adnak arra, miért nehezebb a toldalékolt szavak aktiválása: ezeknek az alakoknak az „összetevőit” két külön helyről kell lehívnunk, majd bizonyos szabályok szerint összeillesztenünk (Gósy 2005). Az ehhez szükséges többletenergia okozhatja az ilyen alakok produkciójának és a feldolgozásának nehezítetttségét. A szótövek és a toldalékok elkülönülő tárolása esetén a szófelismerés során analitikus megközelítés feltételezhető. A két ellentétes elmélet között léteznek átmeneti modellek is, amelyek szerint bizonyos szóalakok egészen teljesen tárolódnak (pl. együttesen gyakrabban aktivált szótő-toldalék párok), míg más szavak a toldalékoktól különálló bemenettel rendelkeznek (Vannest–Boland 1999). A legvalószínűbb válasz az, hogy a nyelvek morfológiai sokfélesége miatt lehetetlen univerzális működésmechanizmust meghatározni (Libben–Jarema 2004). Vélhetően a gazdag morfológiájú, agglutináló nyelvek beszélőinek mentális lexikona másként strukturálódik, mint a nem agglutináló nyelvet beszélőké.

## 2. A lexikális hozzáférés és a szógyakoriság kapcsolata

A lexikális hozzáférés folyamatát – különös tekintettel a temporális jellemzőkre – befolyásoló hatások közül a legtöbbet kutatott a szógyakoriság tényezője, amely mind a produkciót, mind pedig a szófelismerést is érinti (Pléh–Lukács 2015). A szógyakoriság különböző módokon befolyásolhatja a lexikális aktivációt: a gyakoribb szavakat vagy szó szerkezeteket rövidebb idő alatt vagy kevesebb erőfeszítéssel hívjuk elő (Jurafsky et al. 2002). Ebből adódik, hogy számos pszicholingvisztikai vizsgálat támaszkodik a nyelvstatisztika eredményeire és módszereire (vö. Miller 1951; Baayen et al. 1997; Kostić 1995), illetve tekinti kiindulópontjának a Zipf-törvényt (Zipf 1949).

A szógyakoriság lexikális hozzáférésre gyakorolt hatásával kapcsolatos kutatásokra vonatkozóan több módszertani kérdés is felmerül. Egyrészt ezek a kutatások jellemzően írott nyelvi korpuszok elemzésén alapulnak, így nem sokat árulnak el a beszélt nyelvben előforduló szavak gyakorisági eloszlásairól. Másrészt az egyes szavak használati gyakorisága és ezzel összefüggésben ezek aktiválási sajátosságai az idő múlásával és az egyén környezetének függvényében nagymértékben változhatnak (Pléh–Lukács 2015). Harmadrészt pedig elképzelhető, hogy a lexikális hozzáférést egyáltalán nem az adott nyelv lexikális egységeinek statisztikai eszközökkel mért gyakorisága, hanem ezek egyéni nyelvhasználatban való előfordulása határozza meg. Elsőként Caroll (1971) foglalkozott annak a lehetőségével, hogy a szógyakoriságról szubjektív benyomásunk is lehet. Vizsgálatában a kísérleti személyeknek szavak gyakoriságáról kellett ítéletet hozni. Kísérletének érdekessége, hogy a nyelvhez különböző módon viszonyuló adatközlőket vont be. Az eredmények azt mutatták, hogy a szótárírók szubjektív ítéletei erősen korreláltak az objektív gyakorisággal, míg az egyetemisták ítéletei kevésbé. Az előbbi csoport eredményeit magyarázhat-

ja – többek között – a feltételezett metanyelvi tudásuk. Hasonló kísérletet végzett Desrochers és Bergeron (2000), amelyben a kísérleti személyek egy hétfokozatú skálán osztályozták francia főneveket. Ferrand és munkatársai (2008) arra hívták fel a figyelmet, hogy (a kanadai franciában) bizonyos szemantikai mezőhöz kapcsolódó szavakat alulbecsülnek, míg más mezőkhöz tartozókat túlbecsülnek, így a szubjektív gyakorisági megítélésre ez a tényező is hatással lehet.

### 3. Szövegnehézségi mutatók

A szövegnehézség fogalmának számos eltérő definíciója ismeretes. Az egyik legáltalánosabb meghatározás szerint a szövegnehézség olyan tényező, amely megmutatja, hogy milyen nehezen vagy könnyen értünk meg egy adott szöveget az írás *stílusának* köszönhetően (Klare 1963). McLaughlin (1969) definíciójában a szöveg különböző aspektusai mellett már figyelembe veszi a szöveg olvasóját (vagy hallgatóját is): a szövegnehézséget a szöveg és a befogadó viszonya mentén írja le, és megemlíti, hogy fontos tényező az is, hogy milyen figyelemfelkeltőnek és érdekesnek találunk egy adott írott vagy hangzó közlésegyiséget. Hargis és munkatársai (1998), Klare-hez hasonlóan, általános definíciót adnak a szövegnehézségre, amelyben a befogadó viszonya a szöveghez nem jelenik meg: a szövegnehézség azt mutatja meg, hogy egy szöveg szavait és mondatait milyen könnyedén tudjuk elolvasni. Fontos kiemelni, hogy a szövegnehézség (*readability*) és az olvashatóság (*legibility*) nem azonos terminusok: utóbbi nem a szöveg nyelvi összetettségére (szavak, mondatok jellemzői, a szöveg stílusa stb.) utal, hanem a szöveg megjelenésére, például a betűtípusra vagy a betűk méretére.

Az 1920-as évektől kezdődően számos oktatás- és olvasáskutató foglalkozott azzal a kérdéssel, hogy egy adott szövegegység szókinccse és mondatainak hossza miként befolyásolja a szöveg nehézségét, illetve hogy miként hozható létre egy olyan egységes mérőszám, amely ezeket a tényezőket figyelembe véve számszerűsítve kifejezheti az adott szöveg nehézségi szintjét. Az így létrehozott mérőszámokat számos különböző területen alkalmazták, beleértve az újságírást, az egészségügyet, illetve a különböző jogi és biztosítási szolgáltatásokat. Az amerikai hadsereg saját szövegnehézségi mérőszámai alapján állította össze oktatóanyagait (DuBay 2004). Az első, tankönyvek szövegnehézségének meghatározására létrehozott formula Lively és Pressey (1923) nevéhez köthető. Mérőszámukban figyelembe vették, többek között, az ezer szóra eső különböző szavak számát. Az egyik legismertebb és a tudományos kutatásokban használt, a szövegyanyagok nehézségének meghatározására is leggyakrabban alkalmazott mérőszám a Fog-index (Gunning 1952), amely mindössze két változót vesz figyelembe: az átlagos mondathosszt és a kettő (majd későbbi változataiban három) szótagnál hosszabb szavakat. A 20. század második felében további mérőszámokat dolgoztak ki, ezeknél már újabb tényezőket is figyelembe vettek, így a szövegnehézség kutatásának új irányvonala rajzolódott ki. Többek között fontos változás, hogy a különböző formulák kidolgozásában már nyelvészeti és kognitív kérdések is felmerültek, illetve számos oktatáskutató úgy

vélte, hogy a szöveg nehézsége pontosabban megítélhető, ha az – valamilyen módon – nem mentes a szubjektivitástól (DuBay 2004). A szövegnehézség meghatározására használatos mérőszámok elsősorban az angol nyelvre alkalmazhatók, hiszen ezek főként az adott nyelv morfológiai, szintaktikai és szövegszervezési sajátosságain alapulnak. A Fog-index, bár használható magyar nyelvű szövegeken, nem veszi figyelembe a magyar nyelv agglutináló jellegét, így a szöveg nehézségéről adott kép torzulhat.

Az egyén mentális lexikonának tartalmát, illetve a lexikális hozzáférés folyamatát számos tényező befolyásolja. Ezeknek a tényezőknek az egyike, amely összefügg az egyén élethelyzetének változásaival, illetve az egyén környezetéből érkező nyelvi információ összetételével, a szavak gyakorisága. A szógyakoriság többek között meghatározza a mentális lexikon elemeinek aktiválásához szükséges erőfeszítést és a hozzáférés temporális jellemzőit. A ritkábban előforduló szavak előhívása bizonytalanabb, amely a nyelvi alapú információ globális feldolgozását nehezíti. A szógyakoriságokról alkotott szubjektív ítéletek mint lexikális mérőszámok így hozzájárulhatnak egy, a különböző szövegek nehézségét mérő index kialakításához.

Kutatásomnak két célja volt: (i) különböző szófajú és morfológiai felépítésű szavak gyakoriságérzetének vizsgálata 10–11 és 13–14 évesek körében és (ii) egy, a lexikális mérőszámokat kiindulópontként felhasználó, a magyar nyelv morfológiai jellegzetességeit figyelembe vevő szövegnehézségi index alapjainak a kialakítása.

A kutatáshoz két hipotézist állítottam fel:

1. A különböző szavak aktiválása függ az életkortól.
2. A gyakorisági döntések egyazon életkori csoporton belül nem függetlenek a szókategóriáktól.

#### 4. Adatközlők, anyag, módszer

A kutatásban 50 adatközlő vett részt, akiket két különböző életkori csoportból (10–11 évesek és 13–14 évesek) válogattunk ki véletlenszerű mintavételi módszerrel. Az adatközlők mindegyike magyar anyanyelvű, tipikus fejlődésű és ugyanazon vidéki általános iskola 5. és 7. osztályába jár. A kísérletbe való bevonásukban az osztályok magyartanára volt segítségemre.

A kutatáshoz egy 45 elemből álló szólistát alkalmaztam. A szólista elemei különböző szókategóriákhoz tartoznak. A szókategóriák kialakításában és elemeinek a kiválasztásában elsődleges szempont az volt, hogy minden szófaj arányosan képviselve legyen. Az egyes, általam meghatározott szókategóriák nem egyeznek teljesen a szófaji határokkal, hiszen bizonyos kategóriák szűkebb szótípust fednek le, illetve bizonyos esetekben a szókategóriák kiválasztásánál nem a szófaji besorolást vettem alapul, hanem egyéb tényezőket, például a szó eredetét (idegen eredetű szavak) vagy a szó morfológiai szerkezetét (toldalékoltszavak, igeekötős igeik). A vizsgálatban alkalmazott szókategóriák meghatározása így szófaji besorolás, morfológiai felépítés, illetve a szó eredete szerint történt. Ennek megfelelően a szólista elemeit 11 kate-

góriába osztottam: konkrét jelentésű egyedi és anyagnevek toldalék nélkül (*bokor*) és legalább két toldalékkal ellátva (*szekrényekkel*), konkrét jelentésű gyűjtőnevek (*közönség*), elvont értelmű köznevek toldalék nélkül (*emberszeretet*) és legalább két toldalékkal ellátva (*jólneveltségeteknek*), hétköznapi igék (beleértve a főnévi ige- neveket) időjel nélkül és időjellel (*beszéltek, úsztatsz, mászni*), igekötős igék (*meg- nyomja*), igekötős igék, amelyek jelentése teljesen eltér az igekötő nélküli alakétól (*megeőrltet, felhajt*) gyakori melléknevek toldalék nélkül és toldalékkal ellátva (*bo- rús, elolvasandó, legvagányabbik*), határozószók (*holnapután, benn, gondosan*) és idegen eredetű szavak (*konzol, megreguláz, monitor*).

Az adatközlőket arra kértem, hogy a szólista minden elemét osztályozzák egy 1-től 5-ig terjedő Likert-skálán aszerint, hogy az adott szót egyéni nyelvhasználatukban milyen gyakran ejtik vagy hallják (1 = nagyon ritkán használom és hallom; 2 = rit- kán használom és hallom; 3 = közepesen gyakran használom és hallom; 4 = eléggé gyakran használom és hallom; 5 = nagyon gyakran használom és hallom). A kísérletet megelőzően a kérdőív használatát egy 5 adatközlőből álló mintán próbáltam ki. A próbakísérletek azt igazolták, hogy már a 10 évesek is képesek a feladat megoldására, vagyis választ tudnak adni arra, hogy egy szót milyen gyakran (vagy ritkán) mondanak ki, illetve hallanak a környezetükben. A kérdőív kitöltése az iskolai ma- gyarórán történt, a kísérletvezető és a magyartanár felügyelete mellett. A gyerekek- nek pontosan megmagyaráztuk, hogy mit jelentenek az egyes osztályzatok.

Az eredmények elemzése során az egyes szavakra, illetve szó kategóriákra kapott gyakorisági átlagértékeket hasonlítottam össze mindkét életkori csoportban, különös tekintettel a toldalékoltsági és igekötős alakokra kapott értékek életkorok közötti kü- lönségeire, illetve az általam ritkábbnak és ebből adódóan nehezebben feldolgoz- hatónak ítélt szavakra kapott értékek eltéréseire. Az utóbbi szavak kiválasztásánál egyrészt Gósy és Kovács (2001) eredményeit vettem figyelembe, ahol az általuk gyűjtött korpuszban ritkábban előforduló szavakat tekintettem nehezebbnek (ige- nevek). Másrészt figyelembe vettem a szavak eredetét (lásd idegen eredetű szavak), illetve a szójelentés absztraktságát (lásd elvont értelmű köznevek).

A statisztikai elemzéshez az SPSS 20.0 szoftvert használtam (Khi-négyzet próba, ahol a független változó az életkori csoport volt, a függő változó pedig a gyakorisági ítélet).

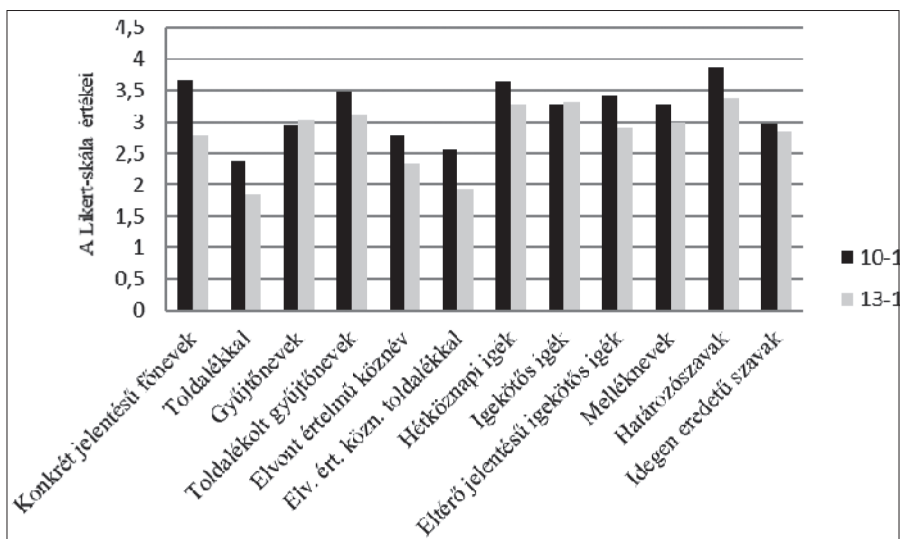
## 5. Eredmények

A Khi-négyzet próba számos szó esetében igazolt összefüggést az életkor és a gya- korisági ítéletek között (1. táblázat). Az életkorral összefüggést mutató gyakorisági ítéletek nem korlátozódnak egy adott szó kategóriára sem (konkrét jelentésű egyedi és anyagnevek, elvont értelmű köznevek, igék, igekötős igék, igenevek, határozó- szók, idegen eredetű szavak).

1. táblázat. Az életkorral szignifikáns összefüggést mutató gyakorisági ítéletek

Szavak	10–11 évesek gyakorisági átlagértékei	13–14 évesek gyakorisági átlagértékei	Életkor és gyakorisági ítélet közötti összefüggés szignifikanciaszintje
<i>bokor</i>	3,4	2,44	0,029
<i>játszótér</i>	3,16	2,36	0,052
<i>templom</i>	4,92	3,28	0,000
<i>emberszeretet</i>	3	2,28	0,048
<i>úsztatsz</i>	2,76	2	0,018
<i>felhajt</i>	3,16	2,36	0,043
<i>párnázott</i>	2,96	1,96	0,040
<i>elolvasandó</i>	3,44	3	0,017
<i>gondosan</i>	3,64	3,08	0,040
<i>konzol</i>	3,84	3,6	0,031

Kivétel nélkül minden szóhoz az idősebb korosztály rendelt alacsonyabb gyakorisági értékeket. A legalacsonyabb szignifikanciaértéket a *templom* szó esetében adtuk. Megvizsgáltam a két életkori csoport átlagolt gyakorisági ítéleteit szókatégoriák szerint (1. ábra). Két szókatégoriától eltekintve (gyűjtőnevek és igeikötős igeik) ebben az összehasonlításban is azt találtam, hogy a fiatalabb korosztály rendelt magasabb értékeket a szavakhoz. A legjelentősebb eltérést a konkrét jelentésű főnevek esetében kaptam (10–11 évesek: 3,66; 13–14 évesek: 2,78), jelentős eltérés mutatkozik azonban a határozószók (3,85; 3,37), a toldalékkal ellátott konkrét jelentésű főnevek

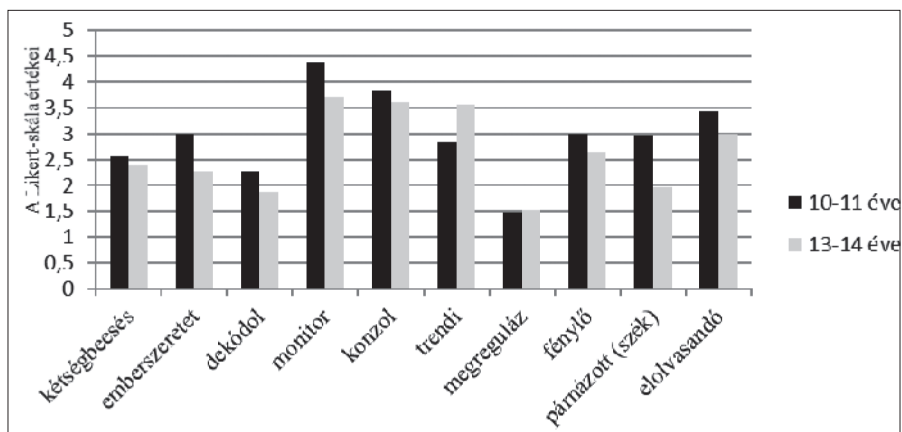


1. ábra. Szókatégoriák gyakoriságának megítélése a két életkori csoportban



(2,37; 1,85), a toldalékkal ellátott gyűjtőnevek (3,48; 3,1) és a toldalékkal ellátott elvont értelmű köznevek (2,78; 2,34) kategóriájában is. Várakozásaimmal ellentétben azonban nem adatoltunk jelentős eltérést több szókategória átlagértékében. Feltételeztem, hogy az idősebb korosztály mindennapi nyelvhasználatának szervezettebb részét képezik az idegen eredetű szavak, mint a fiatalabb korosztályét, az életkori csoportok közötti különbség azonban elenyésző (2,97; 2,86). Szintén elhanyagolható különbséget kaptam az igeikötős igeik (3,26; 3,31) esetében.

A ritkábban használtak, illetve ebből fakadóan nehezebben feldolgozhatónak vélt szavakra adott gyakorisági ítéletek összevetését mutatja a 2. ábra. Jól látható, hogy az életkorral szignifikáns összefüggést mutató szavak közül négy ebben a felsorolásban foglal helyet (bár ezek várakozásaimmal ellentétes tendenciát mutatnak). Egyedül a *trendi* szó esetében mutatkozik jelentősen magasabb értékű gyakorisági ítélet az idősebb korosztálynál (2,84; 3,56); a *megreguláz* szót az idősebb korosztály bár gyakoribbnak jelölte, a különbség gyakorlatilag elhanyagolható (1,48; 1,52). A legkiugróbb különbség a *párnázott* szónál figyelhető meg (2,96; 1,96), ezt követi az *emberszeretet* (3; 2,28), illetve a *monitor* (4,4; 3,72). Fontos kiemelni, hogy bár az *emberszeretet* és a *kétségbeesés*, a *dekódol*, *monitor*, *konzol*, *trendi*, *megreguláz*, illetve a *fénylő*, *párnázott* és *elolvasandó* azonos szókategóriába tartoznak (elvont értelmű köznevek, idegen eredetű szavak és igeik [azon belül is igenevek]), az egyes szavak közötti gyakorisági eltérések nem kiegyenlítettek. Míg az *emberszeretet* esetében például 0,72 a két életkori csoport közötti különbség, a *kétségbeesés* esetében már alig 0,16. A *megreguláz* esetében elmondható, hogy bár a két korcsoport gyakorisági ítéletének átlaga kiegyenlített, az messze alul marad a többi idegen eredetű szó értékéhez képest, így feltételezhető, hogy egyik korcsoport számára sem ismert, gyakoriságának megítélése esetleges volt.



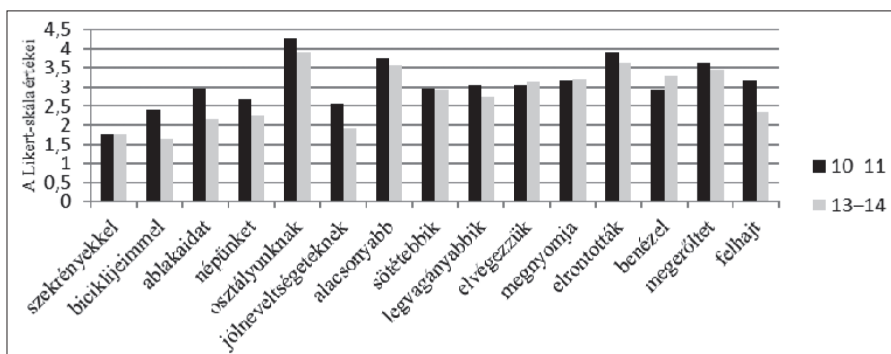
2. ábra. Nehezebben feldolgozható szavak megítélése a két korcsoportban

A 3. ábrán a toldalékkal szótalakokra adott gyakorisági ítéleteket láthatjuk a két életkori csoportban. Bár a statisztikai elemzés csak a *felhajt* szó esetében igazolt



összefüggést az életkor és a gyakorisági ítéletek között, egyértelmű trend látszik ki-rajzolódni: a fiatalabbak – a *szekrényekkel*, az *elvégezzük*, a *megnyomja*, és a *benézel* szavak kivételével – mindenhol magasabb gyakorisági ítéletet rendeltek a lista elemeihez. Az átlagolt gyakoriságiítéletek közötti legjelentősebb eltérést a raggal ellátott főnevek esetében láthatjuk (*biciklijeimmel*: 2,4; 1,64 és *ablakaidat*: 2,96; 2,16).

Azoknak a szavaknak az esetében, ahol az idősebb korosztály jelölt magasabb gyakorisági ítéletet, a különbség nem számottevő az átlagolt értékekben. Fontos megfigyelés, hogy ezeknek a szavaknak mindegyike az igekötős igék csoportjához tartozik (*elvégezzük*: 3,04; 3,12, *megnyomja*: 3,16; 3,20, illetve *benézel*: 2,92; 3,28). Az általam meghatározott szókategóriákban külön csoportot képeztek azok az igekötős igék, amelyek az igekötővel teljesen más jelentést vesznek fel a szótő jelentéséhez képest. Ezeknek a szavaknak az esetében – ellentétben az olyan igekötős igékkel, amelyeknél az igekötő csupán módosítja a szótő jelentését – a fiatalabb korosztály jelölt magasabb gyakorisági ítéleteket (*megerőltet*: 3,64; 3,44 és *felhajt*: 3,16; 2,36).



3. ábra. Toldalékkal ellátott szavak megítélése a két életkori csoportban

## 6. Következtetések

Kutatásomban azt vizsgáltam, hogy két életkori csoportban (10–11 évesek és 13–14 évesek) miként változik a különböző szófajú és morfológiai felépítésű szavak gyakoriságának szubjektív megítélése. Ehhez egy ötfokozatú Likert-skálát alkalmaztam, ahol a magasabb értékek jelölik a szavakhoz rendelt magasabb szubjektív gyakorisági ítéletet.

Az első hipotézis, miszerint a különböző szavak aktiválása függ az életkortól, igazolódott. Az eredmények azt mutatják, hogy – kevés kivételtől eltekintve – a fiatalabb korosztály konzekvensen gyakoribbnak ítélte a kísérletben használt szavakat, mint az idősebb korosztály. Azoknak a szavaknak az esetében, amelyeknél a statisztikai elemzés is összefüggést igazolt az életkor és a gyakorisági ítéletek között, szintén azt láthatjuk, hogy a fiatalabb korosztály magasabb értékeket jelölt. Fontos azonban

kiemelni, hogy bár a hipotézis igazolódott, az eredmények nem állnak összhangban a gyakorisági ítéletekben kirajzolódó trend előzetesen feltételezett irányával. Feltételeztem, hogy bizonyos, általunk ritkábbnak vagy nehezebben feldolgozhatónak vélt szavak és szóalakok esetében (pl. idegen eredetű szavak, toldalékolt szóalakok, igenevek, elvont értelmű köznevek, eltérő jelentést felvevő igekötős igék) az idősebb korosztály fog magasabb értékeket jelölni.

A fiatalabb korosztály által jelölt magasabb értékek magyarázhatók azzal, hogy a szavak gyakoriságának, illetve a gyakorisági tényezők keresztül a szavak érthetőségének szubjektív megítélését az életkoron, illetve az anyanyelvi tudatosság eltérő szintjén kívül más tényezők is befolyásolják. Ilyen tényező lehet – többek között – a gyermek szociokulturális környezete, az intézményesített oktatás útján szerzett nyelvi tapasztalat vagy a gyermek számára biztosított nyelvi környezet hatása (Gósy 2005). A szociokulturális környezet mentális lexikonra gyakorolt hatását bizonyítja a kísérletben alkalmazott *templom* szó gyakoriságítéletében mutatkozó jelentős eltérés (4,92; 3,28): a fiatalabb korosztály – az őket körülvevő iskolai környezetben beálló változásnak köszönhetően – mentális lexikonjában tárolt, vallással kapcsolatos elemek gyakrabban aktiválódnak a mindennapi szóhasználatban, mint az idősebb korosztály esetében, így a fiatalabbak a *templom* szót gyakoribbnak érzik.

Az eredmények további lehetséges magyarázata, hogy a fiatalabb korosztály a szavak gyakoriságának megítélése során kevésbé analitikusan gondolkodik, mint az idősebb korosztály. Elképzelhető, hogy a fiatalabbak kisebb mértékben vették figyelembe a szavak morfológiai struktúráját (lásd toldalékolt szavakra kapott eredmények), sokkal inkább egészelegesen tekintették a különböző alakokat, és így kevésbé tettek különbséget a toldalékolt, illetve a toldalék nélküli szóalakok között. A toldalékolt alakokra kapott eredmények azzal is magyarázhatók, hogy a gyakoriságok meghatározása során a fiatalabbak elsősorban a szótöveket vették figyelembe, és a toldalékolt szóalakok megítélése a toldalék nélküliekével hasonló módon történt. A morfológiailag összetettebb alakok eredményeinek értelmezésénél így nem elhanyagolható az a szempont, hogy valójában az adatközlők a toldalék nélküli szavakhoz hasonló módon kezelték ezeket. Az eredmények alapján továbbá feltételezhető, hogy a fiatalabb vizsgálati személyek az egyes szavak jelentéstartalmát is kevésbé tudatosan elemezték, mint az idősebbek (lásd idegen eredetű szavakra, elvont értelmű köznevekre, igenevekre kapott eredmények). Bár nem zárható ki, hogy a fiatalabb korosztályt érő nyelvi inputban is előfordulnak az általunk nehezebben feldolgozhatónak és ritkábbnak vélt szavak, és így ezek az adatközlők a valós nyelvi tapasztalatukat vették alapul a válaszadás során, elképzelhető, hogy az elemek használatakor azok szemantikai tartalmát nem vagy csak részben alkalmazzák adekvát módon.

Az idősebb korosztály eredményeit ezzel szemben magyarázhatja a magasabb szintű metanyelvi tudatosság, illetve az ennek eredményeképpen alkalmazott analitikus nyelvi gondolkodás. Ez a megfigyelés összecseng Imre és Hertendi (2013) kutatásának eredményeivel, akik az óvodások, kisiskolások és idősebbek szódefiniációinak elemzése során azt találták, hogy az életkor előrehaladásával fokozatosan csökken a szódefiniálási feladatokban jelölt „nincs válasz” kategória előfordulása, amely a lexikális tudatosság fokozatos emelkedését jelzi (elsősorban iskoláskortól).

Ennek kialakulásában nem elhanyagolható az iskolai nyelvtanoktatás során szerzett tudás szerepe. Az idősebb korosztály tagjai 7. és 8. osztályos tanulók, akik már tanultak mondatelemzést, illetve a különböző toldalékok sorrendje és funkciója is több alkalommal visszatérő eleme a tananyagoknak. Ezzel szemben a fiatalabb korosztály tagjainak jelentős része a kísérlet időpontját megelőzően léptek felső tagozatba, így ők a morfológiai alapfogalmakról jelentősen kevesebb ismerettel rendelkeznek. Az igekötős igékre kapott eredmények – szemben más szó kategóriákkal, például a főnevekkel – a két életkori csoportban kiegyenlítették (az idősebb korosztály átlagértékei valamivel magasabbak, de a különbség elhanyagolható). Ezek az eredmények magyarázhatók azzal a megfigyeléssel, miszerint a kisiskolások szókincsét kezdetekben főnévi dominancia jellemzi, később fokozatosan növekszik az igék és más szófajok aránya (Nagy 1978). Mivel az igék aránya fiatalabb életkorban alulmarad a főnevekhez képest, feltételezhető, hogy – függetlenül az adott szótól – a morfológiailag összetettebb, igekötővel ellátott igék gyakoriságának megítélésében is nyomon követhető a szókincs összetételének ezen jellemzője, amely az életkor előrehaladásával később változik.

A második hipotézis, miszerint a gyakorisági döntések egyazon életkori csoporton belül nem függetlenek a szó kategóriáktól, szintén igazolódott. Az eredmények alapján megállapíthatjuk, hogy egyazon életkori csoport adatközlői jelentősen eltérő értékeket rendeltek a különböző kategóriájú szavakhoz. A legalacsonyabb és a legmagasabb átlagértékek tekintetében nagyon hasonló eredmények adathatók: mind a fiatalabbak, mind pedig az idősebbek a toldalékkal ellátott konkrét jelentésű főnevek esetében jelölték a legalacsonyabb értékeket (2,37; 1,85, lásd 1. ábra), a legmagasabb értékeket pedig a határozószók esetében adták (3,85; 3,37, lásd 1. ábra). Megfigyelhető, hogy a főnevekhez köthető szó kategóriák esetében mindkét életkori csoportban rendre markánsan alacsonyabb értékeket kaptak a toldalékolt szóalakok az ugyanazon kategóriához tartozó, toldalék nélküli szóalakokhoz képest. Valószínűsíthető, hogy bár jelentősen eltérő mértékben, de mindkét életkori csoportban szerepet játszott a szavak megítélésében azok morfológiai felépítése. Ez a megfigyelés nem mond ellent az első hipotézis kapcsán levont következtetéseknek, pusztán árnyalja azokat. A kísérletben alkalmazott szavak gyakoriságának megítéléséből arra következtethetünk, hogy mindkét életkori csoport – eltérő mértékben – tudatosan döntött a gyakorisági ítéletek meghozatalakor, figyelembe véve a szavak (összetettebb vagy egészen egyszerű, hétköznapi) jelentéstartalmát is: a határozószók, a hétköznapi igék és a konkrét jelentésű főnevek – az adott életkori csoporton belül – közel azonos értékeket mutattak. Ez a fajta tudatosság azonban nem feltétlenül követhető nyomon minden szó kategóriában. A fiatalabb életkori csoport egyformán gyakorinak ítélte például a gyűjtőneveket és az idegen eredetű szavakat, az idősebb korosztály pedig a toldalékkal ellátott konkrét jelentésű főneveket látta el jóval alacsonyabb értékkel, mint az idegen eredetű szavakat.

Mindezeket figyelembe véve elmondható, hogy az idősebb korosztály a különböző szó kategóriába tartozó, illetve a morfológiai összetettség tekintetében is eltérő szavak gyakoriságának a megítélésben – a fiatalabb korosztályhoz képest – tudatosabban és analitikusabban döntött. Ez a szilárdabb anyanyelvi tudatossággal, a magasabb szintű

metanyelvi ismeretekkel, illetve a nyelvi tapasztalat kvantitatív különbségével magyarázható. A fiatalabb korosztály gyakorisági ítéletei alapján, bár megfigyelhető bizonyos szintű elemző megközelítés és szemantikai disztingválás ebben az életkorban is, alapvetően nem egyértelmű ennek a kísérletnek az eredményei szerint, hogy milyen tényezők alapján döntenek a lexikális elemek gyakoriságáról.

A jelen kutatás korlátai között fontos megemlíteni egyfelől az adatközlők relatíve alacsony számát, illetve a gyakorisági ítéletekhez alkalmazott szavak mennyiségét. Az egyes szókatégoriák aktiválására vonatkozó következtetéseket árnyalná és pontosítaná, ha azokat több lexikális elemről alkotott gyakorisági ítéleten keresztül vizsgálnánk. A kutatás eredményeiből nyilvánvalóan nem derül ki, hogy az egyes szavakra adott értékek tükrözik-e azok ismeretét és értését. A cél azonban nem is ennek megítélése volt, hanem egyfajta szubjektív döntésnek az elemzése, amely ugyanakkor utalhat az adott szavak jelentésének az ismeretére is. A kutatás eredményei adalékkal szolgálnak a 10–14 évesek mentális lexikonának a megismeréséhez.

## SZAKIRODALOM

- Aitchison, Jean 2003. *Words in the mind. An introduction to the mental lexicon*. Blackwell, Oxford.
- Baayen, Harald R. – Dijkstra, Ton – Schreuder, Robert 1997. Singulars and plurals in Dutch. Evidence for a dual parallel route model. *Journal of Memory and Language* 37: 94–119. <https://doi.org/10.1006/jmla.1997.2509>
- Bóna Judit – Imre Angéla 2017. Szójelentések meghatározása 5 és 9 éves kor között. *Beszéd-kutatás 2017*: 185–204.
- Bunta Ferenc – Bóna Judit – Gósy Mária 2016. HU-LARSP: Assessing children’s language skills in Hungarian. In: Fletcher, Paul et al. (eds.): *Profiling grammar: More languages of LARSP. Communication disorders across languages 15*. Multilingual Matters Ltd., Bristol, 80–98. <https://doi.org/10.21832/9781783094875-007>
- Butterworth, Brian 1983. Lexical representation. In: Butterworth, Brian (ed.): *Language production. Vol. 1*. Academic Press, London, 125–53.
- Caramazza, Alfonso – Miceli, Gabriele – Silveri, Caterina M. – Laudanna, Alessandro 1985. Reading mechanisms and the organisation of the lexicon: Evidence from acquired dyslexia. *Cognitive Neuropsychology* 2: 81–114. <https://doi.org/10.1080/02643298508252862>
- Carroll, John B. 1971. Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior* 10: 722–29. [https://doi.org/10.1016/S0022-5371\(71\)80081-6](https://doi.org/10.1016/S0022-5371(71)80081-6)
- Cser János 1939. *A magyar gyermek szókincse*. Magyar Paedagogiai Társaság, Budapest.
- Desrochers, Alain – Bergeron, Mylène 2000. Norms of subjective frequency of use and imagery for a sample of 1,916 French nouns. *Canadian Journal of Experimental Psychology (Revue canadienne de psychologie expérimentale)* 54: 274–325. <https://doi.org/10.1037/h0087347>
- DuBay, William H. 2004. *The Principles of readability*. Impact Information. Costa Mesa, California. <http://www.impact-information.com/impactinfo/readability02.pdf> (Letöltés ideje: 2020. 02. 12.)
- Ferrand, Ludovic – Bonin, Patrick – Méot, Alain – Augustinova, Maria – New, Boris – Pallier, Christophe – Brysbaert, Marc 2008. Age-of-acquisition and subjective frequency

- estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods* 40: 1049–54. <https://doi.org/10.3758/BRM.40.4.1049>
- Gósy Mária 1998. Szókeresés a mentális lexikonban. *Magyar Nyelvőr* 122: 189–201.
- Gósy Mária 2005. *Pszicholingvisztika*. Osiris, Budapest.
- Gósy Mária – Kovács Magdolna 2001. A mentális lexikon a szóasszociációk tükrében. *Magyar Nyelvőr* 125: 330–54.
- Gunning, Robert 1952. *The technique of writing*. McGraw-Hill, New York.
- Hargis, Gretchen – Hernandez, Ann K. – Hughes, Polly – Ramaker, James – Rouiller, Shannon – Wilde, Elizabeth 1998. *Developing quality technical information: A handbook for writers and editors*. Prentice Hall, Upper Saddle River, New Jersey.
- Horváth Viktória 2016. Kisiskolás gyermekek spontán beszédének jellemzői. *Alkalmazott Nyelvtudomány* 16/1. <http://real.mtak.hu/85393/> (A letöltés ideje: 2019. 12. 04.)
- Imre Angéla – Hertendi Márta 2013. *A beszédtervezés stratégiai különböző életkorban*. Előadás a XV. Pszicholingvisztikai Nyári Egyetemen. Balatonalmádi, 2013. május 26–30.
- Jurafsky, Daniel – Bell, Alan – Girand, Cynthia 2002. The role of the lemma in form variation. *Laboratory Phonology* 7: 3–34. <https://doi.org/10.1515/9783110197105.1.3>
- Klare, George R. 1963. *The measurement of readability*. Iowa State University Press, Ames, Iowa.
- Kostić, Alexander 1995. Information load constraints on processing inflected morphology. In: Feldman, Laurie B. (eds.): *Morphological Aspects of Language Processing*. Lawrence Erlbaum Associates. Hillsdale, New Jersey, 317–44.
- Kovács László 2013. *Fogalmi rendszerek és lexikai hálózatok a mentális lexikonban*. Tinta Könyvkiadó, Budapest.
- Lengyel Zsolt 2008. *Magyar Asszociációs Normák Enciklopédiája I*. Tinta Könyvkiadó, Budapest.
- Lengyel Zsolt 2011. A szóasszociációs vizsgálatok újabb eredményei. *Hungarológiai Évkönyv* 12: 179–91.
- Libben, Gary – Jarema, Gonia 2004. Conceptions and questions concerning morphological processing. *Brain and Language* 90: 2–8. <https://doi.org/10.1016/j.bandl.2003.12.005>
- Lively, Bertha A. – Pressey, Sidney L. 1923. A method for measuring the 'vocabulary burden' of textbooks. *Educational administration and supervision* 9: 389–98.
- Lukács Ágnes – Pléh Csaba – Kas Bence – Thuma Orsolya 2014. A szavak mentális reprezentációja és az alaktani feldolgozás. In: Pléh Csaba – Lukács Ágnes (szerk.): *Pszicholingvisztika*. Akadémiai Kiadó, Budapest, 167–251.
- McLaughlin, Harry G. 1969. SMOG grading – a new readability formula. *Journal of Reading* 22: 639–46.
- Marthy Annamária 2010. Szóasszociációs vizsgálatok a Semmelweis Egyetem külföldi diákjainak körében. *THL2* 1–2: 124–34.
- Miller, George A. 1951. *Language and communication*. McGraw Hill, New York. <https://doi.org/10.1037/11135-000>
- Nádor Orsolya 2013. „Magyar” a magyar és a nem-magyar mentális lexikonban. Egy szóasszociációs vizsgálat tapasztalatai. *THL2* 1–2: 42–54.
- Nagy J. József 1979. A szófajok gyakorisági jellemzői a 8–10 éves tanulók nyelvhasználatában. *Magyar Nyelv* 74: 186–204.
- Navracsics Judit 2001. Kétnyelvűek mentális lexikonának jellegzetességei. *Alkalmazott Nyelvtudomány* 1: 51–61.
- Pléh Csaba – Lukács Ágnes 2015. *Pszicholingvisztika*. Akadémiai Kiadó, Budapest. <https://doi.org/10.1556/9789630594998>

- Somos Júlia 2015. Szintagmatikus kapcsolatok a mentális lexikonban – egy anyanyelvűekkel és magyarul tanuló felnőttekkel végzett szóasszociációs vizsgálat eredményei. *THL2* 1–2: 15–29.
- Stanners, Robert F. – Neiser, James J. – Hernon, William P. – Hall, Roger 1979. Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behaviour* 18: 399–412. [https://doi.org/10.1016/S0022-5371\(79\)90219-6](https://doi.org/10.1016/S0022-5371(79)90219-6)
- Vannest, Jennifer – Boland, Julie E. 1999. Lexical morphology and lexical access. *Brain and Language* 68: 324–32. <https://doi.org/10.1006/brln.1999.2114>
- Zipf, George K. 1949. *Human behavior and the principle of least effort*. Addison-Wesley Press, Cambridge, Massachusetts.

Csárdás László  
ELTE BTK  
Alkalmazott Nyelvészeti Doktori Program  
<https://orcid.org/0000-0003-4339-783X>

## SUMMARY

*Csárdás, László Bence*

### **Subjective word frequency judgements in childhood**

One of the basic factors of language comprehension is the relationship between the listener's or the reader's mental lexicon and the vocabulary of the linguistic input. The activation of the elements stored in the mental lexicon, however, greatly varies across individuals. The present research focuses on the subjective word frequency judgements of children between 10 and 14 years of age. 50 children were asked to make frequency judgements (using Likert-scale) on 45 words. The words represent different parts of speech and both suffixed forms and stems were included. Our results show that (i) word frequency judgements depend on age and (ii) are closely linked to the morphological structure and meaning of the word. However, in the case of certain words, frequency judgements seem to be ambiguous. The age-specific lexical indices can contribute to better understanding the processing of the mental lexicon of children between 10 and 14.

**Keywords:** mental lexicon, word frequency, lexical access, language comprehension, readability