

A szófajiság és a kollokáció empirikus vizsgálata a magyarban *

1. rész

1. Bevezetés. Ugyan a *szófaj* műszó többezer éves gyökerekkel rendelkezik az európai nyelvészetben, értelmezése és a szófajiság koncepciójának meghatározott (jelentéstani, alaktani, mondattani) kritériumok szerinti következetes érvényesítése máig kihívást jelent a nyelvi leírás számára. A kollokáció a korpusznyelvészet egyik alapfogalma, amely különböző szavak hajlandóságát jelöli arra, hogy a szövegben egymás közelében előforduljanak (HUNSTON–LAVIOSA 2000). Mind a kollokálódás korpuszban történő mérését, mind a kollokáció és a nyelvi képesség viszonyát illetően számos alternatív elképzelés létezik (l. BARNBROOK et al. 2013). Ebben a tanulmányban a szófajiságot és a kollokációt a beszélők tudása szempontjából vizsgáljuk. Fő kérdésünk, hogy a kollokálódás és az azonos szófaji kategóriába tartozás milyen mértékben kapcsolhatja össze a szavak mentális reprezentációját. Ennek a kérdésnek a megválaszolására lexikai döntési feladatokat végeztettünk magyar anyanyelvű beszélőkkel. A kísérleti személyeknek arról kellett minél gyorsabban döntést hozniuk, hogy létező főnevet, igét látnak-e vagy ún. álszót. A vizsgálat a magyar nyelvben létező főnevek és igék előtt megjelenő, kollokációjukban összefüggő és/vagy szófajiságukban megegyező szavak előfelező hatásának mérésére irányult.

A 2. rész a szófajisággal és a kollokációval kapcsolatos elméleti háttérfeltevéseket, illetve az ezekből adódó hipotéziseket tárgyalja, amelyek vizsgálatára kísérletet terveztünk. A 3. rész a vizsgálat korpuszalapon kialakított anyagát és módszerét ismerteti, míg a 4. az eredményeket közli, amelyeknek értékelésére az 5. részben kerül sor.

2. Elméleti feltevések és hipotézisek

2.1. A szófaj. A szófajisággal kapcsolatban számos különböző koncepció létezik a formális és a funkcionális (használatalapú) modellekben is. A magyar le-

* A Bolyai János Kutatási Ösztöndíj, valamint a Kulturális és Innovációs Minisztérium ÚNKP-23-4 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

író nyelvtan (KESZLER szerk. 2000) a szófajok meghatározásában a szintaktikai szempontok mellett alaktani és jelentéstani kritériumokat is alkalmaz, utóbbiak következetes érvényesítése azonban meglehetősen problematikus (KIEFER 2007: 145–241). A szintén formalista generatív nyelvtanok nem tulajdonítanak nagy jelentőséget az esetlegesnek tűnő lexikai szófajiságnak, és alapvetően szintaktikai természetűnek tételezik a főnévi, igei csoportokat mint kifejezés (phrase) értékű kategóriákat felépítő főnév, ige stb. (elemi alaki) kategóriákat is (É. KISS 1992). A lexikai szófajiság ebben a megközelítésben felfogható bizonyos szavakhoz rögzített szintaktikai információként (vö. KENESEI 2010). A formális modellekből tanulságként mindenképpen levonható, hogy a lexikai értelemben vett szófajiság (amely nem univerzális, hanem sokkal inkább nyelvtipológiai jelenség) a szavaknak a mondattani szerkezetekhez és viszonyokhoz való hozzájárulása felől értelmezhető, vagyis a lexikai egységek szófajiságának empirikus vizsgálatok is számolni kell az adott szófaj szintaktikai beágyazottságával.

A használatalapú kognitív nyelvtan (LANGACKER 1987, 2008) szerint a szófaj jelenségét eredményező mondattani szerkezetek és viszonyok szemantikai tényezőkre vezethetők vissza. A kognitív nyelvtan a bizonyos szófajt képviselő szavakat azonban nem a jelentésük tartalma, hanem szerkezete (a fogalmi hozzáférés módja) szerint tartja egybetartozónak. A szófajiság ebben a tekintetben a jelentések fogalmi megalkotásával, a konstruálással van összefüggésben, ugyanazon fogalmi tartalom különböző szemantikai szerkezetekben dolgozható föl a szófaji jelentésétől (is) függően. Így például a *fut* és a *futás* nyelvi szimbólumokkal a beszélők ugyanarra a referenciális jelenetre, de eltérő módon irányíthatják a beszélőtársak figyelmét, és ez az eltérő mód az, ami miatt különböző szófajú szavaknak tarthatjuk őket. A kognitív nyelvtanban a mondat egy létezőként megkonstruált, főnévvel jelölt dologgal kapcsolatos időbeli eseményt fejez ki, általában egy másik, ugyancsak főnévvel jelölt dologgal viszonyban (TOLCSVAI NAGY 2017: 391–466). A főnévi jelentésben a fogalom dologszerűen konstruálódik meg, a főnév jelentése atemporális és fogalmilag autonóm (TOLCSVAI NAGY 2017: 353–361); az igei jelentés ezzel szemben temporális és fogalmilag függő, hiszen egy folyamat szereplőket és körülményeket is feltételezhet (TOLCSVAI NAGY 2017: 310–352). Ezzel összefüggésben a magyarban az ige önmagában képes kifejezni a mondat szereplő- és időviszonyait, vagyis tulajdonképpen (az önhasonlóság elvének megfelelően) mondat a mondatban (IMRÉNYI 2017: 664–715).

Az utóbbi évtizedekben egyre nagyobb teret nyerő használatalapú nyelvtanok a szófajiság fogalmához képest sokkal inkább előtérbe helyezték a konstrukció fogalmát. A konstrukciós nyelvtanokban (CROFT 2001; GOLDBERG 1995) a nyelv primitívumai a konstrukcióknak nevezett, önálló formával és jelentéssel bíró szerkezetek. Nem csupán a szavak és az idiómák tartoznak a konstrukciók körébe, hanem a szavaknál sematikusabb (mind jelentésükben, mind formájukban kevésbé részletezett), de azokhoz hasonlóan tárolt reprezentációk is. Konstrukciós sémát valósít meg például a következő szerkezet, amelyben egy ágens (*Peti*) és egy beneficiens (*Pisti*) között egy páciensre vonatkozó transzfer megy végbe: *Peti Pistinek passzolta/ütötte/dobta/ajándékozta/vette* stb. *a labdát*. A konstrukciós nyelvtani elméletek szerint lexikon és grammatika szerves egységet képez, és

a jelentés szerves része a grammatikának. Ezek a nyelvi modellek azt feltételezik, hogy a konstrukciók hálózata (az ún. konstruktikon) nyelvenként különbözik, ezért a szófajok is nyelvspecifikusak (bár tipológiai értelemben generalizálhatóak, és a nyelvspecifikus kategóriák közötti hasonlóságoknak lehet funkcionális, kognitív és szemantikai magyarázatuk). Konstrukciós nyelvtani szempontból tehát nincs szükség annak a feltételezésére, hogy egy adott nyelvben bizonyos szófajoknak és szófajok közötti viszonyoknak biztosan meg kell jelenniük, a szófajok nem grammatikai kategóriák, hanem funkcionális prototípusok (CROFT 2001: 103), amelyeknek a(z esetleges) konkrét megvalósulása a konstrukciók nyelvspecifikus hálózata felől (egy adott nyelv nyelvtana felől) értelmezhető. A konstrukció fogalma azon a meglátáson alapszik, hogy a nyelvi szerkezetek mind az idiomatikusság egy bizonyos fokán állnak, vagyis ahhoz, hogy egy szó megjelenhessen egy szintaktikai szerkezetben, a szófajiságnál általában sokkal specifikusabb (elsősorban szemantikai) kritériumoknak kell eleget tennie (FILLMORE et al. 1988). Így például a (*passzol*, *ad*, *dob*, *vesz*, *ajándékoz* igékhez hasonlóan) tranzitív *biztat*, *tanul* és *ölel* igék nem kompatibilisek a fentebb említett ditranzitív konstrukcióval, mivel nem tartalmazzák a konstrukció kidolgozásához elengedhetetlen 'transzfer' jelentéskomponenst: **Peti Pistinek biztatta/tanulta/ölelte a labdát*.

Konstrukciós nyelvtani szempontból kiemelendő, hogy az igék a vonzatszerkezetüknek köszönhetően a magyarban nagy mértékben meghatározzák a mondat struktúráját, vagyis erőteljesen asszociálnak bizonyos konstrukciókat. Minthogy a használatukban nem kapcsolódó igék különböző konstrukciós tudást feltételeznek, elképzelhető, hogy ezen igék a főneveknél kevésbé kapcsolódnak egymáshoz a lexikonban, amelyek konstrukciós szempontból alulspecifikáltabbak, vagyis többféle konstrukcióban is megjelenhetnek. A főnév és az ige grammatikai (a konkrét megvalósulásoktól fogalmilag elvonatkoztatott) jelentésének különbségéből fakadhat, hogy korábbi kísérletek szerint a főnévi és az igei jelentés a magyarban eltérően dolgozódik föl (THUMA–PLÉH 1995). Egy olyan kísérletben, amelyben a kísérleti személyeknek kétértelmű szavakra végződő mondatok elolvasása után kellett minél gyorsabban szemantikailag és/vagy szófajilag kapcsolódó szavakat kiolvasniuk, az igei jelentést előfeszítő mondatkontextust követően (lásd *Te is tudod, hogy János mindig nagyot dob*) a személyek a célszavakat (pl. a *hangszer* szót) (átlagosan 42 ms-mal) gyorsabban olvasták ki, mint a főnévi jelentést követően (lásd *Az ütősök között János kedvence a dob*). Továbbá csak a főnévi olvasat esetében volt megfigyelhető lexikai előfeszítő hatás. Ez azt jelenti, hogy a kísérleti személyek például *Az ütősök között János kedvence a dob* mondat után gyorsabban olvasták ki a szemantikailag és szófajilag is kapcsolódó *hangszer* szót a szemantikailag kapcsolódó, de eltérő szófajú *hajít* és a kontroll szóként használt *labda* szóhoz képest. Ugyanakkor a *Te is tudod, hogy János mindig nagyot dob* mondat után a *hajít* szót nem olvasták ki gyorsabban a résztvevők, mint a szemantikailag kapcsolódó *hangszer*, valamint a kontroll *labda* szavakat (THUMA–PLÉH 2001). Ennek akár az is oka lehet, hogy két szinonim, hasonló konstrukcióvonzással rendelkező ige között (pl. *dob*, *hajít*) nagyobb a távolság a mentális lexikonban, mint két szinonim szereplőket kódoló főnév (pl. *dob*, *hangszer*) között.

A konstrukciós nyelvtanok szerint a szójelentésről való tudásunk összefüggésben van a szavak használatára vonatkozó tudásunkkal, és egy szó ismerete magába foglalja azoknak a konstrukcióknak (és a konstrukciókban előforduló más szavaknak) az ismeretét, amelyekben a szó használatos (CROFT–SUTTON 2017). Megkülönböztethetjük ugyanakkor a konstrukciófüggő, grammatikalizáció útján létrejövő lexikai elemek (névelők, névutók stb.) és a fogalmilag részletező, nagyon különböző konstrukciókat is kidolgozni képes szavak (igék, főnevek stb.) csoportjait. GOODMAN és munkatársai (1981) azt találták, hogy létezik egy ún. szintaktikai előfeszítő hatás, amely szerint a szintaktikai viszonyokat jelölő nyelvi elemek (pl. névelő, birtokos névmás) előfeszítik az őket a mondatban potenciálisan követő főneveket (pl. *my* 'enyém' – *turkey* 'pulyka', *our* 'miénk' – *night* 'éjszaka'). Ez a hatás azonban csak akkor volt kimutatható a kísérletben, ha a kérdéses szópárok a szemantikailag asszociálódó szavakkal (pl. *door* 'ajtó' – *window* 'ablak', *cut* 'vág' – *knife* 'kés') nem vegyesen, egymással keveredve jelentek meg, hanem egy olyan külön blokkban, ahol más típusú előfeszítő hatás nem jelentkezett. A vegyes blokkokban ugyanis a szintaktikai előfeszítés hatása nem mutatkozott, csak a szemantikaié, tehát ez utóbbi „elnyomta” az előzőt. Ennek a felismerésnek a jelen kísérlet szempontjából azért is van jelentősége, mert befolyásolta a vizsgálatunk módszertanának kialakítását: az idézett eredmények tanulságai alapján az alább bemutatandó kísérletben a két vizsgált előfeszítési hatást, a kollokációs és szófaji előfeszítést mi is két külön blokkban teszteltük (l. 3.2.).

2.2. Kollokáció. Mint az előző részben már szó volt róla, konstrukciós nyelvészeti megközelítés szerint a(z absztrakt) szófaji azonosságnál fontosabbnak látszanak a használattal összefüggő, végső soron szemantikai háttérű (specifikus, a világról való tudásunkhoz közvetlenül kapcsolódó) kapcsolatok. A szavak használatában az utóbbi évtizedekben ebben a vonatkozásban a kollokáció vizsgálata bír alapvető jelentőséggel. A kollokálódás számszerűsítése és mérése annak a megragadására szolgál, hogy különböző szavak milyen valószínűséggel jelennek meg egymás közelében. Természetesen pusztán az a tény, hogy egy adott korpuszban két szó gyakran megjelenik egymás mellett, még nem jelenti azt, hogy az együttes előfordulásuk valóban magas lenne az adott nyelvben. A legtöbb magyar szó például előfordul valamely nagyon gyakori szó, például egy kötőszó környezetében, ez azonban önmagában még nem bizonyíték a gyakori kollokálódásra, hiszen könnyen előfordulhat, hogy a kötőszó a saját gyakoriságához képest szemlélve alig fordul elő a másik szó környezetében. A kollokálódás (azaz az együttjárás valószínűsége) mérésének több eszköze is van (HUNSTON 2002). A kölcsönös információ (MI) mérése az együttes előfordulás véletlenszerűségét kívánja kiküszöbölni, és a szavak gyakorisága alapján várható (jósolt) és megvalósult előfordulási szám összevetésére épül. A T-Score érték pedig azt mutatja meg, hogy egy-egy szó csomópontként (node) milyen szavakat vonz (háttérbe szorítva azt az információt, hogy az illető szót mely más szavak hívják elő). A logDice (RYCHLÝ 2008) mérőszám a csomópont (azaz a célszó) és a kollokált szó gyakoriságán, illetve a két szó együttes előfordulásának gyakoriságán alapszik. A jelen tanulmányban ez utóbbi, nagy korpuszok esetében preferált mérőszámot alkalmaztuk a vizsgált szavak kollokációs összefüggéseinek számszerűsítésére,

amely szerint erős és gyenge kollokálódást állapítottunk meg úgy, hogy az utóbbi csoportba azok a szópárok tartoztak, ahol lényegében a kollokációs kapcsolat hiányát mutatta a mérőszám (l. 3.1.).

Ami a kollokáció pszicholingvisztikai vonatkozásait illeti, számos modell a nyelvi képesség lényegi elemeként hivatkozik rá. Eszerint a nyelvhasználók tömbösödött egységeket tárolnak a hosszútávú memóriájukban, vagyis elsajátítják annak a tapasztalatát, hogy két szó milyen valószínűséggel jelenik meg együtt, amiben fontos szerepe van az egymás közelségében előforduló szavak közötti asszociatív kapcsolatoknak (ELLIS 2001). A kollokáció fontosságát hangsúlyozó, de az elemektől független szabályszerűségek szerepét is elfogadó nézetek szerint a nyelvhasználók szükség esetén képesek felbontani a több szóból álló tömböket, de alapvetően egységként (szabályok alkalmazása nélkül) hívják őket elő (ELLIS 2001). A kollokációnak ennél központibb szerepet tulajdonítanak azok a modellek, amelyek szerint minden szó előfeszíti a magas valószínűséggel kollokált (gyakori) szókörnyezetét, és a grammatika nem más, mint szókombinációk újra és újra visszatérő mintázata (HOEY 2005). A vegyes modellek a nyelvben egy szabályelvű és egy kollokáció elvű mechanizmust is feltételeznek. A produkcióban a szavakra vonatkozóan például megkülönböztethető az ún. nyitott választás elve (open-choice principle) és az idióma elve (SINCLAIR 1991). A nyitott választás elve szerint a beszélő egyéni döntése, hogy egy grammatikai szerkezetben milyen szavakat használ, például az ún. *okozott mozgás* konstrukció (GOLDBERG 1995) kidolgozásakor (l. *leverte a poharakat az asztalról, lefújta a szél a virágokat a teraszról*) a beszélő dönthet új, elsőre akár szokatlannak tűnő szavak használata mellett (KALIVODA 2021; PALÁGYI 2022), például: *letáncolná a fákról a virágot, leimádkozzák a bugyit a nőkről* (forrás: MNSz2, ORAVECZ et al. 2014). Az idióma elve szerint azonban a szavakat a beszélők általában a szokásos konstrukciós és lexikai környezetükben használják, és elsődlegesen nem is szavaknak, hanem a szavak társulásainak van jelentésük (melyekből a szójelentéseket utólag lehet elvonatkoztatni). A két elv elkülönítése történhet modularista alapon, a grammatikai szabályok és a memória szerepének éles megkülönböztetésével (vö. PAWLEY–SYDER 1983; SKEHAN 1998). Bizonyos modularista elképzelések szerint alapbeállításnak az idióma elvű, kollokációra építő működés tekinthető (WRAY 2002). Az előző részben említett konstrukciós nyelvtani megközelítések a két elv érvényességét a nyelvi működésben nem tagadják, azonban egységben szemlélik azokat (vö. SINCLAIR 1996: 81), minthogy az újszerű, szabályos, kreatív nyelvi produktumokat is a szavaknál magasabb általánossági szinten lévő, de a memóriában tárolt konstrukciós sémák megvalósításaként (esetleg kiterjesztéseként) értelmezik.

A jelen tanulmányban bemutatott kísérletben vizsgált kollokációs kapcsolat az idióma elv, a szófajiság pedig a nyitott választás elvének érvényességével hozható összefüggésbe (minthogy léteznek olyan konstrukciók, amelyeknek sematikus részeit az adott szófajhoz tartozó lexikai elemek többsége kitöltheti), így – legalábbis látszólag – alapvetően eltérő viszonyokról van szó a két esetben. Ugyanakkor konstrukciós szempontból a két elv ugyanazon kontinuumon helyezhető el, így ez az eltérés vélhetően ugyancsak nem alapjaiban eltérő nyelvi működéseket, illetve összefüggéseket feltételez a szavak egyes (kollokációs

vagy szófaji alapú) csoportjai között: bár a nyelvi szerkezetek esetében fennáll a szabad választás lehetősége, azaz egy adott szintaktikai szerkezetben különböző szavak is előfordulhatnak, ehhez a szavaknak a szófajiságnál általában specifikusabb (elsősorban szemantikai) kritériumoknak kell megfelelniük (vö. FILLMORE et al. 1988).

A szavak közötti szemantikai előfeszítés meglétét már több kutatásnak sikerült kimutatnia (FISCHLER–GOODMAN 1978, MEYER et al. 1975), amit hagyományosan az aktivációterjedés elméletével magyaráznak (ANDERSON 1996/2013). Az aktivációterjedés elmélete szerint a szótaláláskor számos másik szó aktiválódik az adott beszédeseményben a később kimondott szó mellett (AITCHISON 2003: 84–101). Bizonyos mértékig a hasonló vagy éppen ellentétes jelentésű szavak, az alá- vagy fölérendelt kategóriákat szimbolizáló szavak, a vele gyakran együtt előforduló más szavak és kifejezések, a tudáskeret, a forгатókönyv elemei, tehát egy egész fogalmi hálózat aktiválódik, amelyből a végül kimondott szó hozzáférhetővé válik, és egyúttal természetesen hozzáférhetővé teszi magát az aktiválódó hálózatot is. (A *desszert* szó például az ÉTTEREM és/vagy az ÉTKEZÉS forгатókönyvéhez biztosít csomópontként referenciapontot, tehát az ezzel összefüggő szavakat is aktiválja.) Módszertani szempontból kiemelendő, hogy a pszicholingvisztikai kutatások alapját képező nyelvi anyagokban a szavak közötti szemantikai összefüggéseket általában és hagyományosan introspekcióval, a saját nyelvi produkció megfigyelésével (szerencsésebb esetben szinonímiára vonatkozó szótári hivatkozással) állapítják meg.

A disztribúciós szemantikai megközelítés szerint a szavak tartalma nagyrészt jellemezhető a szó kontextuális reprezentációjával, amely azon nyelvi kontextusok absztrakciójaként határozható meg, amelyekben a szóval találkozunk (MILLER–CHARLES 1991). A szemantikai viszony (az azonos szemantikai mezőbe való tartozás) a disztribúciós szemantika elvei szerint az azonos szófajú szavak esetében lényegében azonos a kollokáció alapú összefüggéssel. Disztribúciós szemantikai alapvetés ugyanis, hogy azok a szavak, amelyek hasonló kontextusban jelennek meg, hasonlóak a jelentésükben (erre a belátásra épülnek például a napjainkban népszerű nagy nyelvmodellek). A kollokáció szempontjából ez azt jelenti, hogy a hasonló kollokáltakkal rendelkező szavaknak hasonló a jelentésük. Ez a szemantikai összefüggés azonban nézetünk szerint nem azonos a szinonimitással vagy felcserélhetőséggel.¹ A szavak grammatikai viselkedésének és kollokációjának feltérképezésére szolgáló NoSketchEngine korpuszlekérdező platformon a Magyar Webkorpuszt vizsgálva például kimutatható a *kutya* és a *gyerek* szó kváziszinonimitása. Ennek az az oka, hogy tárgyi viszonyban mindkét szó erősen kollokálódik az *etet* igével, alanyi viszonyban az *eszik* és a *felne* igével, predikatív melléknévként pedig gyakran járnak együtt a *kíváncsi*, *okos*, *egészséges*, *boldog*, *beteg* melléknévekkel. Mindkét főnévnek gyakori jelzői bővítményei továbbá az *éves*, *korú*, *kicsi* melléknévek, a *szomszéd* és az *ismerős* szavak birtokos jelzőként jelennek meg előttük, ők maguk pedig a *viselkedés*, *élet* és *nevelés* főnevek birtokos

¹ Az emberi lexikális memória pszicholingvisztikai elméletei által ihletett WordNet (FELLBAUM 1998) lexikai referenciarendszerben például a hiperonímia játszik meghatározó szerepet.

jelzői. A szó kollokációs jellemzőit és jelentését ugyanakkor nem azonosítjuk egymással, ezért a továbbiakban is kollokáció alapú összefüggésről beszélünk a használatukban megegyező szavak kapcsán, nem pedig szemantikai összefüggésről (amely nézetünk szerint „csupán” motiválja a kollokáció alapú összefüggést, például azt, hogy a *kutya* és a *gyerek* szó hasonló környezetekben jelenik meg).²

2.3. Hipotézisek. A jelen tanulmányban ismertetett kísérletünkben a szófajiságot és a kollokációt mentális tényezőként közelítettük meg. Azt vizsgáltuk, hogy a kollokációs alapon és szófajilag is összefüggő szavak jobban előfeszítik-e egymást, mint a kollokációs alapon nem összefüggő és szófajilag is különböző szavak, illetve azt is, hogy a két hatás egymáshoz képest hogyan alakul. Az előfeszítést egy olyan lexikai döntési feladattal vizsgáltuk, amelyben az előfeszítés erejét a célszavak létező vagy álszavakként történő felismerésének gyorsaságával számszerűsítettük. A döntési helyzetben (valódi szó vagy álszó) mért reakcióidő alapján vizsgáltuk a különböző valószínűséggel kollokálódó szavak előfeszítő erejét három szinten (erős, gyenge, nincs kollokálódás), a szófajiság előfeszítő erejét (azonos vagy eltérő szófajú szavak), illetve ezek lehetséges kombinációit. Az elméleti háttérfeltevésekből és a korábbi kutatási eredményekből következően az alábbi hipotéziseket fogalmaztuk meg a jelen vizsgálatra vonatkozóan:

1. A kollokációs alapon erősen összefüggő szavak (pl. *zöltség-eszik*, *termel-gazda*), az azonos szófajú szavak (pl. *siker-jármű*, *cserél-eszik*), illetve a kollokációs alapon erősen összefüggő és azonos szófajú szavak (pl. *szoba-asztal*, *beszél-hallgat*) erőteljesebben előfeszítik egymást (így alacsonyabb reakcióidő-értékeket találunk ezeknek a szópároknak ez esetében), mint a kollokációs alapon gyengén és szófajilag nem összefüggő szavak (pl. *szerel-erdő*, *sodrás-olvas*). Ezen belül az erősebb kollokációs és szófaji kapcsolattal rendelkező szavak (pl. *szoba-asztal*, *beszél-hallgat*) erőteljesebben elő is feszítik egymást, mint a csak erősebb kollokációs kapcsolattal rendelkezők (pl. *zöltség-eszik*, *termel-gazda*), amelyek viszont erőteljesebben előfeszítik egymást a csak szófaji kapcsolattal rendelkező szavakhoz képest (pl. *siker-jármű*, *cserél-eszik*). Az álszavak (*föröl*, *kölész*) feldolgozása és elutasítása pedig lassabb, mint a valódi szavaké (lásd JANSEN et al. 2017).
2. A főnevek és az igék előfeszítő hatása eltér úgy, hogy
 - a) a kollokációs alapú előfeszítés a főnevek (pl. *szoba-asztal*) esetében erősebb, mint az igék (pl. *beszél-hallgat*) esetében, tehát gyorsabb reakcióidőkhöz vezet a lexikális döntési feladatban, továbbá
 - b) a kollokáció szempontjából gyengén kapcsolódó főneveknek (pl. *siker-jármű*) nagyobb az előfeszítő hatása egymásra (szófaji alapon), mint a kollokáció szempontjából gyengén kapcsolódó igéké (pl. *cserél-eszik*).

² A kollokáció alapú összefüggések háttérben ráadásul különböző szemantikai összefüggések is állhatnak. A mintában például az azonos szófajú szavak esetében a kollokálódást az azonos fogalmi kerethez való tartozás (*művész*, *színház*), a hiperonímia (*kocsi*, *jármű*) és a szinonímia (*eszik*, *fogyaszt*) egyaránt motiválhatja.

3. Anyag, módszer, kísérleti személyek

3.1. Anyag. A kutatás hipotéziseinek vizsgálatára egy ún. lexikai döntési kísérletet terveztünk. A résztvevők feladata az volt, hogy a képernyőn megjelenő szavakról döntsék el, létező magyar szavakat látnak-e. A kísérlet anyagát képező szavak listáját korpuszalapú elemzés segítségével alakítottuk ki. A vizsgálatban résztvevők a lexikai döntést a célszóhoz (target) kapcsolódóan hozták meg, amelyet minden esetben megelőzött egy előfeszítő szó (prime). A kutatás céljának megfelelően olyan prime-target szópárok összeállítására volt szükség, amelyek a) kollokáció szempontjából erősebben kapcsolódnak egymáshoz (tehát nagyobb valószínűséggel kollokálódnak), b) gyengébben kapcsolódnak egymáshoz (tehát alacsony valószínűséggel kollokálódnak), illetve c) nem kapcsolódnak egymáshoz (az együttjárásnak nincs valószínűsége, mert a célszó nem valódi szó, hanem ún. álszó, melynek gyakorisága ugyancsak nulla). Az álszók a magyar hangsor-építési szabályszerűségeknek megfelelő hangalakú és a valódi szavakkal azonos szótagszámú (két szótagú) szavak voltak. A szópárokat aszerint is variáltuk, hogy annak tagjai egyeztek-e szófajukban. Vizsgáltunk főnévi és igei célszavakat azonos vagy eltérő szófajú (ugyancsak főnévi vagy igei) prime szavakkal. Az álszavak egyértelműen beazonosíthatóan szintén főnévi vagy igei szófajúak voltak, ugyanis gyakori és/vagy produktív főnév-, illetve igeképzőket tartalmaztak (pl. *pátság, hülget, szüvény*). A kísérletben tehát az 1. táblázatban bemutatott feltételeket vizsgáltuk. A továbbiakban a kísérleti feltételek kialakítását, illetve a szópárok összeállításának menetét ismertetjük.

1. táblázat

Példák a kísérletben használt anyagra (zárójelben a kondícióknak az adatok ábrázolása során használt kódjával) a kísérleti feltételek szerint

		Az előfeszítő szó (prime)	
Célszó (target)	A kollokálódás erőssége	Főnév (n)	Ige (v)
Főnév (n)	erős (a)	(a_nn) pl. <i>kocsi-jármű</i>	(a_vn) pl. <i>idéz-emlék</i>
	gyenge (b)	(b_nn) pl. <i>csillag-gazda</i>	(b_vn) pl. <i>nevel-asztal</i>
	nincs (c) (a target álszó)	(c_nn) pl. <i>holttest-kölész</i>	(c_vn) pl. <i>fogyaszt-szüvény</i>
Ige (v)	erős (a)	(a_nv) pl. <i>tempó-halad</i>	(a_vv) pl. <i>fogyaszt-eszik</i>
	gyenge (b)	(b_nv) pl. <i>csuklya-hallgat</i>	(b_vv) pl. <i>romlik-figyel</i>
	nincs (c) (a target álszó)	(c_nv) pl. <i>hála-tepel</i>	(c_vv) pl. <i>vádol-mörcög</i>

Mint a 2.2. pontban már utaltunk rá, a kollokáció alapú összefüggés mást jelent az azonos és a különböző szófajú szavak esetében. Különböző szófajú szavak esetében a kollokálódás erőssége a használati összefüggést (tehát az együttes előfordulást) mutatja – ez a kollokálódás a gyakori együttes előfordulásra épül. Azonos szófajú szavak esetében viszont nem az a kérdés, hogy két szó gyakran használatos-e együtt, hanem az, hogy ugyanazon lexikai és szintaktikai környezetben használatos-e (akár úgy, hogy együtt valójában ritkán fordulnak elő) – ez a kollokálódás a hasonló lexikogrammatikai környezetben való előfordulásra épül.

Az azonos szófajú szavak esetében a használati összefüggést a Sketch Engine (KILGARRIFF et al. 2014) korpuszelemző Thesaurus funkciójának segítségével alapítottuk meg. A Thesaurus funkció egy szó (szóalakjainak) használatbeli hasonlóságának meghatározásához az azonos szófajú szavak szómintázatait hasonlítja össze, és a szavakat aszerint rangsorolja, hogy milyen magas a közös kollokáltak aránya. Két azonos szófajú szót akkor tekintettünk hasonló kollokációs mintázatúnak, gyakran kollokálódónak, ha a Magyar Webkorpusz 2.0 alapján az elemző szerint hasonló (szintaktikai és lexikai) környezetben fordulnak elő. A hasonlóság kritériuma az volt, hogy Y szó (a célszó) X szó (a prime) Thesaurus listájának első tíz helyének valamelyikén foglaljon helyet, például: *kocsi* → *jármű*, *fogyaszt* → *eszik* (vö. 1. táblázat).

Két azonos szófajú szót akkor tekintettük kollokáció alapján gyengén kollokálódóknak, ha nem jelentek meg egymás Thesaurus listájában, továbbá az MNSz2-ben³ is mínusz értékben volt a kollokációjuk logDice értéke, vagyis nem találtunk bizonyítékot a kollokálódásukra, például: *siker* → *jármű*, *cserél* → *eszik*.

A nem azonos szófajú szavak esetében az összefüggést ugyancsak a NoSketch Engine platformba integrált Magyar Webkorpusz 2.0-ban és az MNSz2-ben végzett vizsgálatra alapoztuk. A nem azonos szófajú szavak esetében a magas használati összefüggés megállapításának két feltétele volt. Egyfelől az, hogy a NoSketch Engine felületen kimutatható legyen, hogy a két szó jellemzően előfordul együtt valamilyen szintaktikai viszonyban. A hasonlóság kritériuma egyfelől az volt, hogy Y szó (célszó) X szó (a prime) Word Sketch listájában igék esetében az alanyi vagy tárgyi listájának, főnevek esetében pedig az igei listájának első tíz helye egyikén foglaljon helyet. Másfelől az, hogy az MNSz2-ben 5,0 és 6,5 közötti logDice értéket mutassanak a kérdéses szavak megfelelő (vagyis a lexikai döntés feladatban kialakított) irányban 3 token távolságon belül, például: *parkol* → *jármű*, *zöldség* → *eszik*. A kollokáció-jelöltek küszöbértékének (alsó határértékének) tehát az 5,0 logDice értéket határoztuk meg (FRANKENBERG-GARCIA et al. 2019 alapján), felső határértékének pedig a 6,5-öt. Az 5,0 és 6,5 közötti érték gyenge közepes kollokálódásnak tekinthető (KIM-OH 2020 alapján). A relative alacsony felső határértékkel az idioma elv esetlegesen túl erős hatását szerettük volna kiküszöbölni (l. 2.2.). Nézetünk szerint ugyanis például a *rész* lemmával jobbra kollokált *vesz* ige erős 10,867 logDice értéke (3 token távolságban) éppen

³ A számos duplumot tartalmazó személyes alkorpusz *személyes-közösségi* és *személyes-fórum* részét az MNSz ajánlásának megfelelően kizártuk a vizsgálatból. Az MNSz2-ben történő vizsgálatok így a *beszélmyelvi*, *hivatalos*, *sajtó*, *tudományos*, *hivatalos* és *szépirodalmi* alkorpuszokra korlátozódtak.

arra utal, hogy a két szó tömbösödött egységként hívódik elő (*részt vesz*), így nem valószínű, hogy ezek a szavak önmagukban előfeszítenék egymást.

Két nem azonos szófajú szót akkor tekintettünk ritkán kollokálódóaknak, ha a Sketch Engine segítségével a Magyar Webkorpusz 2.0-ban nem volt kimutatható, hogy a két szó jellemzően előfordulna együtt bármilyen szintaktikai viszonyban, egyúttal az MNSz2-ben mínusz logDice kollokációs értéket mutattak a megfelelő irányban 3 token távolságon belül. Ez ugyanis azt jelenti, hogy nem találtunk bizonyítékot a kollokálódásukra, például: *roppan – jármű, módszer – eszik*.

(Folytatjuk.)

PALÁGYI LÁSZLÓ

ELTE Eötvös Loránd Tudományegyetem

DEME ANDREA

ELTE Eötvös Loránd Tudományegyetem

ELTE Nyelvtudományi Kutatóközpont

MTA–HUN-REN NYTK „Lendület” Neurofonetikai Kutatócsoport