

A *bealszik* mintázatainak korpusznyelvészeti feltárása – egy módszer lehetőségeiről

A dolgozat a korpusznyelvészet területén belül egy előtanulmány eredményeit mutatja be: a mintázatnyelvtan (Pattern Grammar, I. Hunston–Francis 2000) magyar nyelvre történő alkalmazásához kidolgozott módszertani eljárásokat ismerteti, illetve bizonyos alapvető mintázatok leírását és értelmezését tűzi ki célul. A mintázatok listázása és részletes elemzése megalapozhatja egy, a magyar mint idegen nyelv tanításában használható oktatási segédanyag létrehozását. A dolgozat a segédanyag kidolgozását a mintázatnyelvtan alapvetéseinek ismertetésével, a mintázat fogalmának definiálásával, valamint egy konzekvens metanyelv és egy módszertani bázis prezentálásával készíti elő.

Az előtanulmány második fele a *bealszik* példáján keresztül mutatja be az alkalmazni kívánt módszertant: az adott szóalak mintázatjelöltjeit olyan módszerrel tárja fel, amely több szinten is generalizál, ugyanakkor igyekszik minimalizálni az előzetesen elfogadott előfeltevések számát.

Kulcsszavak: mintázatnyelvtan, korpusznyelvészet, korpuszalapú kutatás, korpuszvezérelt kutatás, oktatási segédanyag, szubmerzív jelentés

1. Bevezetés: a kutatás megalapozása

A dolgozatom egy olyan előtanulmányt mutat be, amely egy, a magyar nyelvre jellemző tipikus mintázatokat összegző munka előkészítésében érdekelt. A mintázatok számbavételének célja, hogy megalapozzák egy oktatási segédanyag kidolgozását, amely a magyart idegen nyelvként tanulók, illetve a magyart származási nyelvként beszélők számára nyújthat segítséget.

Az előtanulmányom először felvázolja a mintázatnyelvtan (Pattern Grammar, I. Hunston–Francis 2000) főbb alapvetéseit és eszköztárát, meghatározza a mintázat fogalmát, kialakít egy, a magyar nyelvre is alkalmazható metanyelvet, majd állást foglal a mintázatok korpuszadatokból való kinyerésének egy lehetséges módja mellett. A kutatás nyelvelméleti és korpusznyelvészeti előfeltevéseinek rögzítését és az azokra adott reflexiót tehát a választott módszertani keretek korpuszból nyert mintázatokkal való szembesítése egészíti ki. A dolgozat ehhez egy olyan *be* igekötővel képzett igekötős igealakot vizsgál (*bealszik*), amely feltehetőleg a rendszerváltás után jelent meg, és neologizmusként is elemezhető (I. Ladányi 2007: 258–283). A választott igealak elemzésével a korpuszadatokat megfigyelve, a korpuszban reprezentált aktuális nyelvhasználatból kiindulva azonosítom az adott kifejezés formai és szemantikai mintázatait. A korpusznyelvészeti vizsgálat több szinten generalizálja az eredményeit (vö. Simon 2018: 15–16), és a generalizálás során felhasználja a funkcionális nyelvelmélet terminusait is.¹

¹ A kutatás nyelvtanvezérelt abban az értelemben, hogy – a mentális reprezentációs struktúrák feltárására irányuló nyelvtanalapú kutatásokkal szemben – nem a jelentésekhez rendelt mentális reprezentációs struk-

A következőkben ismertetett kutatás egyben egy korábbi kutatás közvetlen folytatása, amely nyelvtanalapú és korpuszalapú megközelítést alkalmazva tárta fel a *bealszik*, a *beijed* és a *behal*/nyelvi elemek szemantikai reprezentációját (l. Kopcsák 2018, 2019). A jelen dolgozat – egyetlen igealakra fókuszálva – közel ugyanazt a kérdéskört teszi elemzés tárgyává, mint az előzménykutatásai, ehhez azonban eltérő módszertani megfontolásokat vesz alapul. A mentális reprezentációk modellálása és kis számú adat kvalitatív elemzése helyett nagy számú korpuszadatból indul ki, induktívan, lépésről lépésre generalizál, és minimalizálni igyekszik az előzetesen elfogadott elméleti előfeltevéseit.² A vizsgálat során nem számolom fel, azonban igyekszem visszaszorítani az adatelemzés előtt alkalmazott kutatói intuícióimat (ehhez l. még a 2.1. alpontot).

A dolgozat felépítése a következő: A 2. szerkezeti egység a korpusznyelvészeti diskurzus néhány főbb megállapítását mutatja be a korpuszalapú és korpuszvezérelt megközelítésmódok, illetve a mintázatnyelvtan alapelveinek rövid áttekintésén keresztül. A 3. alpont ismerteti a vizsgálat korpuszanyagát és módszerét, mindezzel pedig az előkészítendő kutatás elméletét, módszertanát és metanyelvét dolgozza ki. A 4. szerkezeti egység a *bealszik* igealak korpuszadatainak bemutatásával és elemzésével jut el a *bealszik* mintázatainak feltárásához. Az 5. rész a konklúziókat foglalja össze.

2. Részletek a korpusznyelvészet megközelítésmódjaiból

A korpusznyelvészet a nyelvtudomány prosperáló ága (a korpusznyelvészetről általában lásd: Szirmai 2005). Azzal, hogy nagyszámú adatot tesz szisztematikusan és gyorsan feldolgozhatóvá (Sinclair 1991: 4), bevonja a nyelvtudományt a digitális bölcsészeti diskurzusba, rámutat a számítógép, valamint a statisztikai és kvantitatív módszerek alkalmazásainak előnyére, illetve ami a legfontosabb, egyben új evidenciát és új adattípust szolgáltat a nyelvtudomány számára (utóbbihoz l. Sinclair 1991: 2, Hunston–Francis 2000: 250).

A következőkben a korpusznyelvészet diskurzusából emelek ki néhány, a kutatás szempontjából különösen fontos szegmentumot. A 2.1. szerkezeti egység a korpuszalapúság és a korpuszvezéreltség kérdéskörével foglalkozik, szorosan kapcsolódva a forma és a jelentés leírásának kérdéséhez is. A 2.2. szerkezeti egység a dolgozat fő kiindulópontját, a mintázatnyelvtan alapfelvetéseit vázolja fel.

2.1. Korpuszalapú és korpuszvezérelt megközelítésmód

Tognini-Bonelli (2001) kétféle módszertant különít el, amellyel a korpuszadatok feldolgozhatóvá válnak a nyelvészet számára: a korpuszalapú és a korpuszvezérelt megközelítést. A kor-

túrák feltárására, hanem a nyelvi szerkezetek grammatikai (és a grammatikaitól el nem választható szemantikai) jellegzetességeinek a szisztematikus leírására vállalkozik (Simon 2018: 8; a „grammatikai” mint kategória a kontinuumelvnek megfelelően kezelendő, l. Simon 2018: 3, vö. Langacker 1987: 3). A nyelvtanalapú és nyelvtanvezérelt megközelítéshez l. még: Simon 2016: 179–191.

² Elhagyom például azt a korábban az általam végzett kutatásokban is alkalmazott gyakorlatot, amely előfeltételezte az aspektust mint önálló mentális reprezentációt és mint egymással összefüggő kategóriákra osztható szemantikai teret, majd az így előfeltételezett jelentéstani kategóriákat vetítette rá a korpuszadatokra.

puszalapú módszertan szerint a korpuszadatok egy már kidolgozott elméletet validálnak, szemléltetnek vagy magyaráznak. Az elmélet felállítása ebben a megközelítésben nem az adatok feldolgozásán alapul: a korpusz egy már létező elmélet tesztelésére, finomhangolására, illetve kvalitatív evidenciák felmutatására szolgál. Egy korpuszalapú kutatás korpuszelemzése tipikusan csak kisebb módosításokat eredményezhet az elméleti előfeltevéseken, de az adatok nem kérdőjelezik meg magát az előzetesen posztulált elméletet. Ilyen értelemben a korpuszalapú megközelítés nem kötelezi el magát szigorúan a korpuszadatok mellett (Tognini-Bonelli 2001: 65–80).³ Jelen kutatás előzménye (Kopcsák 2018, 2019) is korpuszalapú vizsgálatot folytatott le: egy előzetesen felállított aspektusmodell validálását végezte el önkényesen kiválasztott korpuszadatok kvalitatív elemzésével.

A korpuszvezérelt megközelítésmód nem egy előzetesen elfogadott elméletből kiindulva kívánja magyarázni az adatokat, hanem épp ellenkezőleg: visszaszorítja az elméleti előfeltevéseket, és a korpuszadatok megfigyeléséből kiindulva tesz elméleti állításokat. Egy korpuszvezérelt vizsgálat célja, hogy nyelvi egységek gyakori disztribúciójából és visszatérő nyelvi mintákból következtessen nyelvészeti kategóriákra. Az adatok generalizálásának alapja, hogy a korpusz vertikális „tengelyén” (tehát az egymás alá rendeződő konkordanciasorokon) olyan nyelvhasználati események figyelhetők meg,⁴ amelyek „formalizálhatók” és csoportosíthatók.⁵ A korpuszvezérelt kutatások szigorúan és szisztematikusan köteleződnek el az adatok mellett. A korpuszvezérelt megközelítés az adatok megfigyelésének fázisában az elméleti előfeltevésekkel együtt a kutatói intuíciót is visszaszorítja: a korpuszbeli evidenciákat a lehető legkevesebb intuícióval értelmezi (Tognini-Bonelli 2001: 84–91). A kutatói intuíció és a korpuszevidencia mint az adatok validitását biztosító eljárás között jelentős és rendszerszerű különbség van. A nyelvről való intuíció túlságosan specifikus ahhoz, hogy megközelítse a korpusz kvantifikálható adatsorait (Sinclair 1991: 4). A korpuszvezérelt megközelítés mindezek ellenére nem értékeli le az intuíciót: a professzionális intuíció mindenképpen szükséges marad a már megfigyelt és kinyert korpuszevidenciák értékeléséhez (Tognini-Bonelli 2001: 185, vö. még Majoros 2013).

Az előző bekezdés nyomán kérdésként merül fel, hogy lemond-e mindennemű elméletről egy korpuszvezérelt kutatás. A korpuszvezérelt megközelítés a korpuszadatok értelmezésekor igyekszik minél inkább visszaszorítani az intuíciót és az a priori elméleti állítások elfogadását, szem előtt tartja azonban, hogy elméletfüggetlen megfigyelés nem létezik (Hunston–Francis 2000: 18, vö. még Halliday 1994: xii).⁶ A korpuszvezérelt felfogás tehát nem utasítja el az elméletet, sokkal inkább állandó felülvizsgálat alá helyezi (Tognini-Bonelli 2001: 179). Az elmélet nem kiindulópont, amelyből a megfigyelések táplálkoznak, hanem olyan köztes állomás, amelyet a korpuszadatok módszeres vizsgálata alakít ki.

Bizonyos értelemben tekinthető a korpuszvezérelt megközelítés sajátos elméleti alapvetésének, hogy elfogadja és a gyakorlati korpuszelemzésre is kiterjeszti Firth (1957: 11) sokat

³ A korpuszalapú kutatások különösen kedveznek például a regisztervizsgálatoknak (Biber 2010: 163).

⁴ A nyelvhasználati események ebben az esetben részlegesen dekontextualizált megnyilatkozásokat vagy azok részleteit jelentik.

⁵ Ilyen csoportosítást vagy „formalizálást” hajt végre Hunston–Francis (2000) is, amikor egy-egy kulcsszóra keresve manuálisan szűri végig az egymás alá rendeződő konkordanciasorokat, hogy megfigyelje azokat a mintázatokat, amelyekben a kulcsszó is részt vesz.

⁶ Hunston–Francis (2000: 250) megjegyzi, hogy az elméleti állítások bizonyos értelemben mégis egy-egy módszer által konstruálódnak meg.

idézett állítását, amely szerint minden szó a saját környezetében ismerhető meg: Firth kontextuális jelentésemelvéte⁷ nyomán a horizontális konkordanciasorok vertikális elrendeződéséből indul ki. A konkordanciasorok nyelvhasználatának felszíni szerkezetét elemzi, majd a felszíni szerkezetek közötti jelentős különbségeket és hasonlóságokat generalizálja (Hunston–Francis 2000: 177). Fontos hangsúlyozni: a felszíni szerkezet nem egy szabályvezérelt nyelvtanra utal (Hunston–Francis 2000: 256), és nem is az emberi önreferencia világon kívüli pozícióját, „felszínen áthatoló” és „mélybe leásó” gyakorlatát feltételezi (a felvázolt episztemológiai szerkezet általános bemutatásához l. Gumbrecht 2010: 25–30). A korpuszvezérelt megközelítés a formai oldal vizsgálatával közvetlenül jut el a jelentés feltáráshoz,⁸ ugyanis a forma és a jelentés között nem tesz különbséget (Sinclair 1991: 7). Jelentés és forma ebben a megközelítésmódban ugyanannak a jelenségnek a különböző aspektusait takarja – ez a jelenség pedig az aktuális nyelvhasználat. A forma vizsgálata a jelentés feltárást is magában rejti (Tognini-Bonelli 2001: 99).⁹ A korpuszvezérelt megközelítés legradikálisabb formája semmiféle lemmatizálást nem fogad el, ugyanis a különböző formák különböző mintázatokban vehetnek részt, és eltérő jelentésekkel választhatnak ki (Sinclair 1991: 8, Tognini-Bonelli 2001: 96, vö. még Simon 2018: 6).

2.2. A mintázatnyelvtanról

A mintázatnyelvtan egy adott nyelv mintázatait tárja fel, és eredetileg az angol nyelvre vonatkoztatva jött létre. Az angol nyelv mintázatait a COBUILD projekt részeként publikálták (l. Francis–Hunston–Manning 1996, Francis–Hunston–Manning 1998; vö. még Szirmai 2005: 64–65);¹⁰ a mintázatok azonosításakor lefolytatott kutatásokat és a vizsgálatok során posztulált állításokat Hunston és Francis (2000) foglalja össze.

Egy vizsgált szó mintázata olyan további szavakból vagy struktúrákból áll, amelyek szavak vagy struktúrák adott csoportjából választódnak ki, viszonylag gyakran fordulnak elő a vizsgált

⁷ Ennek bemutatásához l. Tognini-Bonelli (2001: 157–164). Érdemes megjegyezni, hogy bár a korpusznyelvészet számára Firth kutatásai alapvető jelentőségűek, Firth önmaga valószínűleg nem támogatta volna a számítógép nyelvészeti elemzésre való használatát (Tognini-Bonelli 2001: 157).

⁸ Ebben az értelemben a *felszíni szerkezet* megnevezést elnevezés, ugyanis a *felszín* mint terminus már-már automatikusan hívja elő a mélység, valamint a mélységhez rendelődő magasabb rendűség képzetét, miközben önmaga dehonesztálását végzi el. Az előtanulmányom nem vállalkozhat hosszabb metatudományos kitérőre, röviden azonban érdemes szót ejteni arról, hogy a felszíni szerkezeteknek Hunston és Francis (2000) munkáiban alkalmazott fogalma hogyan viszonyul a felszín-mélység képzeinek a generatív grammatikában szokásos értelmezéséhez. Míg a Chomsky-féle generatív grammatika deduktívan, a mélyszerkezetből kiindulva, algoritmizáló műveletek segítségével jut el a felszíni szerkezethez, addig a mintázatnyelvtan esetében a felszíni szerkezet egyszerre tartalmazza a felszín és a mélység képzetét: a különböző szintű generalizálások az egymás alá rendeződő konkordanciasorokból olvashatók ki, a mintázatok pedig a maguk konkrétságában figyelhetők meg úgy, hogy közben sem a nyers adatokhoz, sem a generalizált mintázatokhoz nem rendelődik pozitív vagy negatív előjel.

⁹ Ehhez szorosan kapcsolódó állítás, hogy a szintaxis, a lexikon és a szemantika nem választható el egymástól (Sinclair 1991: 108). Grammatika, lexikon és szemantika nem kontinuumként írható le (mint például a funkcionális kognitív nyelvészet felfogásában, vö. Langacker 1987: 4), hanem ezek egyszerre aktiválódnak és együtt választódnak ki (Hunston–Francis 2000: 251, vö. még Francis 1995).

¹⁰ Az eredeti publikációk részben módosított változata online is elérhető: <https://grammar.collinsdictionary.com/grammar-pattern/> (Utolsó elérés: 2020. aug. 26.). A dolgozatban használt angol nyelvű példák a linkelt oldalról származnak.

szóval, és gyakran asszociálódnak hozzá úgy, hogy az egymáshoz asszociálódás világosan megragadható, önálló jelentéssel bír (Hunston–Francis 2000: 37–45). Ilyen mintázat például a **V n** (V = a vizsgált igei csoport; n = főnévi csoport) mintázat (a mintázatok jelölését l. alább). A **V n** mintázat néhány lehetséges példája: *They broke the window* 'betörték az ablakot', *my husband is a doctor* 'a férjem orvos'.

A mintázatok jelöléséhez Hunston–Francis (2000) egy olyan sajátos metanyelvet használ, amelynek célja, hogy a laikusok számára is jól érthető legyen.¹¹ Ebből következően a mintázatokat jelölő metanyelv rugalmas (bármilyen frazeológiai egységet képes kódolni), transzparens (kis eszköztárat módosít, csak a felszíni szerkezetet kódolja, nem alkalmaz szintaktikai absztrakciót) és konzisztens (csak egyféle metanyelvet alkalmaz, a felszíni szerkezetek kódolásába nem von be például funkcionális jelölőket). A metanyelv a felszíni szerkezetek közül elsősorban a formailag legkönnyebben megragadható szóosztályokat kódolja (pl. **v** = igei csoport, **n** = főnévi csoport, **adj** = melléknévi csoport, **adv** = határozószói csoport). A kódolás a lexikai elemeket is jelölheti, ez esetben ezeket kurziválni kell (pl. **V in n**: *I don't believe in coincidences* 'nem hiszek a véletlenekben'). A szóköz a szekvenciát, a nagybetű pedig az aktuálisan vizsgált szót jelöli¹² (Hunston–Francis 2000: 33–45). A mintázatokat ikonikusan leképző metanyelv megfelelően magasfokú generalizálást tesz lehetővé.¹³ A megfelelően magasfokú generalizálás biztosítja az átláthatóságot, miközben a túlzottan magasfokú általánosítás (pl. a tradicionális nyelvtanok esetében) a specifikus információ elvesztésével járna (Hunston–Francis 2000: 77–81). Hunston–Francis (2000: 18) a mintázatnyelvtant korpuszvezérelt kutatási irányként határozza meg, ugyanis ez az elméleti előfeltevéseket visszaszorítva¹⁴ von le induktív következtetéseket a korpusz adataiból. Biber (2010) azonban amellett érvel – véleményem szerint jogosan –, hogy a mintázatnyelvtan módszertana a korpuszalapú és a korpuszvezérelt megközelítésmód keveréke. Korpuszalapú, mert előzetesen, a korpuszadatok megfigyelése előtt elfogadja a szófajok létezését,¹⁵ a szófaji kategóriák tételezésével és használatával a mintázatok formalizálása során megengedi a professzionális intuíció használatát; ugyanakkor korpuszvezérelt is, mert a korpuszevidenciából induktívan következtet nyelvhasználatbeli mintákra, a mintázatok azonosítása és egymástól való elhatárolása az adatok korpuszbeli disztribúcióján alapszik. A mintázatnyelvtan, úgy tűnik, nem fogadja el a legradikálisabban korpuszvezérelt kutatások állítását, azt, hogy minden felszíni mintához egyedi jelentés asszociálódik (Biber 2010: 175–176).¹⁶

¹¹ Például a nyelvtanulók számára. A COBUILD-projekt több kiadványa is kimagaslóan sikeressé vált az angolt idegen nyelvként tanulók körében (Szirmai 2005: 6).

¹² A fenti esetekben a **V** jelölte azt, hogy az *explain* szóalak volt a korpuszban lekérdezett szó.

¹³ Ebből a szempontból a mintázatnyelvtanban alkalmazott megközelítés nem fogadja el Sinclair (1991) alap-tételét, amely szerint minden különböző formához különböző lexikai egység asszociálódik.

¹⁴ Hunston–Francis (2000) önreflexiója szerint a mintázatnyelvtan két előzetes elméleti előfeltevést érvényesít: (i) az egyes szavak viselkedése nagy jelentőséggel bír egy nyelvi elemzés számára, illetve (ii) a mintázatba rendeződés csak nagy mennyiségű nyelvi adaton figyelhető meg (Hunston–Francis 2000: 250).

¹⁵ Megjegyzendő, hogy Hunston–Francis (2000) nem tartja a szófajiságot a szavak inherens tulajdonságának; a szófajok fogalmát inkább csak módszertani okokból, a bonyolult körülírások elkerülése miatt fogadja el (Hunston–Francis 2000: 179–197).

¹⁶ Már itt érdemes tisztázni a mintázat fogalmának viszonyát a kollostrukció és a konstrukció fogalmához. Mivel a kollostrukció – átfogó megközelítés szerint – olyan konstrukció, amelynek bizonyos pontjain kollokálódó

A mintázatok azonosításakor két nézőpontból lehet vizsgálni az adatokat: (i) ha egy-egy vizsgált szóra keresünk – ebben az esetben a vizsgált szó a keresőkifejezés, a mintázatok pedig az (L1 vagy R1 távolságra lévő tokenek alapján) alfabetikusan rendezett vertikális konkordanciasorokban figyelhetők meg; (ii) ha egy-egy feltételezett vagy korábban megfigyelt mintázatra keresünk – ebben az esetben a keresőkifejezés több szóalakkal és/vagy korpuszlekerdező nyelv segítségével adható meg, a mintázatban előforduló szavak listája pedig javarészt a találati kifejezésekből nyerhető ki. Egy mintázat elemei megelőzhetik vagy követhetik a vizsgált szót, balra vagy jobbra is elhelyezkedhetnek tőle a konkordanciasoron belül. Nem mindig egyértelmű, hogy mi számít mintázatnak és mi nem. Azonosításkor segítheti a kutatót, ha figyelembe veszi, hogy a mintázat elemei gyakran fordulnak együtt elő, kölcsönösen függnek egymástól, és a sorrendjük rögzített lehet. Utóbbi tulajdonság azonban nem szükséges és nem is elégséges feltétele a mintázatoknak: egy mintázat elemei akár különböző sorrendben is megjelenhetnek. A mintázatoknak nem részei a mintázatbeli elemekkel összekapcsolt vonatkozó mellékmondatok, valamint a teret, időt vagy módot kifejező prepozíciós kifejezések vagy határozószavak (Hunston–Francis 2000: 43–76).

3. Anyag és módszer

A korpuszadatok elemzéséhez az adatokat a Hungarian Web Corpus (huTenTen12)-ből nyertem ki (Suchomel–Pomikálek 2012, Jakubíček et al. 2013). A korpuszt a Sketch Engine felületén keresztül értem el (Kilgarriff et al. 2004, Kilgarriff et al. 2014).¹⁷ A huTenTen12 a TenTen korpuszcsalád egyik tagja. A korpusz speciális korpusz (l. Szirmai 2005: 33) abban az értelemben, hogy az internetes nyelvhasználatot reprezentálja; statikus korpusz (l. Szirmai 2005: 32) abban az értelemben, hogy 2012-es létrehozása óta a magyar korpuszok viszonylatában viszonylag impozáns tokenszáma nem került bővítésre (a korpusz közel 2,5 milliárd szövegszóból áll). A statikus korpusz előnye, hogy lehetővé teszi a lekérdezések megismétlését, hátránya, hogy

lexikai elemek jelennek meg, Hunston–Francis (2000) mintázatfogalma értelmezhető lenne kollostrukcióként is (vö. Stefanowitch–Gries 2003), habár az értelmezés kiterjesztése a kollostrukciók irányába a dolgozatnak – mint pilotkutatásnak – nem közvetlen célja. Az azonosított mintázatok a kollostrukturális elemzés során közvetlenül hasznosíthatók (l. Hunston–Su 2017: 590–591), azonban érdemes rámutatni, hogy a mintázatnyelvtan és a kollostrukturális elemzés eltérő, bár nem összeegyeztethetetlen módszertant alkalmaz a mintázatok, illetve a kollostrukciók azonosítására (vö. Hunston–Francis 2000: 37–81; Stefanowitch–Gries 2003: 213–220). A Goldberg-féle konstrukciós nyelvten konstrukciófogalma (l. Goldberg 2006: 5) – a kollostrukció fogalmával ellentétben – nem azonosítható a mintázat fogalmával. Goldberg konstrukciós nyelvtena és a mintázatnyelvtan közötti fő különbség, hogy az előbbi (i) a formát és a jelentést nem abszolút és elválaszthatatlan egységként, hanem forma-jelentés párokként kezeli; (ii) a konstrukciót mentális reprezentációnak tekinti, ellentétben a mintázatnyelvtannal, amely a mintázatot empirikus adatokon megfigyelhető jelenségként értelmezi (Hunston–Su 2017: 569–571); valamint – az előzőekből adódóan –, hogy (iii) a konstrukció fogalma ebben a megközelítésben tipikusan magasabb absztrakciós szintet képvisel a mintázaténál. Az elméleti distinkciók ellenére a mintázatnyelvtan eredményei a Goldberg-féle konstrukciós nyelvten számára is hasznosíthatók: Hunston–Su (2017) például egy olyan módszertani eljárást dolgoz ki, amely a COBUILD-projekt keretében feltárt mintázatokot generalizálja tovább, hogy közepes szinten absztrakt konstrukciókat azonosítson.

¹⁷ Ehhez l. <http://www.sketchengine.eu> (Utolsó elérés: 2020. aug. 26.).

alacsonyabb mértékben validálja a kutatás korpuszevidenciákon alapuló generalizálását (Jakubíček et al. 2013). A korpusz választása mellett két érv szólt. Egyrészt az, hogy a kutatás előzményében megfigyelt korpuszadatok alapján az általam vizsgált szóalakok relatíve nagy számban fordulnak elő az internetes nyelvhasználatban. Másrészt az, hogy a magyar nyelv reprezentatív korpuszának tekinthető Magyar Nemzeti Szövegtár túlságosan nagy arányban tartalmaz duplumokat, a vizsgálati korpusz készítésekor azonban ezeket bekezdésszinten szűrték (Jakubíček et al. 2013).

A vizsgálat során a *bealszik* igealakot használtam keresőkifejezésként.¹⁸ A kifejezést kutatásom előzményében (l. Kopcsák 2019) részben valós nyelvhasználati eseményeket megfigyelve, részben a szakirodalomra támaszkodva (vö. Nádasdy 2002, Ladányi 2004) választottam. A keresőkifejezés meghatározásakor a korpuszvezérelt megközelítés radikális állítását vettem alapul, amely nem különbözteti meg a formát a jelentéstől, és nem lemmatizál. Emiatt nem a *bealszik* lemmát, hanem a *bealszik* szóalakot használtam keresőkifejezésként (a szóalakra 582 találat volt a kutatás vizsgálati korpuszában). A választást indokolta, hogy a kutatás előzményében végzett korpuszvizsgálat során ez az alak bizonyult az egyik leggyakoribbnak a *bealszik* lemma különböző megvalósulásai közül.

Az alábbi kutatásom korpuszvezérelt, ugyanis a korpuszban megfigyelt adatokból tesz induktív következtetéseket. A kutatás kevésbé radikális – Biber (2010) szóhasználatával élve: „hibrid” – megközelítést alkalmaz, amely abban a tekintetben megegyezik a Hunston és Francis (2000) által felvázolttal (vö. 2.2.), hogy egyedül a szavak és a szófajok létezését fogadja el előzetes elméleti feltevésként. A korpuszadatok megfigyelése és kiértékelése a Simon Gábor (2018) által javasolt módszertant követi: többszintű általánosítással lépésről lépésre halad a generalizálás során, ezzel elkerülve a statisztikai adatok torzulását (az esetlegesen releváns specifikus információk kifejejtésének veszélyét) (Simon 2018: 6). Az általánosítás három lépésben történik (vö. Simon 2018: 15–16).¹⁹

Az 1. lépésben a *bealszik* igealak kollokációs profilját vizsgálom meg.²⁰ A kollokáció méréséhez a vizsgálati kifejezést használok csomópontként²¹ (a *csomópont* és a *kollokált* terminusokhoz vö. Sinclair 1991: 115, Reményi 2010: 69), a magyar szakirodalom ajánlásának megfelelően -4 és +4 kollokációs ablakban (vö. Reményi 2010: 80). A csomópont és a kollokáltak asszociációs értékét Simon (2018) nyomán a logDice statisztikai eljárással mérem.²² A kollokációs profil meghatározásával előkészítem a további lépéseket, és a vizsgált kifejezés

¹⁸ A korpuszlekérdezés során a keresés egyaránt figyelembe vette a nagybetűvel és kisbetűvel kezdődő *bealszik* szóalakokat.

¹⁹ Az itt bemutatott háromlépcsős általánosítási eljárás abban a tekintetben eltér a Simon (2018) által kidolgozott generalizáló lépésektől, hogy a kollokációjelöltek áttekintésének és a mintázat azonosításának lépését különválasztja, leképezési táblázatokat azonban nem alkalmaz.

²⁰ A dolgozat szempontjából az a meghatározás releváns, amely szerint a kollokáció nem más, mint két kifejezés szignifikáns együttes előfordulása (Simon 2018: 5).

²¹ A kollokációmérés a korpuszlekérdezéshez hasonlóan érzéketlen volt nagybetű és kisbetű különbségére (l. a 18. lábjegyzetet).

²² A logDice statisztikai eljárás a többi kollokációmérésre használt eljárással ellentétben viszonylag pontosan kiszűri a hapaxokat és a funkciószavakat (Simon 2018: 14). A logDice további előnye, hogy az érték független a korpusz méretétől, ezért a kutatás folytatásában lehetővé teheti a különböző korpuszok adatainak összevetését (Rychlý 2008: 9).

szerkezetének néhány jellegzetességét taglalom. Mivel a csomóponthoz a kollokációjelöltek gyakran asszociálódhatnak, a csomópont és a kollokációjelöltek közös mintázatban is nagy eséllyel vehetnek részt. Ez a lépés tehát eltér a Hunston–Francis (2000) által prezentált mintázatazonítási módszertől. Míg Hunston–Francis (2000) egy-egy szóalakot használ keresőkifejezésként, majd a korpuszfelületen talált konkordanciasorokból közvetlenül olvassa ki azokat a mintázatokat, amelyekben a szóalak szerepel, addig a kutatásom a manuális szűrés időigényes folyamatát a kollokációjelöltek statisztikai alapú szűrésével részlegesen meggyorsítja (bár sem a kollokációmérést, sem a mintázatazonosítást nem automatizálja).

A 2. lépés során a csomópontot és annak egy-egy kollokáltját egyszerre tartalmazó konkordanciasorokban, a konkordanciasorok horizontális kotextusában²³ megfigyelhető mintázatjelölteket azonosítom. Ehhez először lementem a csomópontot és annak kollokáltjait tartalmazó konkordanciasorokat. Ezután a mintázatjelölteket a mintázatok metanyelve alapján kódolom, azonban további vizsgálat függvényévé teszem, hogy valóban azonosíthatóak-e mintázatként. A mintázatjelölt mint fogalom és mint módszertani lépés bevezetése nem csak egy lehetséges tautológiát kerül el.²⁴ A mintázatjelöltek külön kódolása azt mutatja fel, hogy egy csomópont és egy kollokált közösen vehet részt a mintázatokban, azonban a mintázatazonosítás az elemző és a professzionális intuíció hatáskörébe tartozik.²⁵ A mintázatjelöltek szoros olvasatából következtetek a mintázatokra. Mintázat alatt forma és jelentés abszolút, totális és elválaszthatatlan egységét, valamint emergáló új minőségét értem (elfogadva ezzel Hunston–Francis 2000 definícióját; l. 2.2.). Amennyiben több mintázatjelölt is azonos metanyelvi kódot (szófajokat, alszófajokat) tartalmaz, és ezek közel azonos szórendet mutatnak, valamint a metanyelvi kódokon belül közel azonos jelentésű, illetve azonos jelentéstani csoportba tartozó²⁶ szavak választódnak ki, a mintázatjelölteket mintázatként kezelem.

A mintázatok és mintázatjelöltek kódolásakor a mintázatnyelvtanban javasolt metanyelvet alkalmazom azzal a változtatással, hogy (i) az angol kódolást magyarra módosítom (pl. **v** helyett **ig** 'ige', **n** helyett **fn** 'főnév', **adj** helyett **mn** 'melléknév'), (ii) a névutó szófaját is beveszem a kódolásba (**névut** jelöléssel), valamint hogy (iii) a kódolás során jelölöm a főnévi inflexiós toldalékokat (pl. **n-helyviszonyrag**). Az (i) és (ii) változtatásra a leíró metanyelv rugalmasságának és transzparenciájának megőrzése miatt volt szükség.²⁷ A (iii) változtatást az angol és a magyar nyelv szóalakjainak eltérő felépítése indokolja (a magyarban az inflexiós viszonyok elvesztek volna a kódolás túlgeneralizálásával).²⁸ A (ii) és a (iii) módosítás bevezetése egyben arra is rámutat, hogy az elemzőnek minden kódolás és minden metanyelv kialakítása során döntenie kell arról, hogy milyen mértékben kíván általános és specifikus jelenségeket megjeleníteni a saját kódolási rendszerében. A dolgozatban használt metanyelv némileg specifikusabb a Hunston és Francis (2000) által javasolthoz képest, de nem annyira specifikus, hogy ne

²³ A kotextus a megnyilatkozásban előforduló szavak nyelvi környezete (Tátrai 2011: 54–55).

²⁴ A mintázatjelölt tehát egy olyan potenciális mintázat, amelyet az elemző a csomópontot és annak kollokáltjait tartalmazó konkordanciasorok manuális értelmezése során azonosít.

²⁵ A professzionális intuícióhoz l. Majoros (2013), Simon (2018).

²⁶ A mintázatjelöltek jelentéstani csoportját a professzionális intuíció azonosítja.

²⁷ Egy rugalmas, transzparens és konzisztens metanyelv alkalmazásával a magyar nyelv mintázatai (pontosabban: egy adott időpillanatban tipikus mintázatai) is leírhatók lehetnének.

²⁸ A mintázatnyelvtan metanyelve eredetileg is kódol inflexiós toldalékokat (pl. **V-ed**), de néhány esetben akár derivációs toldalékokat is (pl. **V-ing**).

lehesen az eredményeket közepes szinten generalizálni. (Arra, hogy van lehetőség a kódolás specifikus irányba történő elmozdítására, Hunston–Francis (2000: 81) is felhívja a figyelmet.)

A 3. lépésben a lehetséges felszíni mintázatokból – a forma és a jelentés abszolút, totális és elválaszthatatlan egysége, valamint a kettő egyszerre történő kiválasztódása miatt – közvetlenül következtek a mintázatjelöltek, illetve a mintázatok jelentésére. A jelentések klasszifikálásához a funkcionális nyelvelmélet eszköztárából (l. pl. Croft 2012, Tolcsvai Nagy 2015) meríttek. A mintázatjelöltek jelentésének értelmezése egyben további módszertani megerősítésként szolgál a mintázatok azonosításában. Amennyiben adott formákhoz egyértelműen hozzáasszociálódnak adott jelentések, és a mintázat minden eleme részt vesz a forma és a jelentés emergenciájában, úgy a mintázatjelöltek mintázat jellege egyértelmű megerősítésre talál.

4. A *bealszik* szóalak vizsgálata

4.1. Első lépés: a kollokációs profil feltérképezése

A *bealszik* csomópont kollokálódásának mérésekor a logDice asszociációs értéket használva a felület 47 kollokációjelöltet adott ki eredményként. Mivel a korpusznyelvészetben nincs meg egyezés arról, hogy milyen logDice-értéktől számít stabilnak a kollokáció, a vizsgálathoz a 47 kollokációjelölt közül a 20 legstabilabbat választottam ki. A 20 legstabilabb kollokációjelölt kiválasztása során az *n*-best list (vagyis az '*n*-legjobb [kollokációjelölt-]listája') módszert alkalmaztam (ahol *n* a természetes számok halmazából kerül ki). Az *n*-best list módszer lényege, hogy egy adott, tetszőlegesen választható asszociációs érték esetében az elemző az első *n* darab legmagasabb asszociációs értékű kollokációjelöltet választja ki a további elemzés számára (az *n*-best list korpusznyelvészeti alkalmazásához l. Evert–Krenn 2005). A kollokációjelöltek közül az *NV* és a *KV* rövidítés személyneveket jelöl:

- (1) *NV*/fejfájással hamar *bealszik* [...]
- (2) *KV**bealszik* másfél órát [...]

Bár a logDice asszociációs érték alkalmazásának egyik előnye éppen a hapaxok kiszűrése kellene legyen (vö. a 22. lábjegyzettel), az (1) és a (2) példákban is megfigyelhető, hogy az *NV* és a *KV* kollokációjelöltek hapaxok – valószínűleg egy adott megnyilatkozóhoz kötődnek. Az *NV* és a *KV* zajként való megjelenése a kollokációjelöltek listáján azzal is összefügghet, hogy a mama-/szülőblogok valószínűleg felülreprezentáltak a lekért korpuszadatokban.²⁹ A két hapaxot, mivel nem tényleges kollokáltak, a továbbiakban nem elemzem.

A kollokációjelöltek listájából nem lehet egyértelmű szemantikai csoportokat azonosítani, néhány jelzésértékű tendencia azonban így is kirajzolódik. A csomópont a legerősebben a *cicin*, a *beájul*, a *szopizás* és a *szopi* kifejezésekkel asszociálódik. Megfigyelhető, hogy a legstabilabb kollokációjelöltek jelentése (de közvetetten akár még a *berakom* és a *leteszem*

²⁹ A vizsgált konkordanciasorokból úgy tűnik, hogy az *NV* és a *KV* gyakori asszociálódása a *bealszik* igealakhoz abból adódhat, hogy egyetlen nyelvhasználó a mama-/szülőblogok felületén többször is lejegyezte a saját gyerekeivel (akiket az *NV* és a *KV* rövidítésekkel nevezett meg) megtörténő bealvási eseménysort. Az igealak korpuszbeli, egyenlőtlen disztribúciójához és felülreprezentáltságához l. a Mellékletben csatolt ábrát.

kollokációjelölteké is) leginkább a csecsemők táplálásával függ össze. A 18 nem hapax kollokációjelölt közül összeségében úgy tűnik, 7 nyelvi elem függ össze a táplálkozással/táplálással (*szopizás, szopi, szopizik, szoptatás, evés, kaja*). Meglehetősen meglepőnek hat, hogy a *bealszik* csomópont a *beájul* kifejezéssel jellegzetesen gyakran fordul együtt elő, ugyanis a szakirodalom mindkét kifejezés jelentését a szubmerzív tartomány alá sorolja be (vö. Csillag 2013). Az 1. táblázat 18 nem hapax kollokációjelöltje gyenge tendenciaként mutatja fel a hely körülményének kidolgozását (*cicin, kocsiban, autóban*), illetve ennél valamivel erősebb tendenciaként rajzolódik ki a vizsgált csomópont és a főnévi szerkezetben összegződő események együttes előfordulása (*szopizás, evés, fürdés, szoptatás*). Bár a kollokációs profilból egy erősebb tendencia is kirajzolódik (a táplálással/táplálkozással összefüggő kifejezések magas gyakorisága), a lista adatai nem körvonalaznak egyértelmű szemantikai mintázatot.

Kollokációjelölt	logDice
<i>cicin</i>	8,54
<i>beájul</i>	8,52
<i>szopizás</i>	7,19
<i>szopi</i>	6,52
<i>kiságyba</i>	6,03
<i>berakom</i>	5,56
<i>NV</i>	5,30
<i>szopizik</i>	5,05
<i>leteszem</i>	4,81
<i>KV</i>	4,00
<i>evés</i>	3,72
<i>fürdés</i>	2,88
<i>ébred</i>	2,57
<i>kocsiban</i>	2,52
<i>eszik</i>	2,37
<i>kaja</i>	2,17
<i>autóban</i>	2,08
<i>szoptatás</i>	2,07
<i>fáradt</i>	2,03
<i>kezemben</i>	1,93

1. táblázat. A *bealszik* 20 legstabilabb kollokációjelöltje

4.2. Második lépés: a mintázatjelöltek azonosítása

A könnyebb áttekinthetőség érdekében minden mintázatjelölthöz egy-egy korpuszbeli adatrészletet is megadtam. Az eleve gyengének mutatózó mintázatjelölteket aláhúzással jelöltem,³⁰ ezekből 4 fordult elő a vizsgált anyagban, mégpedig a *berakom*, a *fürdés*, a *kaja* és a

³⁰ A mintázatjelöltek mintázatjelöltként való azonosíthatóságát gyengíthette, ha a csomópont és a kollokációjelöltek közös konkordanciaadataiban a mintázatjelölt nem volt kizárólagosan megfigyelhető, ha a mintázatjelöltben kódolt lexikális és szófaji elemek nem voltak stabilak, vagy ha a mintázatjelöltek túl nagy szórendi variabilitást mutattak.

kezemben kollokáltakhoz tartozó mintázatjelöltek esetében. A 14 erősebb mintázatjelöltből 2 mintázatjelölt csak egyszer-ször fordul elő; ezek az **(utalószó) mn, (kötőszó) Ig** és az **Ig, ig felsorolás**³¹ mintázatjelöltek. Mivel a két mintázatjelölt csak egy-egy kollokálthoz köthető, ezek kis valószínűséggel alkotnak a csomóponttal tényleges mintázatot. A fennmaradó 12 mintázatjelölt 4 nagyobb csoportként különíthető el. A 4 csoport közül a leggyengébb az **Ig, ig-tagmondat** és az **ig, Ig-tagmondat**, amelyekhez két-két kollokált köthető (*kiságyba, leteszem*; illetve *szopizik, eszik*), és amelyek között a legfőbb különbség, hogy a csomópont az első vagy a második tagmondatban jelenik meg.³² Az előbb említett két mintázatjelöltnél valamivel erősebb az **Ig fn-helyviszonyrag** csoport, amely 3 kollokálthoz tartozó felszíni szerkezetben fordul elő (*cicin, kocsiban, autóban*). A leggyakoribb mintázatjelölt az **fn névut Ig** (5 példány). A mintázatjelöltek közül egyedül a legutoljára említett **fn névut Ig** mintázatban figyelhető meg, hogy valamelyik szófaji pozíció egy tisztán körülhatárolható lexikai csoporttal asszociálódik. Ebben az esetben az **fn** szófajon belül 4 olyan kifejezés jelenik meg, amelyet a 4.1.-ben táplálkozással/táplálással összefüggő nyelvi elemként azonosítottam (*szopizás, szopi, evés, szoptatás*). Ez a megfigyelés egyrészt előjelezheti azt, hogy az **fn névut Ig** mintázatjelölt viselkedik a legnagyobb valószínűséggel tényleges mintázatként, másrészt összefügghet azzal a 4.1.-ben említett megfigyeléssel, hogy a szülő- vagy kismamablogok erősen felülreprezentáltak a lekért korpuszadatokban.

Kollokáció-jelölt	logDice	Mintázatjelölt	Példa
<i>cicin</i>	8,54	Ig fn-helyviszonyrag	<i>Nyugis a baba, kicsit talán túl nyugis, bealszik a cicin, állandóan ébresztgetni kell.</i>
<i>beájul</i>	8,52	Ig, ig felsorolás	<i>Ilyen a beájul meg a bealszik meg a besértődik meg a bevállal meg még sok másik</i>
<i>szopizás</i>	7,19	fn névut Ig	<i>szopizás közben rendszeresen bealszik kb. tíz perc után</i>
<i>szopi</i>	6,52	fn névut Ig	<i>az éjszakai szopi után pedig bealszik és cumi sem kell</i>
<i>kiságyba</i>	6,03	Ig, ig-tagmondat	<i>míg mélyen bealszik, aztán kiságyba csempészes</i>
<i>berakom</i>	5,56	–	–
<i>NV</i>	5,30	–	–
<i>szopizik</i>	5,05	ig, Ig-tagmondat	<i>Mostanában éjszaka szopizik sokáig, bealszik 5-10 percre</i>
<i>leteszem</i>	4,81	Ig, ig-tagmondat	<i>újra bealszik, de amint leteszem fent van</i>
<i>KV</i>	4,00	–	–

³¹ A zárójel a metanyelven belül azt kódolja, hogy az adott elem nem stabil része a mintázatjelöltnak. Az utalószó és a kötőszó mint kódolt elem egyelőre „kényszermegoldás” a metanyelv alkalmazásakor, további szóalakok vizsgálata szükséges ahhoz, hogy az általánosítás indokoltsága megerősíthető legyen. Az **Ig, ig felsorolás** mint kódolás azt rögzíti, hogy a konkordanciasorokban igék (igealakok) felsorolása történik.

³² Ez nincs ellentmondásban a 30. lábjegyzettel, ugyanis a szórendi stabilitás mint kritérium a 30. lábjegyzetben a mintázatjelölt konkordanciasorban való megfigyelhetőségére vonatkozott, jelen esetben azonban már egy generalizálási szinttel feljebb lépve a mintázatjelöltek tényleges mintázatként való felismeréséről van szó.

<i>evés</i>	3,72	fn névut Ig	<i>Evés közben bealszik és csak nagy nehezen lehet felébreszteni,</i>
<i>fürdés</i>	2,88	–	–
<i>ébred</i>	2,57	Ig, ig-tagmondat	<i>ha egyszer bealszik , akkor sem ébred fel</i>
<i>kocsiban</i>	2,52	Ig fn-helyviszonyrag	<i>úgy kell értelmezni, hogy piásan bealszik a kocsiban</i>
<i>eszik</i>	2,37	ig, Ig-tagmondat	<i>megint eszik egy kicsit és bealszik olyan 9 fél 10 felé</i>
<i>kaja</i>	2,17	–	–
<i>autóban</i>	2,08	Ig fn-helyviszonyrag	<i>Ha véletlenül egyszer-egyszer bealszik az autóban délután</i>
<i>szoptatás</i>	2,07	fn névut Ig	<i>Sok újszülött a szoptatás alatt bealszik a mellen</i>
<i>fáradt</i>	2,03	(utalószó) mn, (kötőszó) Ig	<i>estére annyira fáradt és nyugós hogy bealszik idő előtt</i>
<i>kezemben</i>	1,93	–	–

2. táblázat. A *bealszik* kollokációjelöltjeinek mintázatjelöltjei

4.3. Harmadik lépés: szemantikai csoportosítás

A szemantikai csoportosítás kategóriáinak kialakításakor elsősorban a funkcionális nyelvelmélet eszközkészletére támaszkodtam (vö. Croft 2012, Tolcsvai Nagy 2015). A vizsgált szóalakot állapotváltozást profiláló kifejezésként határoztam meg (a kutatás előzményétől nem függetlenül, vö. Kopcsák 2018). Szemantikai profilt mintázatjelölt nélküli kifejezésnél nem határoztam meg. A szemantikai profilok négy jól elkülönülő csoportba oszthatók, amelyek megközelítőleg egybeesnek a 4.2-ben elhatárolt négy csoporttal (mintázatjelölttel). A négy csoport az állapotváltozás (tehát a bealvás) valamilyen körülményét, résztvevőjét, okát vagy következményét dolgozza ki.³³

Kollokáció-jelölt	logDice	Mintázatjelölt	Példa	Szemantikai csoport
<i>cicin</i>	8,54	Ig fn-helyviszonyrag	<i>Nyugis a baba, kicsit talán túl nyugis, bealszik a cicin, állandóan ébresztgetni kell.</i>	állapotváltozás körülménye (hely)
<i>beájul</i>	8,52	Ig, ig felsorolás	<i>Ilyen a beájul meg a bealszik meg a be-sértődik meg a bevállal meg még sok másik szopizás közben rendszeresen bealszik kb. tíz perc után</i>	nyelvi metareflexió
<i>szopizás</i>	7,19	fn névut Ig	<i>az éjszakai szopi után pedig bealszik és cumi sem kell</i>	állapotváltozás körülménye (idő) és okozója
<i>szopi</i>	6,52	fn névut Ig	<i>míg mélyen bealszik, aztán kiságyba csempészés</i>	állapotváltozás körülménye (idő) és okozója
<i>kiságyba</i>	6,03	Ig, ig-tagmondat	–	állapotváltozás következménye
<i>berakom</i>	5,56	–	–	–
<i>NV</i>	5,30	–	–	–

³³ Nem sorolható csoportba a nyelvi metareflexió kategóriája. Ezek a metareflexiók *be* igekötővel képzett neológ szerkezeteket tartalmaznak, és a felsorolás felszíni szerkezetében figyelhetők meg (vö. 2. táblázat). A metareflexió szemantikai kategóriája magyarázattal szolgál arra, hogy a *bealszik* csomópontnak miért a *beájul* a második legerősebb kollokáltja.

<i>szopizik</i>	5,05	ig, Ig-tagmondat	<i>Mostanában éjszaka szopizik sokáig, bealszik 5-10 percre</i>	állapotváltozás előzménye, oka
<i>leteszem</i>	4,81	Ig, ig-tagmondat	<i>újra bealszik , de amint leteszem fent van</i>	állapotváltozás következménye
<i>KV</i>	4,00	–	–	–
<i>evés</i>	3,72	fn névut Ig	<i>Evés közben bealszik és csak nagy nehezen lehet felébreszteni,</i>	állapotváltozás körülménye (idő) és okozója
<i>fürdés</i>	2,88	–	–	–
<i>ébred</i>	2,57	Ig, ig-tagmondat	<i>ha egyszer bealszik , akkor sem ébred fel</i>	állapotváltozás következménye
<i>kocsiban</i>	2,52	Ig fn-helyviszonyrag	<i>úgy kell értelmezni, hogy piásan bealszik a kocsiban</i>	állapotváltozás körülménye (hely)
<i>eszik</i>	2,37	ig, Ig-tagmondat	<i>megint eszik egy kicsit és bealszik olyan 9 fél 10 felé</i>	állapotváltozás előzménye, oka
<i>kaja</i>	2,17	–	–	–
<i>autóban</i>	2,08	Ig fn-helyviszonyrag	<i>Ha véletlenül egyszer-egyszer bealszik az autóban délután</i>	állapotváltozás körülménye (hely)
<i>szoptatás</i>	2,07	fn névut Ig	<i>Sok újszülött a szoptatás alatt bealszik a mellen</i>	állapotváltozás körülménye (idő) és okozója
<i>fáradt</i>	2,03	(utalószó) mn, (kötőszó) Ig	<i>estére annyira fáradt és nyugós hogy bealszik idő előtt</i>	állapotváltozás előzménye, oka
<i>kezemben</i>	1,93	–	–	–

3. táblázat. A *bealszik* mintázatjelöltheihez asszociálódó szemantikai csoportok

A négy szemantikai csoportot a 3. táblázat adatai mutatják be. Az első szemantikai csoporthoz az **Ig, ig-tagmondat** mintázatjelölt tartozik. Ebben a csoportban az első tagmondat a bealvás folyamatát, a második tagmondat pedig az állapotváltozásra következő, az állapotváltozás által okozott eseményt dolgozza ki.³⁴ A (3)-ban a példában dőlt betűvel kiemelt **Ig, ig-tagmondat** mintázatjelölt első tagmondata a bealvás folyamatát dolgozza ki, amelynek közvetlen következménye az azt időben követő és a mintázatjelölt második tagmondatában megkonstruálódó esemény, az elsődleges figura fel nem ébredésének megfigyeltetése:

(3) Mondjuk, szerencsém van vele, *mert ha egyszer bealszik , akkor sem ébred fel*, ha mondjuk a kenguruban aludt el [...]

Az első csoporttal szorosan összefügg a második azonosított szemantikai csoport, amely a folyamat előzményét dolgozza ki. Ide tartoznak az **ig, Ig-tagmondat**, valamint az **(utalószó) mn, (kötőszó) Ig** mintázatjelöltek. Ebben az esetben a második tagmondat dolgozza ki a bealvás folyamatát, miközben az első tagmondat a folyamat időbeli előzményének hozzáférhetővé tétele mellett egyben a folyamat indokát is implikálja,³⁵ amely arra is utalhat, hogy az ige által megnevezett folyamat olyan állapotváltozás, amelyet az elsődleges figura szenved el (vö. Kopcsák 2018). A (4) adat első tagmondatában megkonstruálódó jelenet, a csecsemő időben elhúzódó szoptatása például a második tagmondatban kidolgozott bealvás folyamatá-

³⁴ Az ok-okozat viszonyt kidolgozó tagmondatok, illetve az események ikonikusan motivált egymásra következtetésének összefüggéséhez l. Kugler (2017: 819–820).

³⁵ Vö. az előző, 34. lábjegyzettel.

nak közvetlen időbeli előzményeként és ezzel együtt – mivel az anyatejjel való elteltődés álmósító hatással bír – a bealvás folyamatának indokaként alkotódik meg:

(4) *Mostanában éjszaka szopizik sokáig, bealszik 5-10 percre, aztán elkezdi nyűglődni [...]*

A harmadik csoportban a folyamat egyik körülménye, a hely profilálódik, amely az **Ig fn-helyviszonyrag** mintázatjelölt felszíni szerkezetében figyelhető meg. A folyamat helyének kidolgozása szintén szolgálhat az állapotváltozás egyik n-edik rendű, implicit okaként. Az (5) példában az *autóban* nyelvi elem profilálja a bealvás folyamatának helyét. A megnyilatkozó az *autóban* főnévvel a folyamat olyan körülményét dolgozza ki, amellyel implikálja, hogy a bealvás mint állapotváltozás a járműben való tartózkodás miatt megy végbe,³⁶ illetve kényszerül rá az elsődleges figurára (aki ebben a kontextusban is egy csecsemő):

(5) Ha véletlenül egyszer-egyszer *bealszik az autóban* délután, sokkal később sikerül csak este lerakni.

A negyedik csoportban a 4.2-ben legerősebb mintázatjelöltként felismert **fn névut Ig** szemantikai profilja figyelhető meg. A szemantikai profil és a felszíni szerkezet közös jellemzője, hogy a **névut** pozíció az állapotváltozás időviszonyítását, a **fn** szófaj pedig az állapotváltozás okozóját, kezdeményező szerepű résztvevőjét teszi hozzáférhetővé. A (6) és a (7) megnyilatkozásokban a **névut** pozíciójú *közben* és *után* elemek a mintázatjelölt **fn** szófajú *szopi* és *evés* eseményeihez viszonyítják a bealvás folyamatát. A **fn** szófajú *szopi* és *evés* elemek a mintázatjelöltben a táplálkozás eseményszerűségét, egyben a bealvás mint állapotváltozás okozóját dolgozzák ki – a táplálkozás a csecsemőként profilálódó elsődleges figura számára (a (4)-hez hasonlóan) ezekben az adatokban is álmósító hatással bír:

(6) Cumival alszik el este, de amint elaludt kiköpi, szóval ez nem zavarja, az éjszakai *szopi után* pedig *bealszik* és cumi sem kell.

(7) *Evés közben bealszik* és csak nagy nehezen lehet felébreszteni, hogy megegye a neki szánt mennyiséget (+/-120ml).

A mintázatjelöltek azonosítása során a **fn névut Ig** tűnt a legerősebb mintázatjelöltnak. A 3. táblázat jelentéstani szempontok szerinti elemzése után azt állítom, hogy a vizsgált szóalak esetében a **fn névut Ig** mintázatjelölt tényleges mintázatként írható le, és nagy valószínűséggel az **Ig fn-helyviszonyrag** mintázatjelölt is valós mintázat. Ugyanis egyfelől több kollokációjelölt vesz részt a két elemzett mintázatban, másfelől a mintázat egy olyan új, nem megjósolható, emergáló minőséget hoz létre a forma és a jelentés szintjén, amelyben minden mintázatban kódolt szófajnak, nyelvi elemnek szerepe van. Formailag azért, mert a *bealszik* szóalakból nem jósolható meg a két azonosított mintázat; jelentéstaniilag pedig azért, mert a mintázatok az állapotváltozás mellett annak helyét és idejét is kidolgozzák úgy, hogy egyben a folyamat okára való utalást is magukban foglalják. A mintázatok az esemény tér- és időbeli

³⁶ Habár az (5)-ben ez nincs explikálva, lehetséges ok lehet az autó kényelme és/vagy monoton mozgása.

körülményeinek megnevezésével implicit vagy explicit módon megadják az állapotváltozás indokát úgy, hogy a **fn névut Ig** esetében a főnév az állapotváltozás indokát, eseményszerű előzményét, a névutó az állapotváltozás előzményéhez való időviszonyítást dolgozza ki, míg az **Ig fn-helyviszonyrag** esetében a helyviszonyraggal inflexált főnév az állapotváltozás helyét és annak többedrangú, implicit indokát teszi hozzáférhetővé. A bealvás mint esemény feltételezhetőleg nem szokványos az eseményt előidéző indok kidolgozása nélkül, az esemény indokának kidolgozására pedig azért lehet szükség, mert a bealvás mint esemény (i) nem felel meg egy tipikus (felnőtt vagy legalábbis nem csecsemő) humán résztvevő életritmus szerinti alvási idejének vagy helyének, illetve (ii) az állapotváltozást elszenvedő személynek (feltételezett) szándéka vagy önmonitorozási folyamatai ellenére megy végbe.

5. Összegzés

A dolgozat egy olyan pilotvizsgálatot mutatott be, amely egy jövőbeli kutatást készít elő. A kutatás folytatásának célja, hogy létrehozzon egy olyan oktatási segédanyagot, amely a magyar nyelv legfontosabb mintázatait listázza és jellemzi röviden. Ez a lista mint a magyar nyelv oktatásában indirekt módon alkalmazható segédanyag segíthet a magyart idegen nyelvként tanulók (vagy a magyart mint származásnyelvet beszélők) számára abban, hogy könnyebben kifejezzék és megértessék magukat az adott nyelven. A jövőbeli kutatás előkészítése során a dolgozat módszertani és elméleti megállapításokat is megfogalmazott. Áttekintettem a korpusznyelvészet néhány állítását és irányzatát, ismertettem a mintázat fogalmát, bevezettem egy konzekvens és minimális előzetes nyelvi-nyelvészeti tudással is érthető metanyelvet, továbbá érveltem a mintázatok feltárásának egy lehetséges módja mellett. Bevezettem a mintázatjelölt fogalmát, amely fontos szerepet játszik a mintázatazonosítás folyamata során. A mintázatazonosításhoz Hunston–Francis (2000) és Simon (2018) alapján egy olyan módszert dolgoztam ki, amely több szinten generalizál: feltérképezi a szóalakhoz tartozó kollokációs profilt, a kollokáltak és a csomópont közös konkordanciasoraiban megfigyeli a csomópont mintázatjelöltjeit, majd a mintázatjelölt-lista értelmezése és szemantikai elemzése után, az utolsó generikus szinten meghatározza a szóalakhoz tartozó mintázatokat.

A 4. fejezet több szinten generalizáló lépésekkel vizsgálta a *bealszik* igealak kollokáltjait és mintázatjelöltjeit. A mintázatjelöltek listáit elemezve és a mintázatjelöltek szemantikai jellemzőit megvizsgálva két tényleges mintázatot figyeltem meg, amelyek a *bealszik* igealakot tartalmazzák. Ez a két mintázat a **fn névut Ig** és az **Ig fn-helyviszonyrag** mintázat volt.

Bár túlságosan ambiciózus vállalás lenne a magyar nyelv mintázatainak teljes körű leírását célul tűzni ki a kutatás jövőbeli folytatásában, a módszerben rejlő lehetőségek jelzésértékűek. Az alkalmazott és a további kutatás során finomítandó metanyelv transzparenciája segítheti, hogy a kutatás bekapcsolódhasson a nyelvészetten kívüli diskurzusokba, a magyar mint idegen nyelv tanításának módszertani gyarapításában, valamint szerepet vállaljon a laikus olvasók nyelvi készségeinek gyarapításában.

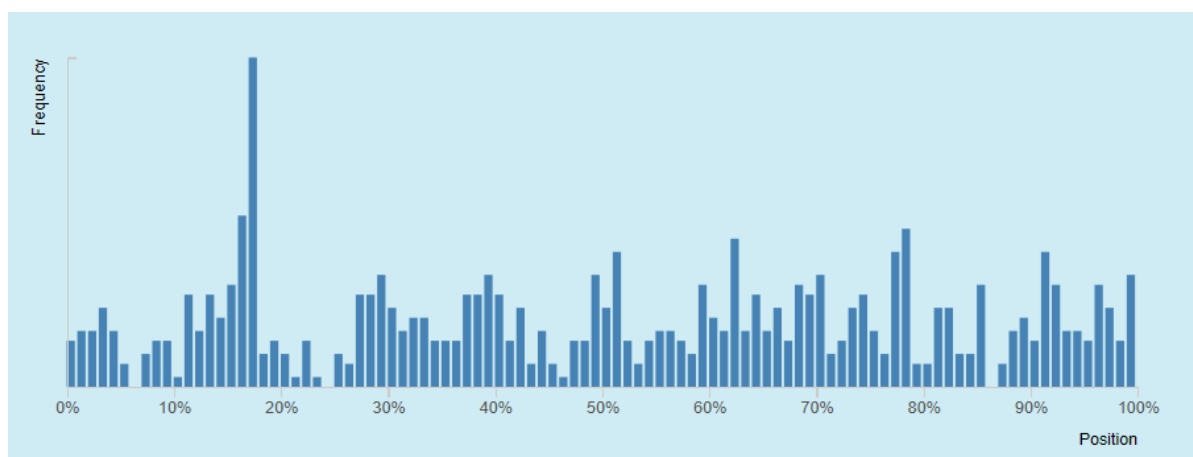
Hivatkozott szakirodalom

- Biber, Douglas 2010. Corpus-based and corpus-driven analysis of language variation and use. In Heine, Bernd – Narrog, Heiko (eds.): *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press. 159–192.
- Croft, William 2012. *Verbs: aspect and causal structure*. Oxford: Oxford University Press.
- Csillag Andrea 2013. Mit tanulhatunk a fiatalok nyelvéről? A *be-* igekötő kognitív szemantikai elemzése néhány újabb keletű példa alapján. *Képzés és Gyakorlat* 11: 7–16.
- Evert, Stefan – Krenn, Brigitte 2005. Using Small Random Samples for the Manual. Evaluation of Statistical Association Measures. *Computer Speech and Language* 19(4): 450–466.
- Firth, John Rupert. 1957. A synopsis of linguistic theory, 1930–1955. In Firth, John Rupert (ed.): *Studies in Linguistic Analysis. Special Volume. Philological Society*. Oxford: Blackwell. 1–32.
- Francis, Gill 1995. Corpus-driven grammar and its relevance to the learning of English in a cross-cultural situation. In Pakir, Anne (ed.): *English in Education: Multicultural perspectives*. Singapore: Unipress.
- Francis, Gill – Hunston, Susan – Manning, Elizabeth 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
- Francis, Gill – Hunston, Susan – Manning, Elizabeth 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.
- Goldberg, Adele 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Gumbrecht, Hans Ulrich 2010. *A jelenlét előállítása. Amit a jelentés nem közvetít*. Ford. Palkó Gábor. Budapest: Ráció Kiadó.
- Halliday, Michael A. K. 1994. *An Introduction to Functional Grammar*. 2nd edition. London: Edward Arnold.
- Hunston, Susan – Francis, Gill 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Studies in Corpus Linguistics 4. Amsterdam: John Benjamins Publishing Company.
- Hunston, Susan – Su, Hang 2017. Patterns, Constructions, and Local Grammar: A Case Study of 'Evaluation'. *Applied Linguistics* 40(4): 567–593.
- Jakubíček, Miloš – Kilgariff, Adam – Kovář, Vojtěch – Rychlý, Pavel – Suchomel, Vít 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*. 125–127.
- Kilgariff, Adam – Baisa, Vít – Bušta, Jan – Jakubíček, Miloš – Kovář, Vojtěch – Michelfeit, Jan – Rychlý, Pavel – Suchomel, Vít 2014. The Sketch Engine: ten years on. *Lexicography* 1: 7–36.
- Kilgariff, Adam – Rychlý, Pavel – Smrž, Pavel – Tugwell, David 2004. Itri-04-08. The sketch engine. *Information Technology* 105: 116.
- Kopcsák Róbert 2018. A magyar aspektusrendszer és a szubmerzív jelentés – a *bealszik* példája. *Ösvények* 2018: 30–44.
- Kopcsák Róbert 2019. *A szubmerzív jelentés a magyar aspektusrendszerben (kognitív nyelvészeti megközelítés)*. Szakdolgozat. Budapest: ELTE BTK.
- Kugler Nóra 2017. Az összetett mondat. In Tolcsvai Nagy Gábor (szerk.): *Nyelvtan. A magyar nyelv kézikönyvtára* 4. Budapest: Osiris Kiadó. 806–895.
- Ladányi Mária 2004. Rendszer – norma – nyelvhasználat. Igekötős neologizmusok „helyi értéke”. In Büky László (szerk.): *Nyelvleírás és nyelvművelés, nyelvhasználat, stilsztika*. Szeged: Szegedi Egyetem. 97–118.
- Ladányi Mária 2007. *Produktivitás és analógia a szóképzésben: elvek és esetek*. Segédkönyvek a nyelvészet tanulmányozásához 76. Budapest: Tinta Könyvkiadó.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar. Volume I. Theoretical prerequisites*. Stanford, California: Stanford University Press.
- Majoros Krisztián 2013. Metafora és kollokáció. In Váradi Tamás (szerk.): *Alknyelvdok7. Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest: MTA Nyelvtudományi Intézet. 131–142.

- Nádasdy Ádám 2002. *Besír, beröhög. Magyar Narancs* 51(35): 41.
- Reményi Andrea Ágnes 2010. Kollokációk korpuszalapú vizsgálata. *Fordítástudomány* 12: 67–95.
- Rychlý, Pavel 2008. A lexicographer-friendly association score. In Sojka, Petr – Horák, Aleš (eds.): *RASLAN 2008. Recent Advances in Slavonic Natural Language Processing*. Brno: Masaryk University. 6–9.
- Simon Gábor 2016. A metaforikus jelentés nyelvtanvezérelt megközelítése. *Magyar Nyelvőr* 140(2):178–199.
- Simon Gábor 2018. Az igei jelentés metaforizációjának mintázatai. Nyelvtan- és korpuszvezérelt esettanulmányok. *Jelentés és nyelvhasználat* 5: 1–36.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stefanowitsch, Anatol – Gries, Stefan Thomas 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209–243.
- Suchomel, Vít – Pomikálek, Jan 2012. Efficient web crawling for large text corpora. In *Proceedings of the seventh Web as Corpus Workshop (WAC7)*. 39–43.
- Szirmai Mónika 2005. *Bevezetés a korpusznyelvészetbe*. Budapest: Tinta Könyviadó.
- Tátrai Szilárd 2011. *Bevezetés a pragmatikába. Funkcionális kognitív megközelítés*. Budapest: Tinta Kiadó.
- Tognini-Bonelli, Elena 2001. *Corpus linguistics at work*. Studies in Corpus Linguistics 6. Amsterdam: John Benjamins Publishing Company.
- Tolcsvai Nagy Gábor 2015. *Az ige a magyar nyelvben. Funkcionális elemzés*. Segédkönyvek a nyelvészet tanulmányozásához 181. Budapest: Tinta Könyvkiadó.

Melléklet

A *bealszik* szóalak disztribúciója a vizsgált korpuszt (huTenTen12) alkotó dokumentumokra vetítve a Sketch Engine felületén. Az ábráról az olvasható le, hogy a *bealszik* szóalaknak az egyik dokumentumon belül szokatlanul magas az előfordulása (ezt az x-tengely 10% és 20% pozíciója között kiugró oszlop jeleníti meg, amely egy, az ábrán nem specifikált dokumentumon belül jelöli a vizsgált szóalak relatív gyakoriságát):



KOPCSÁK RÓBERT
Eötvös Loránd Tudományegyetem
kopcsakrobert@gmail.com