

MARKÓ ALEXANDRA^{1,2} – FŐZŐ ESZTER¹ – GRÁCZI TEKLA ETELKA²

¹NBSZ Szakértői Intézet, ²Nyelvtudományi Kutatóközpont
marko.alexandra.phd@gmail.com, fozo.eszter@nbsz.gov.hu, graczi.tekla.etelka@nytud.hu
<https://orcid.org/0000-0003-0301-7134>, [0000-0003-3351-9661](https://orcid.org/0000-0003-3351-9661)

Markó Alexandra – Főző Eszter – Gráci Tekla Etelka: Alkalmazható-e a diskurzusjelölők vizsgálata a beszélői profilalkotásban?

Discourse markers in speaker profiling for forensic linguistic purposes
Alkalmazott Nyelvtudomány, Különszám: Alkalmazott nyelvészet és kriminalisztika, 2022, 74–93.
doi:<http://dx.doi.org/10.18460/ANY.K.2022.006>

Alkalmazható-e a diskurzusjelölők vizsgálata a beszélői profilalkotásban?

Discourse markers in speaker profiling for forensic linguistic purposes

Discourse markers are a functional group of linguistic elements participating in the organization of discourse: they connect discourse segments and denote pragmatic relationships. In earlier studies, age and gender were found to play a role in the occurrence of various discourse markers and individual language behavior, however, speech tasks were also shown to affect the use of discourse markers.

In linguistic profiling, the expert seeks patterns of the language use of the author or speaker, which refer to certain sociodemographic and other characteristics of the person.

In the present study, we investigated 4 discourse markers whether they show a difference in frequency concerning the speaker's gender, age, and education, and according to speech task.

The results suggest that discourse markers may contribute to profiling to a limited extent and that the speech task must also be involved as a factor. In addition, the individual patterns in the usage of discourse markers cannot be ignored.

Keywords: discourse markers, profiling, sociodemographic variables, speech task

Bevezetés

A profilalkotás „bűnügyi elemző-értékelő tevékenység, amely szűkítheti a számításba vehető elkövetők körét, ezzel elősegítheti a nyomozás további folyamatát és megadhatja annak lehetséges irányait” (Alföldi 2012: 981). Ezt a munkát abban az esetben végzik, ha nincs feltételezett elkövető. A nyelvészeti profilkészítés során a szakértő a kérdéses fogalmazó-szerző vagy beszélő olyan nyelvhasználati, illetve beszéde hangzási sajátosságainak megállapítására törekszik, amelyek alapján behatárolhatók az illető bizonyos szociodemográfiai és egyéb tulajdonságai. Ilyen lehet a kérdéses szerző/beszélő neme, életkora, iskolázottsága, származási helye, illetve a rá jellemző csoportnyelv-változat, idiolektus, hangfelvétel esetén az esetleges beszédhiba is (vö. pl. Tatár, 2013; Tatár et al., 2021).

A diskurzusjelölőket olyan nyelvi elemek funkcionális csoportjaként tartják számon, amelyek a diskurzus szervezésében vesznek részt: diskurzusszegmenseket

kötnek össze, illetve pragmatikai viszonyokat jelölnek. A diskurzusjelölők általában nem befolyásolják a megnyilatkozás igazságfeltételeit, ugyanakkor esetenként rendelkezhetnek emocionális és expresszív funkcióval. Különböző szófajú szavakból alakulhatnak diskurzusjelölővé: lehetnek eredetileg kötőszók, határozószók, partikulák és módosítószók; de többszavas elemek is szolgálhatnak forrásul, például prepozíciós frázisok vagy idiomatikus szókapcsolatok. Elsősorban beszélt nyelvi elemeknek tartják őket, de az írott nyelvben, illetve az írott beszélnyelviség szövegtípusaiban is előfordulnak bizonyos jellemző diskurzusjelölők (Dér, 2009; Schirm, 2010).

A diskurzusjelölők használata és az egyéni nyelvi viselkedés (idiolektus) közötti összefüggések lehetősége korábbi tanulmányokban is felmerült már különböző szempontokból. Kredens (2000) interjúk vizsgálata alapján azt találta, hogy ötből (*like, I mean, kind of / sort of, you know, actually*) három diskurzusjelölő (*like, I mean, kind of / sort of*) használati gyakorisága eltért két beszélő között (idézi Coulthard–Johnson 2007: 170). Coulthard és Johnson (2007) egy bírósági tárgyalásból származó szövegmintán alapján Tony Blair egyik kedvenc diskurzusjelölőjeként a *look* 'nézze' elemet említi. Egy, a gépi elemzők kriminalisztikai célú felhasználhatóságára irányuló magyar nyelvű vizsgálat más szempontok mellett a diskurzusjelölők gyakoriságát is összevetette négy szerző esetében, és kettejükénél nagyobb, kettejükénél kisebb gyakoriságot mutatott ki (Vincze et al., 2021). Dér és Markó (2010) – ugyancsak négy beszélő szövegei alapján – rámutatott a beszélők diskurzusjelölő-használatának egyedi különbségeire, egyes beszélők diskurzusjelölő-használati preferenciáira. Dömötör (2013) is utalt rá, hogy a megnyilatkozó „szavajárása” lehet egy adott diskurzusjelölő preferált használata. Schirm (2019a) szerint „számos információt hordoz a diskurzusjelölők fajtája és gyakorisága a beszélő nyelvszokásáról” (187).

A diskurzusjelölők használata és az egyéni nyelvi viselkedés összefüggése a szövegtípusokon keresztül is megfigyelhető. A diskurzusjelölők fajtái, arányai és szerepei visszatükrözik a szövegtípus formalitását, stílusát, valamint legfőbb jellegzetességeit (Schirm, 2017; Götz, 2019). Schirm (2019b) így fogalmazott: „nem csupán a szövegtípusok intézményessége határozza meg a diskurzusjelölők gyakoriságát, hanem úgy tűnik, hogy a szöveg célja és a beszédhelyzet egyéb körülményei is hatással vannak az elemek eloszlására” (269). A szövegtípus a kriminalisztikai szövegvizsgálatokban (beleértve az írott és a hangzó szövegek vizsgálatát is) igazoltan szerepet játszik. Nemrég Sztahó és munkatársai (2022) a szerzőazonosításhoz és profilalkotáshoz alkalmazott bekezdésvektor-modellezést (Doc2Vec) használtak spontánbeszéd-minták átiratainak elemzésében. A vizsgálatnak az volt a célja, hogy az írott szövegek esetén alkalmazott módszertant teszteljék abból a szempontból, hogy alkalmas lehet-e a beszélőazonosság

megállapításában kiegészítő módszertanként. A hangzó szövegekből készült leiratok eredményei alapján relatíve alacsony hibaarány adódott, amikor azonos típusú szövegeket vetettek össze egymással. Több korábbi kutatással összhangban tehát arra is utalnak az eredmények, hogy az összehasonlítandó minták típusának (beszédhelyzet, műfaj) azonossága vagy különbözősége meghatározó a szerző/beszélőazonosítás sikeressége szempontjából. Mivel a korábbi kutatások különféle terminusokkal illetik, és nem is teljesen azonos módon közelítik meg ezt a tényezőt, illetőleg mivel a jelen vizsgálat egy rendelkezésre álló beszédatbázis részein készült, amelyek a pragmatikai szakirodalom „szövegtípus”-aihoz nehezen társítható, ezen ellentmondások feloldására a továbbiakban a *beszédfeladat* terminust alkalmazzuk a szövegek típusainak átfogó megnevezésére. E beszédfeladatokat a módszertani részben mutatjuk be.

A profilalkotásban a válaszra váró kérdések elsősorban a beszélő/szerző életkorára, nemére és iskolázottságára irányulnak. A magyar diskurzusjelölőket ilyen tekintetben elemző pragmatikai szakirodalomban elsősorban az életkor hatását elemezték. Egyes diskurzusjelölők használati gyakorisága az adatok szerint változik az időben, azaz korosztályfüggő – lásd például Dömötör (2008) megállapítását, amely szerint korábban a *hogy úgy/így mondjam* a leggyakrabban az idősebb korosztályok nyelvhasználatában jelentkezett, majd a XX. század vége felé a *hogy úgy mondjam* szerkezetek helyébe egyebek között az *úgymond* lépett. Markó és Dér (2011) négy diskurzusjelölő, a *hát*, az *így*, az *ilyen* és az *úgyhogy* használatát elemezte 20 fő beszédében, akiknek az életkora 20 és 70 év között volt, és azt találták, hogy a *hát* gyakorisága nőtt az életkorral, míg az *így* gyakorisága csökkent. A másik két diskurzusjelölő esetében nem találtak eltérést a beszélők életkora függvényében. Bóna (2010) adatai szerint a *hát*, az *akkor* és az *ugye* idős (70–85 éves, 10 fő) nőknél gyakrabban fordultak elő, mint a fiataloknál (20–32 év, 10 fő); az *így/úgy*, *ilyen/olyan*, *tehát* használata a fiatalokra volt jellemzőbb. Dér (2021) a *szerintem* gyakoriságának vizsgálatában azt találta, hogy a fiatalabbak minden vizsgált pozícióban szignifikánsan gyakrabban használták ezt a diskurzusjelölőt, mint az idősebbek.

A beszélő nemének és a diskurzusjelölők használatának lehetséges összefüggése ugyancsak kérdésként merült fel egyes elemzésekben. Matei (2011) román nyelvű vizsgálatában megfigyelt nemek közti eltéréseket: eredményei szerint a nők több diskurzusjelölőt használtak, mint a férfiak. Markó és Dér (2011) fent idézett tanulmányában a *hát*, *így*, *ilyen*, *úgyhogy* használatának gyakoriságára nem volt hatása a beszélő nemének. Bóna (2014) azt találta, hogy a beszélő neme csak az életkorral együtt, illetve egyes életkori csoportokban hatott a diskurzusjelölő-használatra, önmagában azonban itt sem volt szignifikáns hatása. Dér (2021) eredményei szerint a férfiak több *szerintem*-et produkáltak, mint a nők, de nem volt

szignifikáns a különbség. A végzettségre vonatkozóan nem találtunk konkrét adatokat a korábbi elemzésekben.

A jelen kutatásban a profilalkotás szempontjait és céljait szem előtt tartva néhány diskurzusjelölő gyakorisági vizsgálatát végeztük el. Az elemzésünk első és átfogó megközelítés, általánosságban arra a kérdésre irányul, hogy a diskurzusjelölők gyakorisági mintázata alapján lehet-e következtetéseket megfogalmazni a profilalkotás során a kérdésként felmerülő szociodemográfiai változókra vonatkozóan, vagyis a vizsgált diskurzusjelölők mutatnak-e gyakorisági eltérést a beszélő neme, életkora, végzettsége tekintetében? A korábbi elemzések e három szempontot együttesen nem vizsgálták, ugyanakkor – mint láttuk – rendelkezésre állnak olyan eredmények, amelyek alapján feltételezhetjük, hogy egyes itt vizsgált tényezők csak valamely másik tényezővel együttesen gyakorolnak hatást az előfordulási gyakoriságra (vö. Bóna, 2014). Vizsgálatunk alapján azt a kérdést is meg kívánjuk válaszolni, hogy a nyelvészeti profilalkotásban alkalmazhatók-e a diskurzusjelölők. Továbbá vizsgáljuk azt a szempontot is, hogy van-e hatása a diskurzusjelölők gyakoriságára a beszédfeladatnak. Ha a válasz igen, abból arra következtethetünk, hogy ezt a tényezőt is figyelembe kell venni a profilalkotásban.

Anyag, kísérleti személyek és módszer

A BEA adatbázis (Gósy, 2012) 78 beszélőjének anyagából spontán és félspontán beszédfeladatok hangfelvételeinek lejegyzett szövegét (Mihajlik et al., 2022) elemeztük. A vizsgált beszélők fele férfi, fele nő. Az életkor tekintetében a kriminalisztikai szempontból releváns intervallumba (Tatár, 2013; Skarnitzl–Vaňková, 2017) tartozó beszélőket válogattunk ki, és nemenként egyenlő számú, azaz 13-13 fős csoportokra osztottuk őket a következő életkori sávok szerint: a 21–25, a 28–39 és a 41–59 éves korúak csoportjába. A végzettség szerint a 78 beszélő közül 32 középfokú, 46 felsőfokú végzettségű volt. Habár a BEA adatközlőinek egyenlőtlen eloszlása miatt nem tudtunk minden csoportba azonos számú közép- és felsőfokú beszélőt beválogatni, ügyeltünk arra, hogy egy-egy nemek és életkor szerint illesztett csoportpárban azonos számban legyenek mind a két végzettséggel rendelkező beszélők (lásd 1. táblázat).

1. táblázat. A résztvevők jellemzői

Férfiak			Nők		
	Középfokú	Felsőfokú		Középfokú	Felsőfokú
21–25 év	8	5	21–25 év	8	5
29–39 év	3	10	28–39 év	3	10
41–59 év	5	8	42–59 év	5	8

Az elemzéshez a magyar nyelvre vonatkozó szakirodalomban kifejezetten gyakorként dokumentált diskurzusjelölőket választottunk ki. Homofónia esetén az egyes elemek diskurzusjelölői és nem diskurzusjelölői funkciója közötti különbség megragadása nehézségekbe ütközhet (Dér–Markó, 2010), ezért olyan elemeket választottunk, amelyek esetében a szóalak (szinte) csak diskurzusjelölői funkcióban jelentkezik. Ennélfogva a szóalakok sikerrel detektálhatók automatikus módszerrel.

A *hát* a különféle korpuszokban a leggyakoribb vagy a második leggyakoribb diskurzusjelölő (Dér–Markó, 2010; Schirm, 2015). Ez az egyik leginkább multifunkcionálisnak talált elem, a diskurzusban betöltött szerepei között említi a szakirodalom a hezitálást, a válaszjelölést, a mondanivaló továbbvitelét, az önjavítást, a magyarázkodást, az evidencijelölést, a következtetést, az ellentét kifejezését, a nyomatékosítást, az attitűdjelölést és így tovább (Schirm, 2017: 2–3).

Az *ugye* korpusztól függően a leggyakoribb vagy a második-harmadik leggyakoribb diskurzusjelölőként jelenik meg (Dér–Markó, 2010; Schirm, 2019b; Molnár, 2019). Funkciói között szerepel az evidencialitás kifejezése, illetve utal a feltételezett (ismert vagy kikövetkeztethető) közös tudásra (pl. Gyuris, 2008); de szolgál kapcsolattartásra is (Schirm, 2019b).

Az *egyébként* szintén viszonylag gyakori, az egyes elemzésekben az összes diskurzusjelölő közül beszédhelyezettől/-típustól függően az első tízben szerepel (Schirm, 2017, 2019b). Funkcióiként elsősorban a téma- és beszédaktusváltást és a mellékinformáció közlését határozzák meg (Schirm, 2019b).

A *szerintem* sokat vizsgált diskurzusjelölő, viszonylag gyakori (bizonyos helyzetekben az első tízben szerepel; Schirm, 2019b). Funkciója lehet a megnyilatkozás aktusa; a következtetés vagy a valószínűsítés; illetve az értékelés, minősítés is (Kugler, 2015).

A diskurzusjelölőket automatikusan detektáltuk a lejegyzett szövegekben. Elemzésük pusztán a gyakoriságukra irányult, nem végeztünk funkcionális vizsgálatot, és nem voltunk tekintettel a kontextusra, így például arra sem, ha az adott diskurzusjelölő más diskurzusjelölőkkel való társulásban jelent meg (akkor sem, ha a társulás másik tagja is a vizsgált elemek között volt, például *egyébként szerintem* – ez esetben külön-külön rögzítettük az egyes példányokat az adott diskurzusjelölő adatai között).

A BEA-felvételek öt beszédfeladatát használtuk fel. Az „interjú” főként kvázi-monologikus spontán beszéd, amelyben a beszélő a saját munkájáról/tanulmányairól, hobbijáról, családjáról beszél. A „vélemény” feladatban a beszélő kap egy hétköznapi témát (pl. karácsonyi készülődés és ünnepek, BKV helyzete és elégedettség, a H1N1 kapcsán a járványok és egészségügyi helyzet), amelyről kifejti véleményét. A témától eltérhet, ha úgy alakul. A felvételvezető ebben a felvételi egységben gyakrabban szólal meg, de nem cél a kétfős társalgás létrejötte, ha az

adatközlő elég közlékeny. A „növény” feladatban és a „Varkocs” feladatban a beszélők meghallgatnak egy-egy rövid felolvasást, amely az előbbi esetben egy ismeretterjesztő szöveg, az utóbbi esetben pedig egy történelmi anekdota, és ezeket kell minél pontosabban összefoglalniuk. Ezt a két felvételi egységet egy általánosított kevert modell eredményei alapján összevontuk, és „összefoglalás”-ként nevezzük meg a továbbiakban. (Egy fő esetében a „Varkocs” feladatot nem rögzítették.) A modellt a statisztikai leírásokban fejtjük ki. Felhasználtuk a „társalgás” feladatot is, amelyben a felvételvezetőhöz és az adatközlőhöz egy harmadik személy, a felvételvezető egy kollégája csatlakozik. Ebben a feladatban egy háromfős társalgás kialakulása a cél. A téma ismét egy hétköznapi esemény köré épül, amely beszélőn belül eltér a „vélemény” feladatban kapott témától.

A lejegyzett szövegekből kinyertük az egyes diskurzusjelölők előfordulásának számát és a teljes szószámot. A megakadásjelenségek (vö. Gósy, 2002) közül a téves kezdést és az újraindítást kizártuk, a szünet a szóban jelenséggel megszakított szóalakot egy szóként vettük számításba. A nem töredékszóval járó hiba típusú megakadásjelenségeket (pl. ismétlés, téves szó) beszámítottuk a szószámba. Az így számított teljes szószámra vetítettük az elemzett diskurzusjelölők gyakoriságát százalékban az adott beszédfeladaton belül.

A statisztikai elemzéseket az R szoftverben végeztük (R Core Team, 2021). Kendall-féle rangkorrelációt és általánosított kevert lineáris modelleket alkalmaztunk az előfordulások lognormál közeli eloszlása miatt. A konfidencia-intervallumot 95%-ban határoztuk meg.

Vizsgáltuk a diskurzusjelölők száma és a szószám korrelációját az egyes beszédfeladatokban, valamint a diskurzusjelölők gyakoriságának (%) korrelációját az egyes beszédfeladatok között, továbbá az egyes diskurzusjelölők között is.

A diskurzusjelölők gyakoriságát (függő változó) elemeztük egyrészt diskurzusjelölőnként külön a beszédfeladat, az életkori csoport, a nem és a végzettség (fix hatások) függvényében; másrészt együtt ugyanezen változók és a diskurzusjelölő mint fix hatások függvényében. Az életkori csoportot, a nemet és a végzettséget együttesen *profilfaktor*ként említjük. A logaritmikus eloszlás miatt az általánosított kevert modelleket (glmm) választottuk (lme4: Bates et al., 2015; lmerTest: Kuznetsova et al. 2017). Minden modell esetében szükséges volt az alap Nelder-Mead optimalizáció helyett „BOBYQA”-t (Bound Optimization by Quadratic Approximation) alkalmaznunk, illetve az iterációk számát 1000000-ra állítottuk. A faktorok közötti interakciót minden modellben engedélyeztük.

A gyakori 0 értéket kezelni kellett a modellekben. Erre egy gyakran alkalmazott eljárást választottunk: A legkisebb, de 0-nál nagyobb érték felét adtuk a 0 előfordulásokhoz. Mivel egy szó előfordulásának valószínűsége valójában nem 0, legfeljebb tart 0-hoz, ezért feltételezzük, hogy ez a módszer ebben az esetben is

alkalmazható. A modellekben a százaléktételekkel dolgoztunk, mert a darabszám (mint a korrelációs elemzések mutatták) összefügg a szószámmal, és így a különböző beszéd típusokban mutatott korrelációkkal is jobban el tudtunk számolni a modellekben. Mindkét modellben log-transzformációval dolgoztunk (Gamma(link = "log") beállítással).

A diskurzuszjelölőket együtt vizsgáló modellben random hatásként a diskurzuszjelölőkre állított random meredekséget szerepeltettük beszélőnként random konstanssal. Mivel néhány beszélőtől hiányzott 1-1 beszéd feladat felvétele vagy az ahhoz tartozó lejegyzés, ezért a beszéd feladatra nem tudtunk random meredekséget illeszteni. Ebben a vizsgálati részben nem alkalmaztunk modellszelekciót, mivel az egyes fix faktorok hatásának felmérése volt a célunk. Az első változatban a kétféle „összefoglalás” külön szerepelt. A Tukey-féle post hoc teszt alapján (emmeans: Lenth, 2021) minden beszéd feladat-pár között szignifikáns eltérést kaptunk, kivéve a két összefoglalást, így ezeket összevontuk. A modellt újrafuttattuk az így már csak négyelemű beszéd feladat faktorról, illetve a további modelleket már csak így készítettük el.

A diskurzuszjelölőnkénti külön modellek esetében az alábbi módon jártunk el. A random hatások között csak beszélőnkénti random konstans szerepeltettünk, random meredekséget (a néhány beszélőnél hiányzó 1-1 beszéd feladat miatt) nem. A modellszelekció során egy-egy fix hatást eltávolítottunk, amennyiben a bővebb modellben nem kaptunk rá szignifikáns hatást. A modelleket anova()-val vetettük össze. Amennyiben a bővebb modelltől nem különbözött szignifikánsan, a (lehető) legkisebb AIC-számú modellt választottuk (Akaike, 1974). A szignifikanciaszinteket az Anova() segítségével (car, Fox–Weisberg 2019), a hatásnagyságokat az sjstats csomag (Lüdecke, 2021) segítségével nyertük ki. A több mint kétszintű faktoroknál Tukey-féle post hoc tesztet végeztünk (emmeans: Lenth, 2021).

Az ábrákon (ggplot2: Wickham, 2016) nem a logtranszformált, hanem a mért gyakoriságokat ábrázoltuk, mivel azok így jobban átláthatóak.

Eredmények

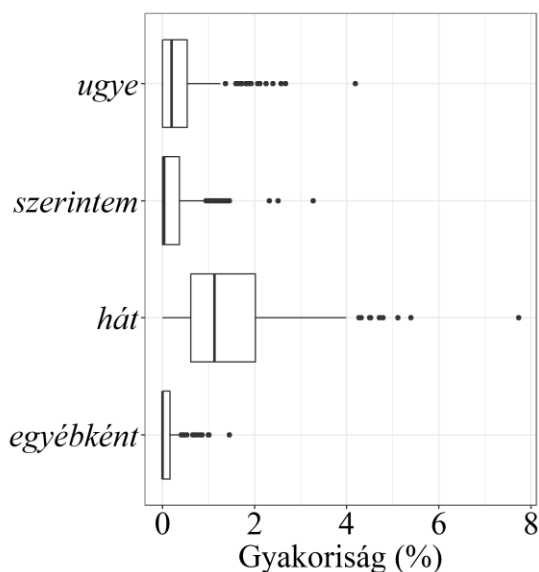
Összesen 194531 szóból 4301-et, azaz 2,21%-ot tett ki a négy vizsgált diskurzuszjelölő. Gyakori volt, hogy egy adott diskurzuszjelölő nem fordult elő néhány beszélőnél egy-egy beszéd feladatban. A *hát* előfordulása volt a legmagasabb. Szinte minden beszélőnél minden beszéd feladatban megjelent, és ezen belül is nagyobb arányban (1. ábra), mint a többi elemzett diskurzuszjelölő. A legkevesebb előfordulást az *egyébként* esetében kaptuk, a beszélők közül is kevesebben használták, és ők is ritkábban, mint a *hát*-ot.

A diskurzuszjelölők előfordulási gyakorisága, azaz darabszáma gyenge vagy közepesen erős korrelációt mutatott az adott beszéd minta (az adott beszélőnek az

adott beszéd feladatban rögzített produkciója) összes szószámával (2. táblázat; Kendall-féle rangkorreláció) kivéve a *szerintem* megjelenését az összefoglalókban és a társalgásban.

A beszéd feladatok között a *hát* minden esetben és az *ugye* majdnem minden esetben szignifikáns korrelációt mutatott. Ez a korreláció gyenge, legfeljebb közepes erősségű, de azt jelzi, hogy aki az egyik beszéd feladatban gyakrabban használja az adott diskurzusjelölőt, az a másikban is. Az *ugye* esetében az összefoglaló és a társalgás tért el ebből a szempontból, ezekre nem találtunk ilyen összefüggést. Az *egyébként* és a *szerintem* esetében kevesebb beszéd feladtpár között találtunk szignifikáns korrelációt, ami arra enged következtetni, hogy ezek használata elsősorban a beszéd feladattól függ, ami a funkciójukkal függhet össze, a beszélőn belüli gyakoriságuk ehhez képest korlátozottabban kikövetkeztethető (3. táblázat, Kendall-féle rangkorreláció).

1. ábra. A vizsgált diskurzusjelölők előfordulási gyakorisága a teljes anyagban



2. táblázat. A diskurzusjelölők darabszáma és a beszédminta teljes szószáma közötti Kendall-féle rangkorreláció (τ) eredménye (a szürke cellák a szignifikáns korrelációt jelölik)

	Interjú	Összefoglaló	Társalgás	Vélemény
<i>egyébként</i>	0,335	0,202	0,363	0,299
<i>hát</i>	0,419	0,183	0,459	0,404
<i>szerintem</i>	0,356	0,072	0,056	0,238
<i>ugye</i>	0,382	0,232	0,532	0,267

3. táblázat. A diskurzusjelölők egyes beszéd típusok közötti gyakorisága (%) Kendall-féle rangkorrelációjának τ -értékei (a szürke cellák a szignifikáns korrelációt jelölik)

	<i>hát</i>	<i>egyébként</i>	<i>ugye</i>	<i>szerintem</i>
Interjú – vélemény	0,496	0,190	0,314	-0,005
Interjú – összefoglaló	0,319	-0,029	0,274	0,247
Vélemény – összegzés	0,287	0,248	0,314	0,093
Interjú – társalgás	0,394	0,143	0,219	0,084
Összefoglaló – társalgás	0,274	0,089	0,162	0,146
Vélemény – társalgás	0,316	0,252	0,360	0,222

Az is elmondható, hogy bizonyos diskurzusjelölők gyakorisága összefügg az egyes beszélőkön belül. A 4. táblázat az elemzett diskurzusjelölők gyakorisága közötti Kendall-féle rangkorrelációkat mutatja be. A korreláció az *egyébként* és a *szerintem*, a *hát* és a *szerintem*, illetve a *hát* és az *ugye* között szignifikáns, bár a *hát* és a *szerintem* között a társalgásban nem. Ezek a korrelációk azonban gyengék, kevésbé jelentősek tehát, mint az azonos diskurzusjelölőkön belüliek. Emellett ugyanakkor kiemeljük, hogy a beszéd feladatok közötti korreláció tekintetében az *egyébként* és a *szerintem*, illetve a *hát* és az *ugye* viselkedett hasonlóan, amit az egymással való korrelációjuk együtthatója megerősített.

4. táblázat. Az egyes diskurzusjelölők gyakorisága (%) közötti Kendall-korreláció (a szürke cellák a szignifikáns korrelációkat jelölik)

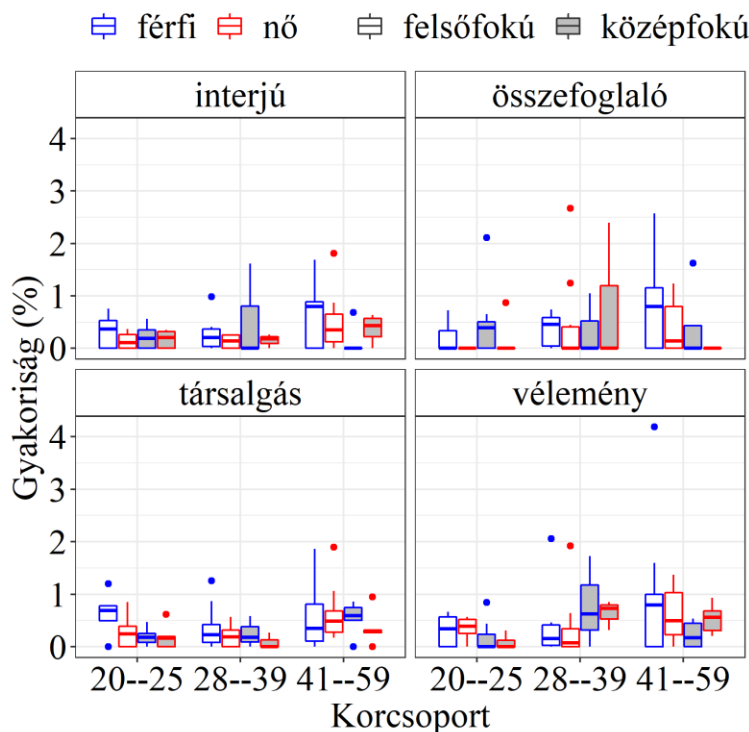
	Interjú	Összefoglaló	Társalgás	Vélemény
<i>egyébként – hát</i>	0,077	0,077	0,066	0,077
<i>egyébként – szerintem</i>	0,126	0,126	0,125	0,126
<i>egyébként – ugye</i>	0,042	0,042	0,04	0,042
<i>hát – szerintem</i>	0,087	0,087	0,079	0,087
<i>hát – ugye</i>	0,157	0,157	0,155	0,157
<i>szerintem – ugye</i>	0,038	0,038	0,041	0,038

Annak vizsgálatára, hogy a felvett változók hogyan befolyásolták a választott diskurzusjelölők előfordulását, kétféle módon készítettünk modellt. Az elsőben a négy diskurzusjelölőt együtt vizsgáltuk és a fix hatások között is szerepeltettük, valamint random meredekséget illesztettünk rá beszélőnként, random konstanssal. Ebben a modellben tehát öt fix hatás szerepelt: a szó maga, a korcsoport, a nem, a végzettség és a beszéd feladat. Az öt faktor interakciója megengedett volt. A modell eredményei alapján az öt faktor interakciója szignifikáns hatással bírt ($\chi^2(18) = 79991024$, $p < 0,001$, $r^2_m = 0,611$, $r^2_c = 0,734$). Az egyes diskurzusjelölők közötti

eltérést Tukey-féle post hoc teszttel vetettük össze. Minden diskurzusjelölő között szignifikáns eltérést találtunk. A választott diskurzusjelölőkről az eddig ismertett eredmények alapján azt mondhatjuk, hogy bár vannak olyan párok, amelyek gyakorisága összefügg a beszélőn belül, ugyanakkor nem mindegyiké, és a kérdéses faktorok mentén eltérő a viselkedésük. Mindezek miatt ahhoz, hogy a profilalkotásban betölthető szerepüket megvizsgálhassuk, külön-külön elemezzük őket a továbbiakban.

Az *ugye* esetében a legbővebb modell (fix hatások: nem, korcsoport, végzettség és beszédfeladat interakcióval, illetve beszélőnkénti random konstans) jellemezte a legjobban az adatokat. A négy fix hatás interakciója szignifikánsan meghatározónak bizonyult ($\chi^2(6) = 13,496$, $p = 0,036$, $r^2_m = 0,233$, $r^2_c = 0,634$). Fontos kiemelni a marginális és kondicionális hatásnagyságok közötti különbséget, ami arra utal, hogy a beszélők közötti nagy variabilitás figyelembevétele a random konstansban jelentősen megnöveli a modell hatékonyságát. Vagyis nagy a beszélők közötti variabilitás. A 2. ábrán szemléltetjük az *ugye* gyakoriságát a különféle beszélői csoportokban a három profilmfaktor alapján együttesen, míg az 5. táblázatban egy-egy profilmfaktor alapján bontva. Az eredmények alapján az *ugye* diskurzusjelölőt a férfiak valamivel gyakrabban használták, mint a nők: három beszédfeladatban (interjú, összefoglaló, vélemény) a gyakoribb előfordulások a legidősebb vizsgált korcsoportban a felsőfokú végzettségű férfiak és a középső korcsoportban a középfokú végzettségű férfiak esetében fordultak elő, a társalgásban pedig csak a legidősebb, felsőfokú végzettségű férfiak körében. Kiugróan magas értékek minden beszédfeladatban, minden korcsoportban megjelentek. Érdeemes megjegyezni, hogy a legidősebb korcsoportban a középfokú és a felsőfokú végzettségű férfiak között, illetve a középső korcsoportban a középfokú végzettségű nők és férfiak között jellemzően eltérőek a gyakorisági mutatók. Az egy-egy profilmfaktor alapján történő elemzés (5. táblázat) alapján az életkori csoportok között van eltérés: a legidősebb korcsoportban három beszédfeladatban is másfél-kétszer gyakoribb volt az *ugye* használata, mint a két másik korcsoportban.

2. ábra. Az *ugye* gyakorisága (%) az egyes beszédfeladatokban a három profilfaktor (nem, korcsoport és végzettség) függvényében

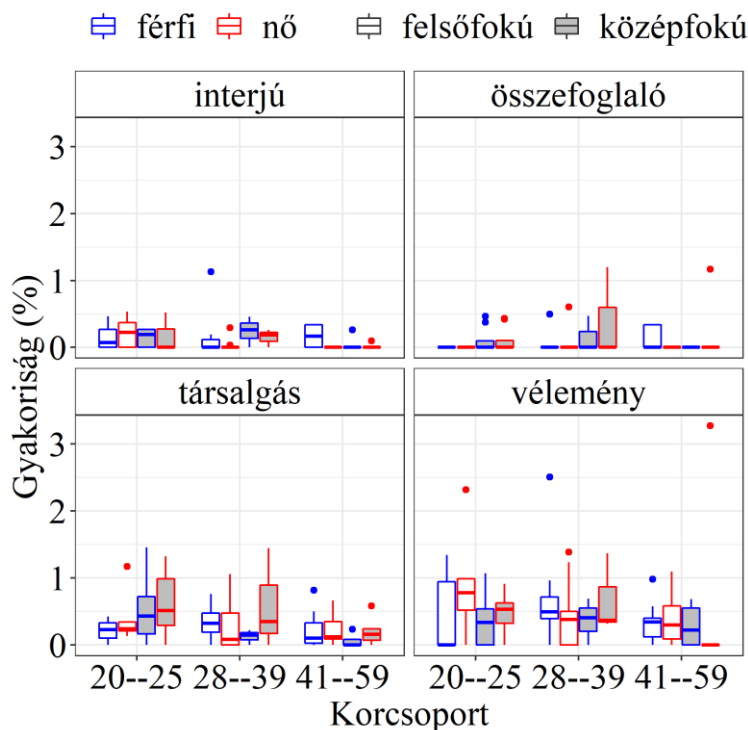


5. táblázat. Az *ugye* gyakorisága (%) az egyes tényezők függvényében

		Interjú		Összefoglaló		Társalgás		Vélemény	
		medián	IQR	medián	IQR	medián	IQR	medián	IQR
Életkor	20–25	0,196	0,352	0	0,352	0,189	0,468	0,034	0,376
	28–39	0,186	0,265	0,090	0,588	0,185	0,324	0,210	0,585
	41–59	0,330	0,768	0	0,795	0,475	0,538	0,490	0,649
Nem	férfi	0,188	0,547	0,358	0,691	0,290	0,507	0,163	0,596
	nő	0,221	0,326	0	0,143	0,241	0,475	0,292	0,575
Végzettség	felsőf.	0,222	0,464	0	0,708	0,305	0,478	0,296	0,624
	középf.	0,190	0,348	0	0,430	0,189	0,427	0,185	0,542

A *szerintem* esetében is a legbővebb modell bizonyult a legjobbnak, és ismét szignifikáns hatást mutatott a négyes interakció ($\chi^2(6) = 2236691,7, p < 0,001, r^2_m = 0,547, r^2_c = 0,639$). A 3. ábra alapján kevésbé látható csoportspecifikus mintázat a *szerintem* használatában.

3. ábra. A szerintem gyakorisága (%) az egyes beszédfeladatokban a három profilmfaktor (nem, korcsoport és végzettség) függvényében



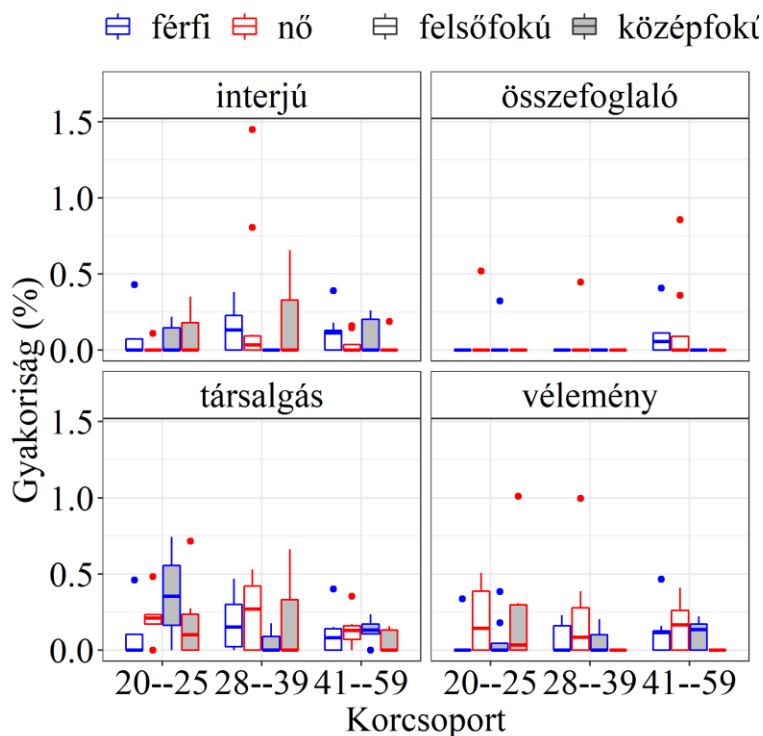
A vélemény beszéd típusban értelemszerűen magasabb értékek is megjelennek minden korcsoportban az interjúhoz és az összefoglalóhoz képest. Ehhez hasonlóan a társalgásban is több beszélői csoportban magasabb az előfordulási arány. Ugyanakkor ezek a különbségek inkább a beszélők közötti nagyobb variabilitásból adódnak, mintsem a csoportok közötti eltérésekből. Az összefoglaló beszéd típus esetében találunk csoportspecifikusabb mintázatot, a középső korcsoport középfokú végzettségű női beszélői esetében jelentek meg gyakrabban az előfordulások. Érdeemes ugyanakkor megemlíteni, hogy az ő esetükben nem figyelhetők meg a többi csoporthoz képest magasabb értékek a társalgásban és a véleményben. A három profilmfaktor együttes (3. ábra) és az egyesével történő elemzése (6. táblázat) esetében is a beszéd feladatok között találtunk lényeges eltérést.

6. táblázat. A szerintem gyakorisága (%) az egyes tényezők függvényében

		Interjú		Összefoglaló		Társalgás		Vélemény	
		medián	IQR	medián	IQR	medián	IQR	medián	IQR
Életkor	20–25	0,118	0,274	0	0	0,340	0,339	0,478	0,764
	28–39	0	0,167	0	0	0,220	0,472	0,433	0,417
	41–59	0	0,071	0	0	0,101	0,238	0,213	0,539
Nem	férfi	0,073	0,266	0	0	0,212	0,404	0,371	0,628
	nő	0	0,065	0	0	0,238	0,429	0,386	0,708
Végzettség	felsőf.	0	0,169	0	0	0,190	0,328	0,389	0,775
	középf.	0	0,262	0	0	0,235	0,505	0,346	0,624

Bár az *egyébként* esetében is a legbővebb modell bizonyult a legjobbnak ($r^2_m = 0,322$, $r^2_c = 0,547$), a négyes interakció csak tendenciaszintű ($\chi^2(6) = 11,514$, $p = 0,074$), míg a nem, a végzettség és az életkor interakciója ($\chi^2(3) = 13,907$, $p = 0,003$) és az életkor és a végzettség interakciója ($\chi^2(2) = 6,463$, $p = 0,034$), valamint az életkor és beszédfeladat interakciója ($\chi^2(2) = 17,022$, $p = 0,009$) is szignifikáns hatással bírt. (Az életkor további interakciókban ugyancsak tendenciaszintű hatással bírt.) Az összefoglaló beszéd típusban alig-alig jelent meg az *egyébként* diskurzusjelölő (4. ábra), a vélemény beszéd típus esetében a legfiatalabb vizsgált korosztályban a nőkre volt jellemző az előfordulása, míg a többi korcsoportban és az interjúban, valamint a társalgásban nagyon változatos mintázatot eredményezett. A társalgás esetében a gyakoribb előfordulás a legfiatalabb vagy a középső korosztályban jellemző, de ezen belül a mintázatok változatosak. A profilmfaktorokat együttesen (4. ábra), illetve egyesével elemezve (7. táblázat) egyrészt a beszéd feladatok közötti eltérő gyakoriság erősíthető meg, valamint az, hogy a beszélők közötti variabilitás jelentősége erőteljesebb, mint a profilmfaktorok egyes szintjei alapján felállítható tendenciák.

4. ábra. Az egyébként gyakorisága (%) az egyes beszédfeladatokban a három profilmfaktor (nem, korcsoport és végzettség) függvényében



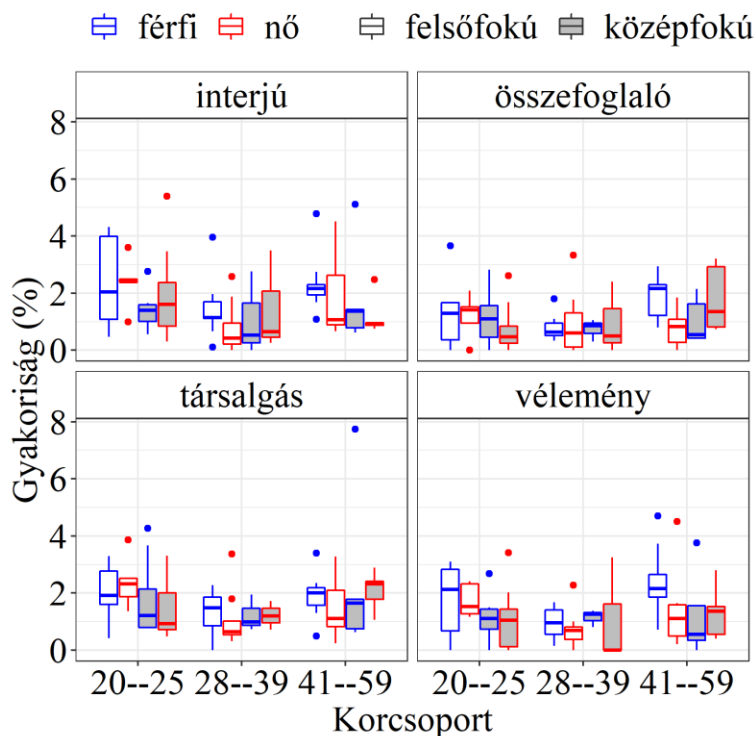
7. táblázat. Az egyébként gyakorisága (%) az egyes tényezők függvényében

		Interjú		Összefoglaló		Társalgás		Vélemény	
		medián	IQR	medián	IQR	medián	IQR	medián	IQR
Életkor	20–25	0	0,122	0	0	0,196	0,386	0	0,264
	28–39	0	0,205	0	0	0,162	0,305	0	0,194
	41–59	0	0,138	0	0	0,112	0,155	0,057	0,167
Nem	férfi	0	0,184	0	0	0,118	0,300	0	0,161
	nő	0	0,107	0	0	0,147	0,270	0	0,271
Végzettség	felsőf.	0	0,138	0	0	0,139	0,287	0,057	0,223
	középf.	0	0,147	0	0	0,132	0,255	0	0,143

A hát esetében az a modell bizonyult a legjobbnak, amelyben a végzettség nem szerepelt, tehát az életkor, a nem és a beszédfeladat szerepelt a fix hatások között. A hármas interakció szignifikáns hatásúnak bizonyult ($\chi^2(6) = 107146, p < 0,001, r^2_m = 0,175, r^2_c = 0,404$). Meg kell azonban jegyezni, hogy a marginális hatásnagyság viszonylag alacsony, vagyis jelentős további hatások állhatnak a háttérben, és mindemellett nagyban javítja az eredményeket a random hatás, tehát jelentős a

beszélők közötti variabilitás. A *hát* gyakorisága jellemzően nagyobb volt, mint a másik három vizsgált diskurzusjelölőé (5. ábra).

5. ábra. A *hát* gyakorisága (%) az egyes beszédfeladatokban a három profilmfaktor (nem, korcsoport és végzettség) függvényében



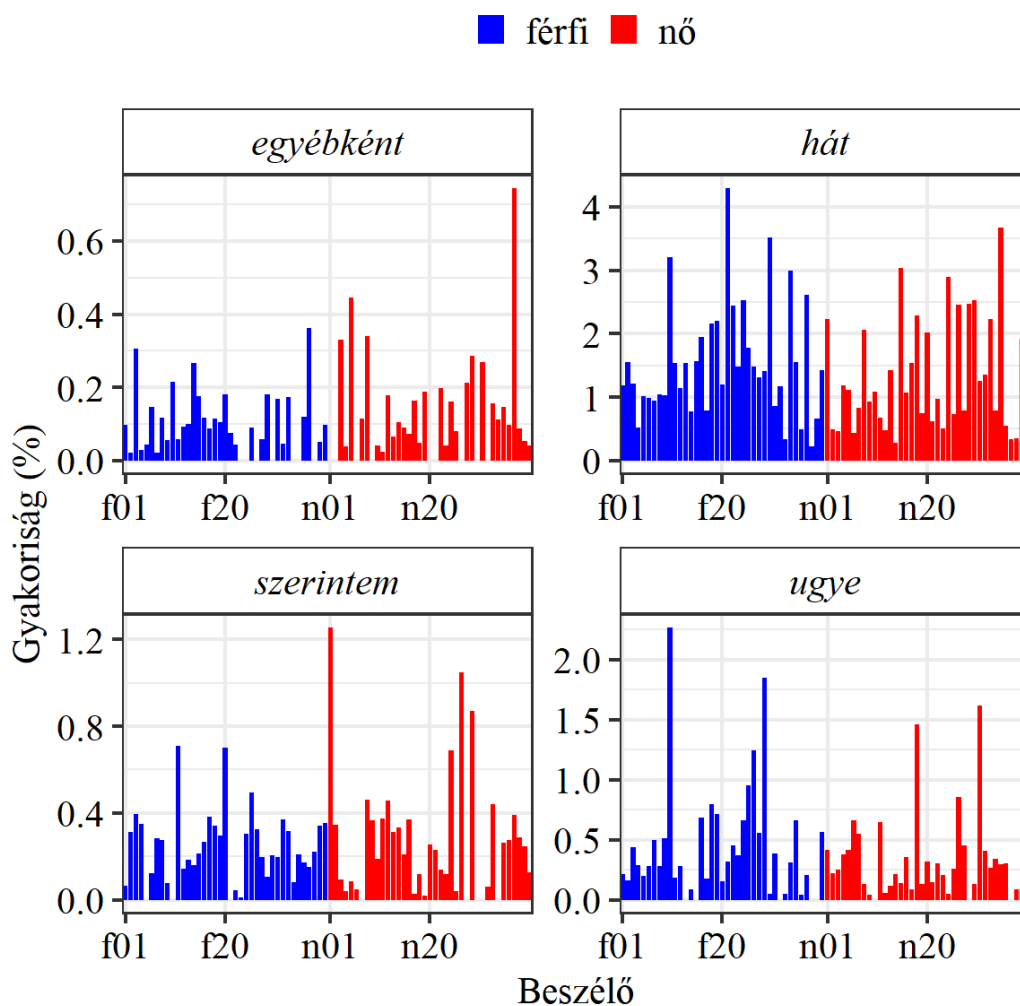
Mint fentebb láttuk, majdnem minden beszélő majdnem minden beszédfeladatában szerepelt, de emellett, ahol szerepelt, gyakoribb is volt. A felsőfokú végzettségűek esetében elmondható, hogy a gyakoribb használat a legfiatalabb és a legidősebb vizsgált korcsoportokban jellemző, és a felsőfokú végzettségűek esetében gyakran egy korcsoporton belül a férfiaknál jelennek meg a magasabb gyakoriságok. A profilmfaktorok alapján egyesével elemezve az adatokat (8. táblázat) nem rajzolódott ki egyik faktor esetében sem a faktorszintek között egyértelmű tendencia, vagyis hasonló a mintázat az egyes beszédfeladatokban megfigyelhető gyakoriságokban.

Mindezek alapján felmerül a kérdés, hogy amennyiben a diskurzusjelölő-gyakoriságokat a vizsgált szociodemográfiai szempontok egyike sem határozza meg önmagában, milyen egyéni mintázatokat találunk, amelyek hasznosulhatnak az elemzésben. Mivel ennek a kérdéskörnek a tárgyalása nem volt a jelen tanulmány célja, csak egy rövid kitekintés erejéig, egy összegző ábrán mutatjuk be a négy vizsgált diskurzusjelölőnek a szószámhoz viszonyított gyakoriságát a beszédfeladatok különválasztása nélkül, beszélőként (6. ábra).

8. táblázat. A *hát* gyakorisága (%) az egyes tényezők függvényében

		Interjú		Összefoglaló		Társalgás		Vélemény	
		medián	IQR	medián	IQR	medián	IQR	medián	IQR
Életkor	20–25	1,614	1,446	1,004	1,249	1,633	1,703	1,240	1,379
	28–39	0,938	1,405	0,635	0,615	1,014	1,075	0,792	0,790
	41–59	1,399	1,412	1,202	1,442	1,784	1,290	1,448	1,598
Nem	férfi	1,423	1,080	1,053	1,254	1,605	1,233	1,248	1,450
	nő	0,959	1,732	0,763	1,152	1,194	1,597	1,009	1,168
Végzettség	felsőf.	1,199	1,466	0,951	1,198	1,595	1,399	1,062	1,505
	középf.	1,168	1,126	0,776	1,211	1,489	1,346	1,062	1,179

6. ábra. Az egyéni gyakorisági adatok alakulása (a beszédfeladatonkénti gyakoriságok átlaga, %)



Következtetések

Vizsgálatunk elsőként elemezte a diskurzusjelölők használatát a profilalkotás nézőpontjából. Három csoportfaktort (nem, életkor, végzettség) elemeztük négy beszédfeladatban (interjú, összefoglaló, vélemény, társalgás), hogy van-e, és ha igen, milyen eltérés a diskurzusjelölők használatában ezen faktorok mentén.

Az elemzett négy diskurzusjelölő eltért a tekintetben, hogy az összes szószámmal korrelál-e a gyakoriságuk, illetve, hogy az egyes beszéd típusok közötti használati gyakoriságuk korrelál-e. A *szerintem* és az *egyébként*, illetve az *ugye* és a *hát* mutattak hasonló mintázatot a szószámra vetített gyakoriságban, és az egymáshoz képesti előfordulási gyakoriságban is (az utóbbiak mutattak nagyobb előfordulási arányt). E párokon belül a két-két diskurzusjelölő gyakorisága is gyenge korrelációt mutatott. A két pár esetében vélhetően a funkciójuk hasonlósága alapján jelent meg hasonló gyakorisági mintázat.

A három profilmfaktor és a beszéd feladat nem önmagukban, hanem különböző interakciókombinációkban mutattak szignifikáns hatást a statisztikai elemzésben, azaz nem jelenthető ki, hogy például az életkor, a nem vagy a végzettség mentén lenne jellemző eltérés egy-egy diskurzusjelölő használatában. Ez azt jelenti, hogy nem bizonyultak jó profiljegyeknek a vizsgált diskurzusjelölők előfordulásai. A beszélők közötti variabilitás ugyanakkor jelentős volt, ami arra enged következtetni, hogy az egyéni nyelvhasználat leírásakor felhasználhatók a diskurzusjelölők gyakorisági mutatói.

Az eredmények megerősítik azt a korábbi kutatások által sugallt feltételezést, hogy a diskurzusjelölők használata idiolektális sajátosságokat mutat, mint ahogyan azt is, hogy ez nem független más tényezőktől, például a beszéd feladattól. Az adatok azt mutatják, hogy összességében az egyes diskurzusjelölők gyakorisága jellemző az adott beszélőre a beszéd típusától függetlenül, ami fontos adalék az adott szerző/beszélő nyelvhasználatának szemantikai és pragmatikai sajátosságainak feltérképezéséhez.

Azt találtuk, hogy az összesített szószámmal gyenge-közepes szignifikáns összefüggésben áll a diskurzusjelölők előfordulási gyakorisága, ami arra utal, hogy a beszéd minta terjedelmétől függ az egyes diskurzusjelölők gyakorisága. A gyakorlatban ez azt jelenti, hogy a hosszabb beszéd minták esetében lehet sikeresebben alkalmazni a diskurzusjelölők elemzését a profilalkotásban – ez egybevág más, a profilalkotásban, illetve a beszélő- és szerzőazonosításban alkalmazott nyelvi és beszéd jellemzők kapcsán tapasztaltakkal.

Az *ugye* esetében látszólagos életkori tendencia fedezhető fel, az idősebb korcsoportokban valamivel gyakoribb a használata. Ez a tendencia azonban a nemek és a végzettségek figyelembevételével, valamint egyértelműen csak egyének közötti eltérésként értelmezhető, mivel a hármast csoportbontásban nem rajzolódik ki

következetes csoportszintű gyakorisági mintázat.

Kismértékben láttunk csoportszintű eltéréseket a *szerintem* előfordulásában, a beszédfeladatok tekintetében azonban kirajzolódtak tendenciák. A *szerintem* várható módon a kvázimonologikus véleményalkotás beszéd típusában volt felülreprezentált, de ugyancsak gyakori volt a társalgásban, ahol inkább előfordulhat, hogy a beszélőnek a többi résztvevő álláspontjához képest kell meghatározni a sajátját. További kutatásokban érdemesnek tartjuk megvizsgálni például, hogy milyen egyéb szavakat, szókapcsolatokat találunk, amelyek az adott diskurzusjelölő funkciójában állnak. A vélemény beszéd típusban az *úgy látom / azt látom* szókapcsolattal, valamint a *véleményem* szóalakkal is találkozhatunk, ám mindkettőt mindössze két-két beszélő használta, így ezek relevanciája elhanyagolható. Ugyanakkor az *úgy gondolom / azt gondolom* szerkezetek előfordulása kiemelkedő, sőt a teljes szósámhoz viszonyítva a szerkezetek előfordulása gyakoribb, mint a *szerintem* diskurzusjelölőé: 194531 szóból 53-szor, ezzel szemben a *szerintem* 47-szer fordult elő összesen.

Az *egyébként* gyakoriságára ugyancsak több tényező együttesen volt hatással. Az *egyébként* diskurzusjelölő ritka megjelenése az összefoglaló beszéd feladatban feltehetőleg azzal függ össze, hogy ebben a beszéd típusban ritka a témaváltás, a témáról való asszociáció vagy az egyéb olyan mozzanat, amely az *egyébként* használatát indokolná.

A korábbi vizsgálatokban általában leggyakoribbnak talált *hát* itt is felülreprezentált volt. Esetében csak a végzettségnek nem volt szerepe a gyakoriság alakulásában, miközben sok tényező együttesét feltételezhetjük a megjelenés háttérében, így itt is a beszélők közötti variabilitás tűnt a legmeghatározóbb tényezőnek.

Összegezve: a diskurzusjelölők a jelen vizsgálat alapján a profilalkotáshoz nem, ugyanakkor az egyéni nyelvhasználat, az idiolektus feltérképezéséhez sokkal inkább hozzájárulhatnak. A beszéd feladat erősen meghatározó tényező, azaz figyelembe kell venni a szövegnyelvészeti szempontú beszélőazonosításnál is, ha eltérő beszéd feladatok között zajlik az összevetés. Az itt vizsgált diskurzusjelölők az egyéni nyelvhasználati sajátosságok azonosításában külön-külön és együtt is (változó mértékben) alkalmazhatók.

A további kutatásokban érdemesnek tartjuk az előfordulásuk pozíciójának, funkciójának és környezetének elemzését is, valamint a diskurzusjelölő-társulások vizsgálatát. Ezek további szempontokat adhatnak hozzá az egyéni nyelvhasználati jellemzők feltárásához, aminek a kriminalisztikai gyakorlatban kiemelkedő jelentősége van.

Köszönetnyilvánítás

A kutatást támogatta a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal az NKFIH FK128814-es projekt, valamint az Európai Unió az RRF-2.3.1-21-2022-00004 azonosítójú, Mesterséges Intelligencia Nemzeti Laboratórium projekt keretében.

Irodalom

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19/6, 716–723.
- Alföldi Ágnes Dóra (2012). A profilalkotás tudományterületi elhelyezkedése és térbeli modelljei. *Magyar Tudomány*, 2012/8, 980–987. Letöltés: http://epa.oszk.hu/00600/00691/00104/pdf/EPA00691_mtud_2012_08_0980-0987.pdf
- Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67/1, 1–48. doi: 10.18637/jss.v067.i01.
- Bóna Judit (2010). Bizonytalansági megakadások idősek és fiatalok spontán beszédében. *Beszédkutató* 2010, 125–138.
- Bóna Judit (2014). Az életkor, a nem és a beszéd típus hatása a diskurzusjelölők használatára. In Havas Ferenc, Horváth Katalin, Kugler Nóra, & Vladár Zsuzsa (szerk.), *Nyelvben a világ. Tanulmányok Ladányi Mária tiszteletére* (388–397). Budapest: Tinta Könyvkiadó.
- Coulthard, M. & Johnson, A. (2007). *An Introduction to Forensic Linguistics. Language in Evidence*. London & New York: Routledge.
- Dér Csilla Ilona (2009). Mik is a diskurzusjelölők? In Keszler Borbála & Tátrai Szilárd (szerk.), *Diskurzus a grammatikában – grammatika a diskurzusban* (293–303). Budapest: Tinta Könyvkiadó.
- Dér Csilla Ilona (2021). Diskurzusjelölő a periférián? A szerintem véleményjelölő beszélt nyelvbéli pozícióiról. *Magyar Nyelvőr*, 145, 312–330. doi: 10.38143/Nyr.2021.3.312.
- Dér, Cs. I. & Markó, A. (2010). A pilot study of Hungarian discourse markers. *Language and Speech*, 53/2, 135–80.
- Dömötör Adrienn (2008). A *hogy úgy mondjam* diskurzusjelölő. *Magyar Nyelv*, 104, 427–442.
- Dömötör Adrienn (2013). Idéző szerkezetből keletkezett diskurzusjelölők – és követőik. In Forgács Tamás, Németh Miklós, & Sinkovics Balázs (szerk.), *A nyelvtörténeti kutatások újabb eredményei VII.* (31–41). Szeged: SzTE Magyar Nyelvészeti Tanszék.
- Fox, J., Weisberg, S. (2019). *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Gósy Mária (2002). A megakadásjelenségek eredete a beszédprodukción tervezési folyamatában. *Magyar Nyelvőr* 126/2. 192–204. Letöltés: <http://www.c3.hu/~nyelvor/period/1262/126206.pdf>
- Gósy Mária (2012). Multifunkcionális beszélt nyelvi adatbázis – BEA. *Általános Nyelvészeti Tanulmányok*, 24, 329–349.
- Götz Andrea (2019). Diskurzusjelölők és kötőelemek gyakorisága írott és beszélt, mediált és nem mediált diskurzusokban. In Laczkó Krisztina & Tátrai Szilárd (szerk.), *Kontextualizáció és metapragmatikai tudatosság* (291–313). Budapest: ELTE Eötvös József Collegium.
- Gyuris Beáta 2008. A diskurzuspartikulák formális vizsgálata felé. In Kiefer Ferenc (szerk.), *Strukturális magyar nyelvtan 4.* (639–682). Budapest: Akadémiai Kiadó.
- Kredens, K. (2000). *Forensic Linguistics and the Status of Linguistic Evidence in the Legal Setting*. PhD dissertation. University of Łódź.
- Kugler Nóra (2015). *Megfigyelés és következtetés a nyelvi tevékenységben*. Budapest: Tinta Könyvkiadó.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82/13, 1–26. doi: 10.18637/jss.v082.i13.
- Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.6.0. Letöltés: <https://CRAN.R-project.org/package=emmeans>

- Lüdecke, D. (2021). *sjstats: Statistical Functions for Regression Models* (Version 0.18.1). doi: 10.5281/zenodo.1284472, <https://CRAN.R-project.org/package=sjstats>.
- Markó Alexandra & Dér Csilla Ilona (2011). Diskurzuszjelölők használatának életkori sajátosságai. In Navracsecs Judit és Lengyel Zsolt (szerk.), *Lexikai folyamatok egy- és kétnyelvű közegben. Pszicholingvisztikai tanulmányok II.* (49–61). Budapest: Tinta Könyvkiadó.
- Matei, M. (2011). The influence of age and gender on the selection of discourse markers in casual conversations. *Bulletin of the Transilvania University of Braşov, Series IV: Philology and Cultural Studies*, 4/1, 213–220. Letöltés: <http://rs.unitbv.ro/BU2011/Series%20IV/BULETIN%20IV%20PDF/32%20MATEI.pdf>
- Mihajlik Péter, Grácsi Tekla Etelka, Kohári Anna, Tarján Balázs, Balog András & Mády Katalin 2022. A BEA továbbfejlesztése és alkalmazása kontrasztív gépi beszédfelismerési kísérletekre. In: Mády Katalin & Markó Alexandra (szerk.), *Általános nyelvészeti tanulmányok 34.: Fonetikai tanulmányok*. Budapest: Akadémiai Kiadó 361–380.
- Molnár Cecília Sarolta (2019). *Speciális kérdések? Az ugye partikulát tartalmazó megnyilatkozások formája és használata*. PhD-disszertáció. Budapest: ELTE.
- R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Letöltés: <https://www.R-project.org/>.
- Schirm Anita (2010). Hogyan (ne) tanítsuk a diskurzuszjelölöket? In Zimányi Árpád (szerk.), *A tudomány nyelve – A nyelv tudománya. Alkalmazott nyelvészeti kutatások a magyar nyelv évében. XIX. Magyar Alkalmazott Nyelvészeti Kongresszus. Eger, 2009. április 16–18.*(389–396). Székesfehérvár & Eger: MANYE & Eszterházy Károly Főiskola.
- Schirm Anita (2015). A diskurzuszjelölők az osztálytermi kommunikáció szövegtípusaiban. In Baditzné Pálvölgyi Kata, Szabó Éva, & Szentgyörgyi Rudolf (szerk.), *Tanóratervezés és tanóratudatosság: A magyar nyelv és irodalom, az idegen nyelvek és a művészetek műveltségi területen* (49–66). Budapest: Eötvös Loránd Tudományegyetem.
- Schirm Anita (2017). A diskurzuszjelölők és a szövegtípusok viszonyáról. *Magyar Nyelv*, 113, 330–341.
- Schirm Anita (2019a). A diskurzuszjelölő-használat életkori sajátosságai a nyelvi interjú szövegtípusában. *Beszédkutatás 2019*, 187–205. doi: 10.15775/Beszkut.2019.187-205
- Schirm Anita (2019b). A diskurzuszjelölők funkciói a szövegtípusok tükrében. In Laczkó Krisztina és Tátrai Szilárd (szerk.), *Kontextualizáció és metapragmatikai tudatosság* (265–290). Budapest: ELTE Eötvös József Collegium.
- Skarnitzl, R., Vaňková, J. (2017). Fundamental frequency statistics for male speakers of common Czech. *Acta Universitatis Carolinae: Philologica*, 2017/3, 7–17.
- Sztahó, D., Beke, A., Szaszák, Gy. & Fejes, A. (2022). Forensic Authorship Classification by Paragraph Vectors of Speech Transcriptions. Berend Gábor, Gosztolya Gábor és Vincze Veronika (szerk.), *XVII. Magyar Számítógépes Nyelvészeti Konferencia* (275–288). Szeged: Szegedi Tudományegyetem Informatikai Intézet.
- Tatár Zoltán (2013). Beszélőprofil-alkotás lehetőségei a kriminalisztikai fonetikában. *Alkalmazott Nyelvtudomány*, XIII/1–2, 121–130.
- Tatár Zoltán, Varga Zoltán és Főző Eszter (2021). Beszélőazonosítás a kriminalisztikában. In Markó Alexandra (szerk.), *Tanulmányok a beszédtudomány alkalmazásainak köréből*. 229–250. Budapest: ELTE Eötvös Kiadó.
- Vincze Veronika, Kicsi András, Főző Eszter és Vidács László (2021). A gépi elemzők kriminalisztikai szempontú felhasználásának lehetőségei. In Berend Gábor, Gosztolya Gábor és Vincze Veronika (szerk.), *XVII. Magyar Számítógépes Nyelvészeti Konferencia* (275–288). Szeged: Szegedi Tudományegyetem Informatikai Intézet.
- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.