

A borderline személyiségzavar felismerése a nyelvhasználat lexikai és grammatikai jellemzői alapján

Felletár Fanni^{1,2}, Babarczy Anna^{1,2}

¹ Nyelvtudományi Kutatóközpont, Kísérleti Szemantika és Pragmatika Kutatócsoport
{felletar.fanni, babarczy.anna}@nytud.hu

² Budapesti Műszaki és Gazdaságtudományi Egyetem, Kognitív Tudományi Tanszék
babarczy.anna@ttk.bme.hu

Kivonat A borderline személyiségzavarban (borderline personality disorder; BPD) érintett személyeket az identitás, a társas kapcsolatok valamint az érzelem- és indulatszabályozás zavarának tartós mintázata jellemzi. Az átlagpopuláció heterogenitása és a tradicionális diagnosztikai módszerek szubjektív jellege miatt a kórkép felismerése nehéz feladatot ró a klinikusokra. Számítógépes nyelvészeti módszerekkel kimutatták, hogy a BPD-vel élők – a kontroll személyekhez (healthy controls; HC) viszonyítva – szignifikánsan gyakrabban használnak egyes számú és tagadást kifejező nyelvi formákat valamint negatív valenciájú szavakat a narratíváikban. Emellett előző kutatásunkban arra az eredményre jutottunk, hogy a beszédprodukciónyelvi enkódolás fázisához köthető hibáinak aránya jelentős magyarázóerővel bír a BPD-ben érintett személyek megjósolásakor. Ezek alapján azt feltételeztük, hogy a BPD és a HC csoport elkülöníthető a nyelvi enkódolási hibák, az egyes számú és a tagadást kifejező nyelvi formák valamint a negatív valenciájú szavak alapján. A manuálisan és a NarrCat segítségével kinyert faktorokat egy logisztikus regressziós modellben használtuk fel, melynek AUC értéke 0,907 lett. A modelltől arra következtethetünk, hogy minél magasabb az említett faktorok aránya egy narratívában, annál nagyobb a valószínűsége, hogy az adott személy a BPD csoportba tartozik.

1 Bevezetés

1.1 A borderline személyiségzavar jellemzői és felismerésének nehézségei

A borderline személyiségzavarral (borderline personality disorder; BPD) élő személyeket az identitás, a társas kapcsolatok valamint az érzelem- és indulatszabályozás zavarának tartós mintázata jellemzi. A DSM-5-TR (American Psychiatric Association [APA], 2022) dimenzionális modellje alapján a diagnózis felállításához az egyénnek a következő 7 személyiségvonásból legalább 4–gyel kell rendelkeznie, melyek közül legalább az egyik impulzivitás, kockázatvállalás vagy ellenségesség: érzelmi labilitás, depresszivitás, szorongásosság, szeparációs bizonytalanság, impulzivitás, kockázatvállalás és ellenségesség. A kórkép központi komponense továbbá az identitászavar, amely ürességérzés, alacsony szorongástűrés, valamint társas interakció hatására változó énkép, célok és értékrend formájában jelenik

meg (APA, 2022). Mivel a nyelvhasználat során az egyén gondolatait, érzelmeit, intencióit fejezi ki, a gondolatokat, érzelmeiket, intenciókat formáló személyiségvonások is tükröződnek a nyelvhasználatban (Du és Sun, 2022; Hirsh és Peterson, 2009; Ireland és Mehl, 2014; Pennebaker és mtsai, 2003). Bár a BPD-nek nem centrális tünete az atipikus nyelvhasználat, a fentiek alapján feltételezhető, hogy a BPD személyiségvonásai megjelennek valamilyen módon a nyelvhasználatban.

Tekintve, hogy a BPD gyakran komorbid egyéb zavarokkal (pl. hangulatzavarok, szorongásos zavarok, függőségek, figyelemhiányos hiperaktivitás-zavar [ADHD], más személyiségzavarok), valamint a DSM – klinikai gyakorlatban továbbra is elterjedtebb – kategoriális modellje szerint a diagnózis felállítható 9-ből bármely 5 diagnosztikai kritérium teljesülése esetén (APA, 2022), a BPD-ben érintettek populációja meglehetősen heterogén (Cavelti et al, 2012; Millon és mtsai, 2004; Smits és mtsai, 2017), így a diagnosztizáció komoly klinikai kihívást jelent. További nehezítő tényező, hogy a tradicionális diagnosztikai módszerek (interjú, tesztek) a páciens belátásán, őszinteségén és emlékezeti képességén valamint a klinikus szubjektív megállapításain alapulnak, így nem teszik lehetővé a tünetek közvetlen és objektív mérését (Cwik és mtsai, 2016; Heinz és mtsai, 2022; Walsh és mtsai, 2020). Ezek a módszerek elengedhetetlenek, hiszen a páciens szubjektív megélése és a klinikus szakértelme a diagnosztizáció központi komponensei, azonban a BPD-vel élők nyelvhasználatával foglalkozó kutatások (Carter és Grenyer, 2012; Lyons és mtsai, 2018; Rosenbach és Renneberg, 2015; Wang és mtsai, 2020) eredményei rámutattak, hogy a számítógépes nyelvészeti módszerek diagnosztikai alkalmazása hasznos kiegészítő eleme lehet a jelenlegi hagyományos eljárásnak.

1.2 A BPD-ben érintett személyek nyelvhasználatának jellemzői

A BPD-vel élő személyek nyelvhasználatának vizsgálata viszonylag új kutatási iránynak számít, azonban néhány jó kiindulópontként szolgáló tanulmány már megjelent a témában. A tanulmányok egy része a diskurzusszinttel foglalkozik, ennek is két típusával: a narratív monológok koherenciaelemzésével (Adler és mtsai, 2012; Bendstrup és mtsai, 2021; Lind és mtsai, 2019) és a párbeszédok szerveződésének vizsgálatával (Wang és mtsai, 2020, 2021). A mondatok szintjén Carter és Grenyer (2012) a mondatok hosszát és komplexitását vizsgálta. A szavak szintjén a lexikai denzitást és diverzitást mérték, valamint szógyakorisági elemzéseket végeztek a Linguistic Inquiry and Word Count (LIWC; Pennebaker és mtsai, 2001) szoftver segítségével a következő kategóriákban: szófajok, beszédtema (pl. család, munka, halál), személyjelölés (E/1, E/3, T/1, T/3), tagadás, pszichés folyamatok (pl. kogníció, érzelmek), “abszolút” szavak (pl. minden, semmi) (Carter és Grenyer, 2012; Lyons és mtsai, 2018; Rosenbach és Renneberg, 2015; Spruit és mtsai, 2022; Wang és mtsai, 2020).

A BPD-ben érintett személyek szóhasználatára fókuszáló kutatások jellemzően a személyjelölést és a szavak valenciáját vizsgálták. Az E/1-es személyes névmásokat tekintve azt találták, hogy a BPD-vel élő személyek narratíváiban szignifikánsan gyakrabban fordulnak elő ezek az alakok, mint az egészséges kontroll személyek (healthy controls; HC) személyek narratíváiban (Carter és Grenyer, 2012; Lyons és mtsai, 2018; Rosenbach és Renneberg, 2015; Spruit és mtsai, 2022). Spruit és

munkatársai (2022) egy különböző mentális zavarokkal (ADHD, autizmus spektrum zavar, bipoláris zavar, BPD, evészavar, generalizált szorongásos zavar, major depresszió, obszesszív-kompulzív zavar [OCD], poszttraumás stressz szindróma, skizofrénia) élő személyeket magában foglaló klinikai csoportot és egy HC csoportot összevetve azt fedezte fel, hogy az E/1-es névmások használata a mentális zavarral élők csoportjának általános jellemzője. Az E/3-as személyes névmásokat azonban nemcsak a HC személyekhez viszonyítva (Carter és Grenyer, 2012; Lyons és mtsai, 2018), hanem a generalizált szorongásos zavarral, a major depresszióval, az OCD-vel és a skizofréniaival élő személyekhez mérten is szignifikánsan gyakrabban használják a BPD-ben érintett személyek (Lyons és mtsai, 2018). A személyes névmások gyakorisága arról ad információt, hogy milyen személy(ek) áll(nak) az adott személy figyelmi fókuszában (Lyons és mtsai, 2018). Így az E/1-es névmások gyakori használatáról befelé irányuló figyelemre, depresszióra, ruminációra következtethetünk (Edwards és Holzman, 2017; Zimmerman et al, 2016), az E/3-as névmások gyakori megjelenése pedig egy az egyén életében kitüntetett helyet elfoglaló személyre való fókuszaltságot jelez (Lyons és mtsai, 2018). Mivel a depresszivitás és a kitüntetett személy irányában érzett szeparációs szorongás a BPD személyiségvonásai közé tartoznak (APA, 2022), ezek az eredmények jól tükrözik a BPD-vel élő személyek működését.

A személyjelölésen túl a BPD-ben érintett személyek nyelvhasználatának meghatározó nyelvi markere az érzelmeket és értékelést kifejező szavak valenciája. Az ezzel kapcsolatos kutatások egybehangzóan arra az eredményre jutottak, hogy a BPD-vel élő személyek narratívái szignifikánsan több negatív valenciájú szót tartalmaznak, mint a HC személyeké (Carter és Grenyer, 2012; Lyons és mtsai, 2018; Spodenkiewicz és mtsai, 2013). Mivel a negatív valenciájú szavak megnövekedett száma a depresszió jele (Baddeley és mtsai, 2011; Newell és mtsai, 2018), és a BPD-re jellemző a depresszivitás (APA, 2022), ez az eredmény érvényesnek tekinthető. A negatív valenciájú szavakhoz hasonlóan a tagadás is egyfajta negativizmust, negatív orientációt fejez ki. Rosenbach és Renneberg (2015) major depressziós, BPD-ben érintett és HC személyeket összehasonlítva azt mutatták ki, hogy a BPD csoport mindkét csoporthoz képest gyakrabban használt tagadást kifejező nyelvi formákat. Wang és munkatársai (2020) kutatása szerint a tagadószavak a bipoláris személyekkel szemben is elkülönítik a BPD-vel élő személyeket.

Előző tanulmányunkban a BPD és a megakadásjelenségek gyakorisága közötti összefüggést vizsgáltuk (Felletár és mtsai, 2022). Számos jellemzőt figyelembe véve, a beszédprodukciónyelvi enkódolás (Levelt, 1989) fázisához köthető hibáinak (grammatikai hibák, kontamináció) aránya bírt a legnagyobb magyarázóerővel a BPD csoport függvényében. A grammatikai hibák morfológiai és szintaktikai szabályoknak ellentmondó jelenségek. A kontaminációk szó- és szerkezetvegyülések, amelyek egyszerre aktiválódó, egymással versenyző szerkezetek eredményei (Gósy, 2002).

1.3 A kutatás célja és menete

Jelen kutatásunk célja a BPD-ben érintett személyek nyelvhasználatának modellálása a HC személyek nyelvhasználatához viszonyítva, számítógéppel támogatott tartalomelemzési módszerrel. A fenti kutatások eredményeire támaszkodva azt feltételezzük, hogy a BPD és a HC csoport jól elkülöníthető a nyelvi enkódolási hibák,

az egyes számú nyelvi formák, a negatív valenciájú szavak valamint a tagadást kifejező nyelvi formák alapján. Tudomásunk szerint kutatásunk az első, amely magyar anyanyelvű mintán foglalkozik a témával.

A grammatikai hibákat, a kontaminációt és az E/3-as nyelvi formákat manuálisan, a többi jellemzőt pedig gépi módszerrel, a NarrCat (Ehmann és mtsai, 2014; László és mtsai, 2013) segítségével annotáltuk. A szógyakoriság alapján kinyert, standardizált, majd kategóriánként aggregált jellemzőket egy logisztikus regressziós statisztikai modellben használtuk fel a két csoport osztályozása céljából. Mivel kutatásunk hosszú távú célja a BPD felismerésének számítógépes nyelvészeti támogatása, a modell kiértékelésekor főként a modell fedését, F-értékét és AUC értékét vettük figyelembe.

2 Módszer

2.1 Résztevők

A BPD-ben érintett és a HC személyek esetében is kényelmi és hólabda mintavételt alkalmaztunk. Mindkét csoport meghatározó részét a közösségi médiában jelen lévő BPD-vel élő személyeknek létrehozott és egyéb (pl. kísérleteket hirdető) csoportokból toboroztuk. A kutatásban már részt vett személyeket arra kértük, hogy osszák meg a részvételi lehetőséget ismerőseikkel. Mindkét csoportnál alapkritérium volt, hogy a résztvevők magyar anyanyelvűek legyenek, illetve ép hallással és beszédszervekkel rendelkezzenek. A BPD-ben érintett személyek esetében a kutatásban való részvétel feltétele a BPD diagnózisát igazoló klinikai dokumentum bemutatása volt, amely tartalmazza az F6030 – Érzelmileg labilis személyiségzavar BNO-kódot (BNO-11, 2022). A komorbid zavarok nem számítottak kizáró oknak, mivel az átlagpopulációban is gyakran előfordulnak. A HC személyként való részvétel feltétele az volt, hogy az egyének ne rendelkezzenek az alábbi zavarok egyikével sem: nyelvi vagy beszédzavar (pl. afázia), agyi vagy idegrendszeri betegség, sérülés (pl. Parkinson-kór, agydaganat), fejlődési zavar (pl. ADHD), pszichoaktív szerhasználat (gyógyszer-, alkohol-, drogfüggőség), mentális zavar (pl. személyiségzavar, skizofrénia). A HC személyek önbevallás alapján adtak számot az egészségi állapotukról. Végül 27 BPD-vel élő és 27 HC személy vett részt a kísérletben. A két csoport illesztésénél a résztvevők nemét, életkorát és sikeresen elvégzett tanéveinek számát vettük figyelembe, amelyek adatait az 1. táblázat foglalja össze.

Jellemzők	BPD	HC
N	27	27
nemi arány; nő:férfi	23:4	23:4
életkor; $M(SD)$	27,26(7,18)	26,59(7,28)
elvégzett tanévek száma; $M(SD)$	14,85(2,38)	15,52(2,85)

1. táblázat. A résztvevők demográfiai adatai.

BPD = borderline személyiségzavar; *HC* = egészséges kontroll; *N* = elemszám; *M* = átlag; *SD* = szórás.

A táblázat alapján láthatjuk, hogy a két csoport adatai kiegyensúlyozottak. Az elemszám ($N = 27$) és nemi arány (23:4) teljesen megegyezik, az átlagéletkor és az elvégzett tanévek átlagos száma pedig közel azonos. A nők ugyan felülreprezentáltak, azonban a BPD-ben érintettek klinikai populációját is hasonló arány jellemzi (APA, 2022). A BPD és a HC csoport életkorának és elvégzett tanévei számának összevetésére – az adatok nem normális eloszlása miatt – Mann-Whitney próbát alkalmaztunk, amely azt mutatta, hogy sem az életkor ($p = 0,590$), sem az iskolázottság ($p = 0,437$) tekintetében nincs szignifikáns eltérés a két csoport között. A hasonló demográfiai adatok lehetővé teszik a valós különbségek megállapítását a statisztikai elemzés során.

2.2 Adatgyűjtés

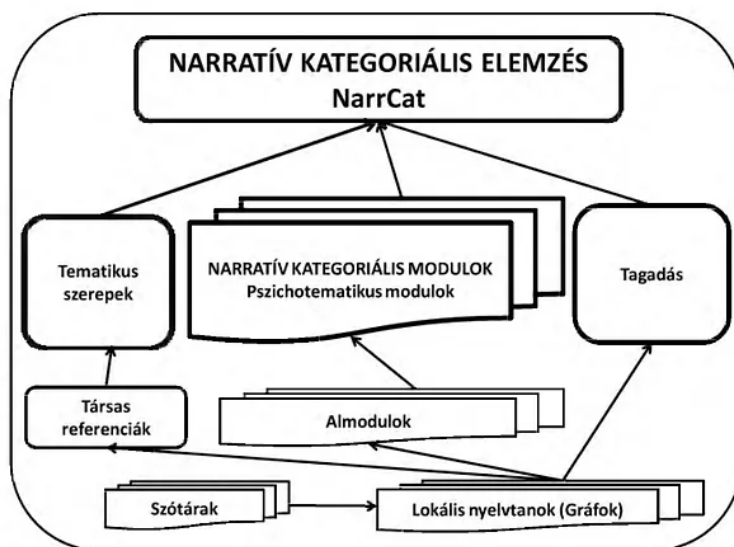
Az adatgyűjtéshez a narratív interjú módszerét választottuk, amely egy személyes életeseményekre fókuszáló, strukturálatlan interjúzási stratégia, melyben az adatrögzítő határozza meg a narratíva témáját, majd hagyja az adatközlőt minél szabadabban kibontakozni. Azért esett a választásunk erre a módszerre, mert 1) a pszichiátriai vizsgálatok során jellemzően strukturálatlan interjút alkalmaznak (Bruchmüller és mtsai, 2011; Cwik és mtsai, 2016; Tringer, 2010), 2) a BPD-ben érintett személyek a HC személyekhez hasonlóan teljesítenek strukturált körülmények között (Edell, 1987; Kernberg, 1975; Millon és mtsai, 2004), 3) így megfelelő közlésmennyiséget kapunk a jellemzők méréséhez. Az adatközlőket arra kértük, hogy meséljék el a tegnapi napjukat: *Kérlek, meséld el részletesen a tegnapi napodat a felébredésedtől kezdve! Hogyan telt a napod? Mit csináltál? Mi történt veled? Hogy éreztél magad?* Nem szabtuk meg minimum és maximum időkeretet, ezzel is elősegítve az adatközlő minél spontánabb megnyilvánulását. A résztvevőket a kísérlet menetéről tájékoztattuk, a beszédfelvételhez és annak tudományos célú felhasználásához írásos hozzájárulásukat adták.

2.3 Adatok előfeldolgozása

Az adatgyűjtés lezárultával a hangfelvételeket megvágtuk és felhangosítottuk az Audacity program segítségével (Audacity® 3.1.3; <https://audacityteam.org/>). A hangfelvételeket ezután lefűttattuk az Alrite automatikus beszédfelismerő programon (Alrite©; <https://alrite.io/ai/hu/>), amelynek eredményeként egy összesen 20.139 tokenből álló szöveges korpuszt kaptunk. Az átiratok tartalmi és nyelvi helyességét manuálisan ellenőriztük, szükség esetén a hibákat kijavítottuk.

Mivel fontosnak tartottuk, hogy kutatásunk módszertanában is konzisztens legyen a bemutatott tanulmányokkal, a nyelvhasználat lexikai jellemzőinek kinyeréséhez a hivatkozott irodalomban alkalmazott LIWC mintájára fejlesztett NarrCat narratív kategoriális tartalomelemző szoftver (Ehmann és mtsai, 2014; László és mtsai, 2013) társas referencia, tagadás (Hargitai, 2004), érzelem (Fülöp és László, 2006) és értékelés (Bigazzi és mtsai, 2006) moduljait használtuk. A NarrCat a Nooj korpusznyelvészeti fejlesztő környezetben (Silberstein, 2005) fut, és a Magyar Nemzeti Szövegtár (Váradi, 2002), a Szeged Korpusz (Csendes és mtsai, 2005) valamint különböző történelmi és

többek között BPD-vel élőkkel gyűjtött pszichológiai tematikájú korpuszokon lett betanítva. A rendszer alapját a felsorolt korpuszokból épített annotált szótárak és a lokális nyelvtanok képezik, amelyek együttesen szolgáltatnak bemenetet az ún. pszichotematikus modulokhoz (1. ábra). A szoftver a narratívákat tagmondatokra bontja, majd az egyes tagmondatokat adatredukció révén vektorokká transzformálja, melyek értékei a pszichológiai kategóriákba tartozó nyelvi formák gyakoriságát reprezentálják (Ehmann és mtsai, 2014).



1. ábra: A NarrCat felépítése (Ehmann és mtsai, 2014).

A társas referencia modulból az E/1-es formákat, az érzelm modulból a negatív érzelmet jelölő szavakat, az értékelés modulból pedig a negatív értékelést jelölő szavakat mérő jellemzőket vettük figyelembe. Érzelmet jelölő szavaknak nevezzük azon főneveket, mellékneveket, igéket és határozókat, amelyek bármely morfológiai formájukban érzelmi (pl. *dühös*), hangulati (pl. *jókedvű*) vagy izgalmi (pl. *fáradt*) állapotot, állapotváltozást (pl. *beleszeret*), illetve ezek viselkedéses megnyilvánulását (pl. *sír*) fejezik ki, melynek az elbeszélő experiense (pl. *izgul*) vagy páciense (pl. *megbánt*) (Fülöp és László, 2006). Értékelést kifejező szavaknak a cselekvést, történést, állapotot (pl. *fekszik*) értelmező főneveket (pl. *munkakerülő*), mellékneveket (pl. *lusta*), igéket (pl. *pihen*) és határozókat tekintjük (Bigazzi és mtsai, 2006). Mivel a társas referencia modul az E/1-es és a T/1-es nyelvi formákat azonosítja, az E/3-as alakokat manuálisan annotáltuk. Az E/3-as formák kézi annotációja során a társas referencia modul nyelvi markerein (személyes névmások, személyragok) túl a közneveket és a tulajdonneveket is jelöltük. Abban az esetben is jelöltük az E/1-es és E/3-as alakokat, ha implicit argumentumként szerepeltek. A lexikai jellemzők annotálását a 2. táblázat szemlélteti.

Szövegszegmens	E/1	E/3	NEZ	NET	TAG
<i>Szóval nagyon-nagyon kemény volt,</i>	0	0	0	1	0
<i>és minden egyes szombatom után ilyen nagyon idegesen megyek haza.</i>	2	0	1	0	0
<i>Jól érzem magam a barátnőimmal,</i>	3	0	0	0	0
<i>tök jót nevetgélünk meg dumálunk,</i>	0	0	0	0	0
<i>de belül ideges vagyok,</i>	1	0	1	0	0
<i>mert nem tanultam semmi újat,</i>	1	0	0	0	2
<i>mert leszólták a filozófiát,</i>	0	0	0	1	0
<i>mert a témaköröm a szexualitás,</i>	1	0	0	0	0
<i>azt is leszólják meg kiröhögik,</i>	0	0	0	2	0
<i>túl nyílnak tartanak,</i>	1	0	0	1	0
<i>az is vagyok,</i>	1	0	0	0	0
<i>az nem zavar,</i>	1	0	0	0	1
<i>de hogy kicsit úgy kezelnek így a tanárok is,</i>	1	0	0	0	0
<i>hogyan én leszek az,</i>	2	0	0	0	0
<i>akivel nehéz dolguk lesz.</i>	0	0	0	1	0

2. táblázat. Lexikai jellemzők annotálása.

NEZ = negatív érzelem; NET= negatív értékelés; TAG = tagadás.

A nyelvi enkódolás fázisához köthető hibák elemzéséhez a megakadásjelenségek két típusát, a grammatikai hibákat és a kontaminációkat annotáltuk. A grammatikai hibák helyét [GRM], a kontaminációkat pedig [KNT] címkével jelöltük manuálisan a szövegben. Ezt követően a jegyek gépi összeszámolásával kiszámítottuk mindkét jelenség gyakoriságát. A grammatikai jellemzők annotálását az alábbi példa szemlélteti:

...gondoltam, hogy így elé megyek, akkor jövünk [GRM] együtt ha- jöjjünk együtt haza, meg ilyesmi, vele voltunk [KNT] (vele voltam × együtt voltunk), aztán én még elmentem...

Miután kinyertük az összes jellemző gyakoriságát, a kapott értékeket arányosítottuk a szószám alapján. A hatásméret növelése és a dimenziócsökkentés érdekében a jellemzőket (grammatikai hiba, kontamináció, E/1, E/3, negatív érzelem, negatív értékelés) faktorokba aggregáltuk: nyelvi enkódolási hibák, egyes szám, negatív valenciájú szavak. A tagadás jellemzőt eredeti formájában hagytuk meg. A jellemzők rendszerét a 3. táblázat foglalja össze.

Jellemzők	Nyelvi markerek	Példák
Nyelvi enkódolási hibák		
grammatikai hiba	morfológiai és szintaktikai hibák	“... <i>azt néztem, ahogy Jamie Oliver scrampie hamburgereket készít.</i> ”
kontamináció	szerkezeti vegyülés	“... <i>ha most nem fogod be ezt a faszágot, akkor meg foglak ütni!</i> ”
Egyes szám		
E/1	Névmások	<i>én, engem, magát, neki</i>
E/3	Személyragok	<i>jött, órái</i>
	Köznevek	<i>a srác, a barátnőm</i>
	Tulajdonnevek	<i>Rodrigo</i>
Negatív valenciájú szavak		
negatív érzelem ^a	Főnevek	<i>szégyenérzet^a, függőség^b</i>
	Melléknevek	<i>szomorú^a, bántalmazott^b</i>
negatív értékelés ^b	Igék	<i>stresszelek^a, utálok^b</i>
	Határozók	<i>idegesen^a, kínosan^b</i>
Tagadás^c	Tagadószók	<i>nem, se</i>
	Igék	<i>nincs, sincsen</i>
	Határozószók	<i>sosem, sehol</i>
	Névmások	<i>senki</i>
	Névtók	<i>nélkül</i>
	Szuffixumok	<i>névtelen, hibátlan</i>
	Prefixumok	<i>unintelligens, antipatikus</i>

3. táblázat. Kinyert jellemzők, a BPD csoport narratíváiból vett példákkal.
^a = negatív érzelmeket kifejező szavak; ^b = negatív értékelést kifejező szavak;
^c = Hargitai (2004) példái.

3 Eredmények

Az adatok előfeldolgozását követően leíróstatistikai elemzést majd hipotézisvizsgálatot végeztünk az R programozási nyelvre épülő JASP statisztikai szoftver segítségével (JASP Team (2022), JASP (Version 0.16.3); <https://jasp-stats.org/>).

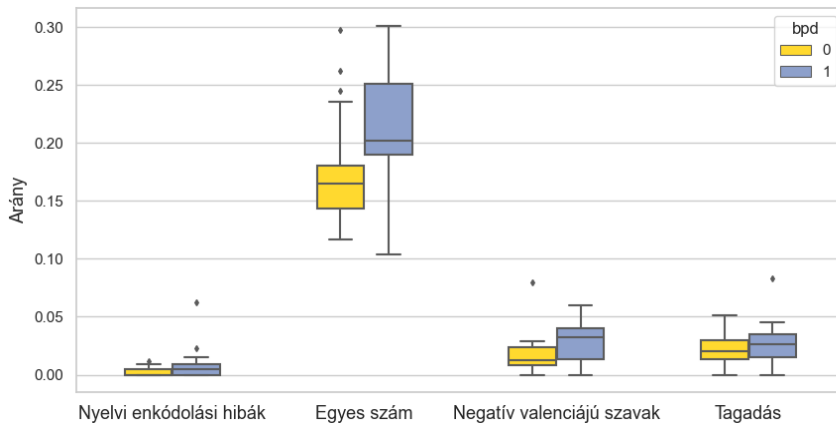
3.1 Leíróstatistikai elemzés

Első lépésként a jellemzők és a faktorok átlagát, szórását és a BPD csoporttal való összefüggését vizsgáltuk meg, melynek eredményeit a 4. táblázat foglalja össze.

Jellemzők	BPD M(SD)	HC M(SD)	r^2
Nyelvi enkódolási hibák	0,008(0,012)	0,002(0,003)	0,137
grammatikai hiba	0,006(0,012)	0,002(0,003)	0,084
kontamináció	0,002(0,004)	0,000(0,000)	0,017
Egyes szám	0,213(0,046)	0,172(0,043)	0,336
E/1	0,170(0,052)	0,134(0,041)	0,260
E/3	0,043(0,025)	0,038(0,029)	0,014
Negatív valenciájú szavak	0,028(0,017)	0,017(0,015)	0,185
negatív érzelem	0,012(0,011)	0,006(0,006)	0,096
negatív értékelés	0,016(0,013)	0,011(0,011)	0,068
Tagadás	0,027(0,016)	0,021(0,013)	0,048

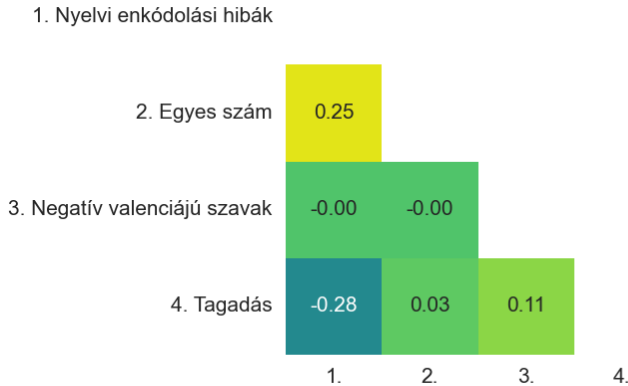
4. táblázat. Leíróstatisztikai eredmények. *BPD* = borderline személyiségzavar; *HC* = egészséges kontroll; *M* = átlag; *SD* = szórás; r^2 = determinációs együttható.

A BPD csoport narratíváiban az összes jellemző átlagosan nagyobb arányban fordul elő, mint a HC csoportban. A nyelvi enkódolási hibák esetében ez az arány ($M_{BPD} = 0,008$) a HC csoportnak ($M_{HC} = 0,002$) a négyszerese. A determinációs együttható alapján, a két csoport közötti különbség az egyes szám tekintetében a legnagyobb ($r^2 = 0,336$), amelyet az alábbi ábra (2. ábra) jól szemléltet a többi faktorhoz viszonyítva.



2. ábra: A faktorok arányának helyzeti középértékei a BPD (kék) és HC (sárga) csoportban.

A hipotézisvizsgálásra használt logisztikus regressziós modell előkészítéseként, a kollinearitás minimalizálása érdekében megvizsgáltuk, a faktorok páronként mennyire korrelálnak egymással. A faktorok kiszűréséhez a korrelációs erősség kritikus szintjét $|0,40|$ -nél húztuk meg, ez alatt ugyanis hagyományosan legfeljebb csak közepes erősségű összefüggésről beszélhetünk (pl. Dancey és Reidy, 2004). Az alábbi korrelációs mátrix alapján láthatjuk, hogy nincs erős korreláció a faktorok között (3. ábra).



3. ábra: A faktorok korrelációs mátrixa.

3.2 Hipotézisvizsgálás

A hipotézisvizsgálást egy logisztikus regressziós modellben, “stepwise” módszerrel végeztük el, melyben az eredményváltozó a csoport volt, a prediktorok pedig a nyelvi enkódolási hibák, az egyes szám, a negatív valenciájú szavak és a tagadás. A első modell csak az interceptet tartalmazta, amelyhez a második modellben ($R^2 = 0,233$; $\chi^2(52) = 10,392$; $p = 0,001$) az egyes szám, a harmadikban ($R^2 = 0,371$; $\chi^2(51) = 7,216$; $p = 0,007$) a nyelvi enkódolási hibák, a negyedikben ($R^2 = 0,527$; $\chi^2(50) = 9,581$; $p = 0,002$) a negatív valenciájú szavak, az ötödikben ($R^2 = 0,624$; $\chi^2(49) = 6,909$; $p = 0,009$) pedig a tagadás adódott hozzá (ld. 5. Táblázat).

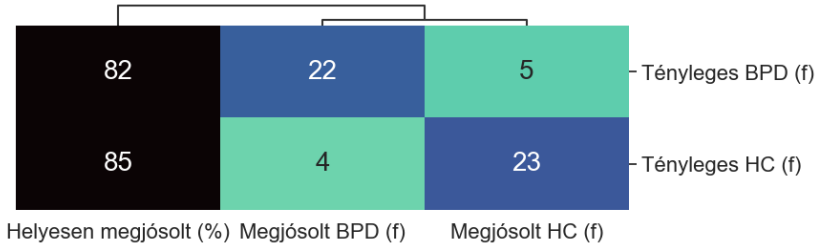
Modell	BIC	R^2	χ^2	df	p
1. Csak intercept	78,849			53	
2. Egyes szám	72,446	0,233	10,392	52	0,001
3. Egyes szám Nyelvi enkódolási hibák	69,219	0,371	7,216	51	0,007
4. Egyes szám Nyelvi enkódolási hibák Negatív valenciájú szavak	63,627	0,527	9,581	50	0,002
5. Egyes szám Nyelvi enkódolási hibák Negatív valenciájú szavak Tagadás	60,707	0,624	6,909	49	0,009

5. táblázat. A modellépítés folyamata.

BIC = Bayesi információs kritérium; R^2 = pszeudo determinációs együttható; χ^2 = a χ^2 -teszt eredménye; df = szabadságfok; p = szignifikanciaérték.

Bár az összes modell erősen szignifikáns, az utolsó modell bejósoló ereje a legpontosabb ($BIC = 60,707$), és a varianciát is ez magyarázza a legmagasabb arányban

($R^2 = 0,624$). A végső modell pontossága 0,833, precizitása 0,846, fedése 0,815, F-értéke 0,830, AUC értéke pedig 0,907 lett. Az alábbi konfúziós mátrixban (4. ábra) láthatjuk, hogy a modell mindkét csoport tagjait közel egyforma, 80% körüli arányban osztályozza helyesen, a HC személyeket azonban valamivel pontosabban kategorizálja.



4. ábra: Az osztályozás konfúziós mátrixa a kimenetek gyakoriságával (f) és a helyes kategorizációk arányával (%).

A Khí-négyzet teszt alapján nem mutatható ki szignifikáns összefüggés a kategorizáció helyessége és a nem között ($\varphi_c = 0,093$; $\chi^2(1) = 0,470$; $p = 0,493$). A nők közel 85%-át és a férfiak 75%-át osztályozta helyesen a modell. A modell megfelel a prediktorok egymástól való függetlenségét mérő multikollinearitás kritériumainak, és mind a négy prediktorának jelentős mértékű és szignifikáns a hozzájárulása a modell bejósoló erejéhez (ld. 6. táblázat).

Változók	b	SE	z	p	Tolerancia	VIF
Intercept	0,485	-2,827	-3,312	<,001		
Egyes szám	1,700	0,599	2,839	0,005	0,630	1,588
Nyelvi enkódolási hibák	3,325	0,349	2,869	0,004	0,625	1,599
Negatív valenciájú szavak	1,261	0,449	2,808	0,005	0,692	1,445
Tagadás	1,294	0,560	2,311	0,021	0,696	1,436

6. táblázat. Az utolsó modell változói.

b = standardizált regressziós együttható; SE = a standardizált regressziós együttható standard hibája; z = a z-teszt eredménye; p = szignifikanciaérték; VIF = Variance Inflation Factor.

4 Összegzés

Jelen tanulmányunkban azt feltételeztük, hogy a BPD-ben érintett személyek jól elkülöníthetőek az egészséges személyek csoportjától a nyelvi enkódolási hibák, az egyes számú nyelvi formák, a negatív valenciájú szavak és a tagadást kifejező nyelvi formák alapján. Egy logisztikus regressziós modell segítségével hatékonyan be tudtuk sorolni a résztvevőket a BPD és a HC csoportba. A szakirodalomban tárgyalt megállapításokkal összhangban, eredményként azt kaptuk, hogy minél magasabb a nyelvi enkódolási hibák, az egyes számú nyelvi formák, a negatív valenciájú szavak és

a tagadást kifejező szavak aránya egy narratívában, annál nagyobb a valószínűsége, hogy az adott személy a BPD csoportba tartozik.

Kutatásunk hosszú távú célja a BPD felismerésének számítógépes nyelvészeti támogatása. A gyakorlatban is hasznosítható modell építéséhez azonban a BPD populáció nyelvhasználatának átfogóbb megismerésére lesz szükség, amit többféle annotációs és elemzési módszer (pl. HuSpaCy, Orosz és mtsai, 2022; emóció- és szentimentszótárak, Szabó, 2014) keresztezésével, változatos szövegtípusok elemzésével, és más pszichiátriai zavarokban érintett populációk bevonásával remélhetünk megvalósítani a középtávú jövőben.

Köszönetnyilvánítás

Felletár Fanni kutatását a Magyar Tudományos Akadémia fiatal kutatói ösztöndíjprogramja támogatta. Köszönjük Kneszl Beáta Carmennek, BPD aktivistának és írónak, hogy részt vett a kutatásban, és emellett népszerűsítette a részvételi lehetőséget a BPD közösségben, Gosztolya Gábornak, az ELKH-SZTE Mesterséges Intelligencia Kutatócsoport munkatársának, hogy közreműködött a jellemzők kinyerésében, Pólya Tibornak, a Természettudományi Kutatóközpont Narratív és Történeti Pszichológiai Kutatócsoportja vezetőjének, hogy biztosította számunkra a NarrCat használatát, valamint a kísérlet összes résztvevőjének, hogy narratíváikkal hozzájárultak az adatgyűjtés sikerességéhez.

Hivatkozások

- Adler, J. M., Chin, E. D., Kolisetty, A. P., Oltmanns, T. F.: The distinguishing characteristics of narrative identity in adults with features of Borderline Personality Disorder. An empirical investigation. In: *Journal of Personality Disorders* 26(4). pp. 498–512 (2012) <https://doi.org/10.1521%2Fpedi.2012.26.4.498>
- American Psychiatric Association: *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). American Psychiatric Association, Washington, DC (2022)
- Baddeley, J. L., Daniel, G. R., Pennebaker, J. W.: How Henry Hellyer's use of language foretold his suicide. *Crisis* 32(5). pp. 288–92 (2011) <https://doi.org/10.1027/0227-5910/a000092>
- Bendstrup, G., Simonsen, E., Kongerslev, M. T., Jørgensen, M. S., Petersen, L. S., Thomsen, M. S., Vestergaard, M.: Narrative coherence of autobiographical memories in women with borderline personality disorder and associations with childhood adversity. *Borderline Personality Disorder and Emotion Dysregulation* 8(18). (2021) <https://doi.org/10.1186/s40479-021-00159-5>
- Bigazzi S., Csertő I., Nencini A.: A személy- és csoportközi értékelés pszicholingvisztikája. In: *IV. Magyar Számítógépes Nyelvészeti Konferencia*. pp. 267–277. Szegedi Tudományegyetem, Szeged (2015)
- Carter, P. E., Grenyer, B. F. S.: Expressive language disturbance in borderline personality disorder in response to emotional autobiographical stimuli. *Journal of*

- Personality Disorders 26(3). pp. 305–321 (2012)
<https://doi.org/10.1521/pedi.2012.26.3.305>
- Cavelti, M., Lerch, S., Ghinea, D., Fischer-Waldschmidt, G., Resch, F., Koenig, J., Kaess, M.: Heterogeneity of borderline personality disorder symptoms in help-seeking adolescents. *Borderline Personal Disord Emot Dysregul.* 8(1):9. (2021 Feb 26) <https://doi.org/10.1186/s40479-021-00147-9>
- Csendes, D., Csirik, J., Gyimóthy, T., Kocsor, A.: The Szeged Treebank. In: *Proceedings of the 8th International Conference on Text, Speech and Dialogue.* pp. 123-131. Karlovy Vary, Czech Republic (2005 September 12-16)
- Cwik, J. C., Papen, F., Lemke, J.-E., Margraf, J.: An Investigation of Diagnostic Accuracy and Confidence Associated with Diagnostic Checklists as Well as Gender Biases in Relation to Mental Disorders. *Front. Psychol.* 7(1813). (2016) <https://doi.org/10.3389/fpsyg.2016.01813>
- Dancey, C., Reidy, J.: *Statistics without Maths for Psychology. Using SPSS for Windows.* Prentice Hall, London (2004)
- Du, X., Sun, Y.: Linguistic features and psychological states. A machine-learning based approach. *Front. Psychol.* 13(955850). (2022) <https://doi.org/10.3389/fpsyg.2022.955850>
- Edwards, T., Holtzman, N. S.: A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality* 68. pp. 63–68 (2017) <https://doi.org/10.1016/j.jrp.2017.02.005>
- Ehmann, B., Csertő, I., Ferenczhalmy, R., Fülöp, É., Hargitai, R., Kóvágó, P., Pólya, T., Szalai, K., Vincze, O., László, J.: Narratív kategóriális tartalomelemzés. A NARRCAT. In: X. Magyar Számítógépes Nyelvészeti Konferencia. pp. 136–147. Szegedi Tudományegyetem, Szeged (2014)
- Felletár, F., Gosztolya, G., Hoffmann, I., Babarczy, A.: A megakadásjelenségek és a temporális paraméterek szerepe a borderline személyiségzavar felismerésében. *Beszédtudomány – Speech Science* 3(1). pp. 207–243 (2022) DOI: 10.15775/Besztud.2022.207-243
- Fülöp, É., László, J.: Az elbeszélések érzelmi aspektusának vizsgálata tartalomlemző program segítségével. In: IV. Magyar Számítógépes Nyelvészeti Konferencia. pp. 296–304. Szegedi Tudományegyetem, Szeged (2006)
- Gósy, M.: A megakadásjelenségek eredete a spontán beszéd tervezési folyamatában. *Magyar Nyelvőr* 126(2). pp. 192–204 (2002)
- Hargitai, R.: A LAS-Vertikum tagadás és self-referencia modulja. In: II. Magyar Számítógépes Nyelvészeti Konferencia. pp. 261–264. Szegedi Tudományegyetem, Szeged (2004)
- Heinz, M. V., Thomas, N. X., Nguyen, N. D., Griffin, T. Z., Jacobson, N.C.: *Technological Advances in Clinical Assessment.* In: Asmundson, G. J. G. (szerk.) *Comprehensive Clinical Psychology*, 2nd edition, Vol. 4. pp. 301–320. Elsevier (2022) <https://dx.doi.org/10.1016/B978-0-12-818697-8.00171-0>
- Hirsh, J. B., Peterson, J. B.: Personality and language use in self-narratives. *Journal of Research in Personality* 43. pp. 524–527 (2009) <https://doi.org/10.1016/j.jrp.2009.01.006>
- Ireland, M. E., Mehl, M. R.: Natural language use as a marker of personality. In: Holtgraves, T. M. (szerk.), *The Oxford handbook of language and social psychology.* pp. 201–218. Oxford University Press (2014)

- László, J., Csertő, I., Fülöp, É., Ferenczhalmy, R., Hargitai, R., Lendvai, P., Péley, B., Pólya, T., Szalai, K., Vincze, O., Ehmann, B.: Narrative Language as an Expression of Individual and Group Identity. The Narrative Categorical Content Analysis. SAGE Open. pp. 1–12 (2013 April-June) <https://doi.org/10.1177/2158244013492084>
- Levelt, W. J. M.: Speaking. From intention to articulation. MIT Press, Cambridge, MA (1989)
- Lind, M., Vanwoerden, S., Penner, F., Sharp, C.: Inpatient Adolescents With Borderline Personality Disorder Features. Identity Diffusion and Narrative Incoherence. Personality Disorders 10(4). pp. 389–393 (2019) <https://doi.org/10.1037/per0000338>
- Lyons, M., Aksayli, N. D., Brewer, G.: Mental distress and language use. Linguistic analysis of discussion forum posts. Computers in Human Behavior 87. pp. 207–211. (2018) <https://doi.org/10.1016/j.chb.2018.05.035>
- Millon, T., Grossman, S., Millon, C., Meagher, S., Ramnath, R.: Personality Disorders in Modern Life (2nd ed.). Wiley, Hoboken, NJ (2004)
- Newell, E. E., McCoy, S. K., Newman, M. L., Wellman, J. D., Gardner, S. K.: You Sound So Down. Capturing Depressed Affect Through Depressed Language. Journal of Language and Social Psychology 37(4). pp. 451–474 (2018)
- Orosz, Gy., Szántó, Zs., Berkecz, P., Szabó, G., Farkas, R.: HuSpaCy: an industrial-strength Hungarian natural language processing toolkit. In: XVIII. Magyar Számítógépes Nyelvészeti Konferencia. pp. 59–73. Szegedi Tudományegyetem, Szeged (2022) <https://doi.org/10.48550/arXiv.2201.01956>
- Pennebaker, J. W., Mehl, M. R., Niederhoffer, K. G.: Psychological aspects of natural language. use. Our words, our selves. Annu Rev Psychol. 54. pp. 547–77 (2003) <https://doi.org/10.1146/annurev.psych.54.101601.145041>
- Rosenbach, C., Renneberg, B.: Remembering rejection. Specificity and linguistic styles of autobiographical memories in borderline personality disorder and depression. Journal of Behavior Therapy and Experimental Psychiatry 46. pp. 85–92 (2015) <https://doi.org/10.1016/j.jbtep.2014.09.002>
- Silberztein, M.: NooJ. A Linguistic Annotation System for Corpus Processing. In: Proceedings of HLT/EMNLP 2005 Interactive Demonstrations. pp. 10–11. Association for Computational Linguistics, Vancouver, British Columbia, Canada (2005)
- Smits, M. L., Feenstra, D.J., Bales, D. L., de Vos, J., Lucas, Z., Verheul, R., Luyten, P.: Subtypes of borderline personality disorder patients: a cluster-analytic approach. Borderline Personal Disord Emot Dysregul 4(16). (2017) <https://doi.org/10.1186/s40479-017-0066-4>
- Spodenkiewicz, M., Speranza, M., Taïeb, O., Pham-Scottez, A., Corcos, M., Révah-Levy, A.: Living from day to day. Qualitative study on borderline personality disorder in adolescence. J Can Acad Child Adolesc Psychiatry 22(4). pp. 282–289 (2013 Nov) PMID: 24223047
- Spruit, M., Verkleij, S., de Schepper, K., Scheepers, F.: Exploring Language Markers of Mental Health in Psychiatric Stories. Applied Sciences 12(2179). (2022) <http://dx.doi.org/10.3390/app12042179>
- Szabó, M. K.: Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai. Nyelv, kultúra, társadalom. Budapest, Hungary (2014)

- Várad, T.: The Hungarian National Corpus. Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002). pp. 385–389. Las Palmas de Gran Canaria (2002)
- Walsh, C. G., Chaudhry, B., Dua, P., Goodman, K. W., Kaplan, B., Kavuluru, R., Solomonides, A., Subbian, V.: Stigma, biomarkers, and algorithmic bias. Recommendations for precision behavioral health with artificial intelligence. *JAMIA Open* 3(1). pp. 9–15 (2020) <https://doi.org/10.1093/jamiaopen/ooz054>
- Wang, B., Wu, Y., Taylor, N., Lyons, T., Liakata, M., Nevado-Holgado, A. J., Saunders, K. E. A.: Learning to Detect Bipolar Disorder and Borderline Personality Disorder with Language and Speech in Non-Clinical Interviews. *INTERSPEECH* 2020. pp. 437–441. Shanghai, China (2020) <https://doi.org/10.48550/arXiv.2008.03408>
- Wang, B., Wu, Y., Vaci, N., Liakata, M., Lyons, T., Saunders, K. E. A.: Modelling Paralinguistic Properties in Conversational Speech to Detect Bipolar Disorder and Borderline Personality Disorder. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 7243–7247 (2021) <https://doi.org/10.48550/arXiv.2102.09607>
- Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., Wolf, M.: First-person Pronoun Use in Spoken Language as a Predictor of Future Depressive Symptoms. Preliminary Evidence from a Clinical Sample of Depressed Patients. *Clin Psychol Psychother* 24(2). pp. 384–391 (2017 Mar) <https://doi.org/10.1002/cpp.2006>