

BALÁZS GÉZA – BEKE OTTÓ – SZŰTS ZOLTÁN

KONTAKTUSNYELV ÉS CHATGPT

A vajdasági magyar nyelvhasználat mesterséges intelligencia általi értelmezésének jelenlegi lehetőségei (Netnyelvészeti esettanulmány)

Kivonat

A magyar nyelvészek már évtizedek óta kutatják a vajdasági magyar nyelvhasználat jelenségeit. A mesterséges intelligencia széles körű elterjedésével azonban tudáshiány lépett fel, amelynek azonosítására és lehetséges megválaszolóra a jelen (eset)tanulmány vállalkozik, amikor a vajdasági magyar nyelvhasználat ChatGPT általi értelmezésének kurrens potenciálját méri fel. A vizsgálat korpuszát három, kontaktusnyelven megfogalmazott prompt alkotja. A dolgozat változatlan formában közli a szerb nyelvből származó lexikai elemeket és egyéb jellegzetességeket tartalmazó felhasználói promptokat, majd azok mesterséges intelligencia általi köznyelvi interpretációját. A köznyelvi változatok értelmezése a jelentésbeli ekvivalencia felmérésén alapul.

Kulcsszavak: vajdasági magyar nyelvhasználat, kontaktusnyelv, prompt, ChatGPT

1. Bevezetés

A mediatizált és digitális világban a nyelv is az: mediatizált és digitális. Technikailag felaprózott, szukcesszív és szegmentális. A közvetítő közeg jellemvonásaihoz, architektúrájához alkalmazkodik. Ezt mélystruktúrája teszi lehetővé, ugyanis egymást követő diszkrét elemekből épül fel. Ez képezi szerkezeti és működési alapját (Saussure 1997). „[A] nyelv szegmentális (diszkrét) elemei”, vagyis a beszédhangok, „pontosabban az ezekből – mindekeleltől a gazdaságosságra törekvő írásgyakorlat által – létrehozott »funkcionális fikciók«, a fonémák, illetve az ezekből felépülő szavak” (Benczik 2001: 24; lásd továbbá Benczik 2006: 12–6; vö. Derrida 1991; Orbán 1994) (számítás)technikai értelemben véve immár valóban digitálissá válnak. Megszabadulnak jelentést hordozó, megkülönböztető „testes” jellegüktől (Benczik 2001: 24), s csupán a szegmentáltság, a tagoltság, illetve a tágabb értelemben vett artikuláltság marad meghatározó. Testetlenné válnak. A digitalizáció és azzal együtt a virtualizáció a nyelvi kód logikájából, absztrakciós potenciáljából táplálkozik (Chomsky 2018: 43–4).

A mediatizált és digitális, illetve a digitálisan rögzített és közvetített nyelv illékonyága még írott formájában is az élőbeszéd-szerűséget idézi fel. Rögzítettsége inkább számítástechnikai, grafematikus, digitális, mintsem ortográfiai és grammatikai. Neveztük másodlagos írásbeliségnek, spontán írásbeliségnek; s küszöbön áll a harmadlagos, akár gépinek is nevezhető írásbeliség és szóbeliség (a megközelítések összefoglalása és továbbvezetése: Balázs G. 2024a, Arató–Balázs 2024b).

A verbális kommunikáció a gomba módra szaporodó online platformokon bontakozik ki és alakul át. Ez a cseppfolyós írásban (Assmann 2009: 156) való beszéd, a „szkriptoralitás” (Orbán 1998: 87), illetve a kreolizáció, az élőbeszédnek az írásba történő befurakodása, egy újfajta nyelvi minőség, avagy létmód: „Ma még írunk, de az új eszközökön sokszor úgy írunk, mintha beszélénk (levélbeszéd, internetes csevegés), a távirati tömörségű SMS pedig sok előszóbeli megnyilvánulást is jellemez. Új nyelvi hatások, változások indultak el, amelyek új nyelvi, gondolkodási minőséget hoznak létre” (Balázs G. 2003: 18; lásd továbbá Balázs G. 2023: 85, 102, 129; Ong 2010; vö. Ferraris 2008). Az életvilág mediatizációja (Berger 2023) a nyelvet is magába szippantja. A változásokra a nyelvészetnek is szükségképpen reflektálnia kell. Ez a netnyelvészet feladata és kihívása (Veszelszki 2017). Az 1990-es évek végén formálódni kezdő magyar netnyelvészet kezdeti kérdései még nyelviek (nyelvhasználatiak) voltak, de a kutatások hamarosan kulturális, antropológiai jelleget öltöttek; és mára megállapíthatóvá vált, hogy nem pusztán kommunikációs-nyelvi változások történnek, hanem egy új evolúciós „megugrás”, „megszaladás” jelei bontakoznak ki (Balázs G. 2023: 297–305).

2. Kontaktusnyelv és ChatGPT

Az élőbeszédet jellemző közvetlenség és esetlegesség különböző módokon nyilvánul meg az online térben. Ezek közé sorolható a kontaktus-, esetenként pedig kifejezetten a keveréknyelveken megfogalmazott írásbeli és szóbeli, vagyis hangüzenetek megjelenése és burjánzása az online platformokon (Balázs L. 2023)

A kontaktusnyelv vagy kontaktuskód két vagy több nyelv érintkezéséből fakadó kölcsönhatásaként létrejött nyelvváltozat (Ladányi–Hrenek 2019: 78–9); esetünkben egy kisebbségi nyelvben a többségi nyelv lexikai, szerkezeti (vonzatszerkezeti) változóinak megjelenése. A keveréknyelv (vagy kevert nyelv) két- vagy többnyelvű beszélők csoportjai közötti intenzív és elszigetelt kontaktus eredményeként kialakuló, nem közvetítő nyelvként szolgáló kontaktusnyelv, amelyet ugyancsak a kapcsolatba kerülő nyelvek ötvöződő lexikai és grammatikai jegyei jellemeznek (Ladányi–Hrenek 2019: 71). Fon-

tos megjegyezni, hogy mindkét, egyébként összekapcsolódó és akár nehezen szétválasztható esetben az átvett nyelvi jelenségek rendszerint kívül maradnak a köznyelvi normarendszeren. Produktumaik és azok írásbeli reprezentációi az internet, különösképpen pedig a web 2.0 és a közösségi média (Szűts 2018: 305–421) elterjedése előtt kevésbé voltak megfigyelhetők, vizsgálhatók; az elemzések eddig kizárólag az élőnyelvi és a sajtónyelvi anyagra terjedtek ki. Az információs és nyelvi kapuőrök útját állták megjelenésüknek, széles körű használatuknak. Egyfajta nyelvi cenzor szerepét is betöltötték. Ami nem felelt meg a sztenderd nyelvváltozat kívánalmainak, az csak esetlegesen bukkant fel a sajtónyelvben, az élőnyelvben pedig ugyancsak nehéz volt tetten érni.

A weben és főleg a social médiában a kapuőrök funkciója háttérbe szorul, minimalizálódik, esetenként meg is szűnik (Szekfü 2007: 125–6). Ennek következtében mindenki úgy beszél, fogalmaz és ír, ahogy szeretne, illetve ahogy tud. Potenciálisan végtelen a megnyilatkozási-publikációs tér és lehetőség. Mindenképpen csökken – ha nem is szűnik meg – a normalizációs nyomás, illetve a (mikro)közösségekben, kisvilágokban jelentkezik (Kovács 2010, 2011; Barabási 2011; Balázs G. 2014, 2017, 2023). Nyelvi, sőt esetenként stílusbeli buborékok, visszhangkamrák alakulnak ki (vö. Szűts 2018: 206; Harari 2018: 194; Pariser 2012), és azok megformálódását, valamint megszilárdulását többek között a mesterséges intelligencia algoritmikus működésmódja segíti elő (Kökuti et al. 2023). A kontaktus- és keveréknyelvi üzenetek megtalálják a maguk helyét és létjogosultságát az online térben. Ez az auditív és írásos tartalmakra egyaránt vonatkozik.

2022 novembere óta széles rétegek számára ingyenesen elérhető a ChatGPT generatív mesterséges intelligencia (MI), a nagy nyelvi modell használata jelentős hatást gyakorol az ember–ember és az ember–gép között zajló kommunikációs folyamatokra és mintázatokra (Harari 2024: 179–268). Ehhez leginkább az alkalmazás konverzációs képességei járulnak hozzá, hiszen hatékony a nyelvi üzenetek, szövegek értelmezésében, valamint generálásában. Ez a felhasználóban a beszélgetés benyomását kelti. A jelzett percepciós hatás kialakulását az is nagyban elősegíti, hogy az egy-egy promptot¹ követő gépi válasz megalkotása – eltekintve az igazán bonyolult és időigényes feladatok megoldásától – csupán másodperceket vesz igénybe, annak megjelenítése azonban írásban karakterről karakterre, betűről betűre történik. Az alkalmazás ugyanis minden esetben az élő párbeszéd élményét igyekszik nyújtani. A felhasználónak ennek megfelelően az a benyomása támad, hogy a válasz valós időben generálódik, és ez javítja a kommunikációs élményt (vö. Csíkszentmihályi 2016; Szűts 2018: 339–42).

¹ Prompt: Szöveges kérdés, utasítás (itt) szöveg, más esetekben kép, fordítás létrehozására (generálására), személyre szabott ajánlásokra stb.

Szerteágazó tudományos diskurzus foglalkozik az MI-k és azon belül a nagy nyelvi modellek jelenségkörének különböző vetületeivel. Annak a kérdésnek a felvetése, majd pedig szisztematikus tárgyalása azonban, hogy a ChatGPT képes-e, és amennyiben igen, milyen szinten és hatékonysággal képes értelmezni az egyes nyelvváltozatok, különös tekintetben a kontaktus- és keveréknyelvi jellemvonások által meghatározott közleményeket, mind a mai napig még a nemzetközi szakirodalomban sem történt meg. A jelen netnyelvészeti esettanulmány arra tesz kísérletet, hogy felmérje a vajdasági magyar nyelvhasználat mesterséges intelligencia általi értelmezésének kurrens potenciálját. A szóban forgó nyelvváltozat vonatkozásában a kérdés megválaszolása kulcsfontosságú, hiszen az adott határon túli közösség nyelvhasználatát a szerbvel mint államnyelvvél történő kapcsolat és közelség nem hagyja érintetlenül.² Ennek eklatáns és egyben közismert példái a szókölcsönzések, de egyéb grammatikai, például szintaktikai és pragmatikai hatások is előfordulnak. (A szerbiai, vajdasági magyar nyelvváltozathoz lásd: Vukov Raffai 2019.)

Valószínűsíthető, hogy a chatbot³ számára a kontaktusnyelven megfogalmazott üzenetek feloldozása és értelmezése az azokban előforduló „nyelvidegen” lexikai elemek, továbbá egyéb nyelvtani jellemzők folytán kihívást jelent, hiszen a tanító adatok e tekintetben hiányosak, illetve korlátosak (lehetnek). Természetesen nem maguknak a szerb nyelvből származó komponenseknek interpretációja, hanem azok magyar nyelvű kontextusa, vagyis beágyazottsága jelentheti a mesterséges intelligencia számára a nehézséget. A ChatGPT-nek a vizsgálatban betöltött, kiemelkedő pozícióját széles körben elterjedt használata indokolja. (Tanulmányunk folytatása egy korábbi hasonló jellegű kutatásnak: Balázs G.–Beke–Szűts 2024.)

A kutatás a kontaktusnyelvi elemeket és egyéb nyelvi jellegzetességeket leíró módon kezeli, vagyis értéksemleges, szociolingvisztikai megközelítésmódot alkalmaz.

3. Kontaktusnyelvű promptok és gépi értelmezésük

Az alábbiakban változtatások nélkül közöljük az empirikus kutatásba bevont, vajdasági magyar kontaktusnyelven megfogalmazott promptokat. Megalkotásukért köszönet illeti Dudás Benjámin és Molnár Emesét, az Újvidéki

² Természetesen a magyar nyelv esetében az összes határon túli, kisebbségi nyelv-, illetve diaszpóráközösség esetében hasonló jelenségek és tendenciák tapasztalhatók. A jelen esettanulmány azonban kimondottan a vajdasági magyar nyelvváltozatra összpontosít.

³ Chatbot: olyan szoftveralkalmazás (robot), amely szöveges vagy szövegről beszédre átalakított online kommunikációt folytat élő személyekkel (ügyfelekkel, ügyintézőkkel) előre betanított szabályok és betáplált információk alapján.

Egyetem szabadkai székhelyű Magyar Tannyelvű Tanítóképző Karának negyedéves kommunikáció szakos hallgatóit, továbbá Slezák Lázárt, Zombor Város Közigazgatási Hivatalának szerb–magyar fordítóját. Az üzenetek összeállítása során nem vehették igénybe a ChatGPT vagy egyéb nagy nyelvi modell szolgáltatásait. Kizárólag saját nyelvi-szövegalkotási kompetenciájukra hagyatkozhattak. Valóban humán ágensek módjára tevékenykedtek, vagyis nem szervezhették ki összetett kognitív funkcióikat technikai apparátusokba. Annak érdekében pedig, hogy az ne befolyásolja stílusukat, nem kaptak felvilágosítást a kutatás kérdésfelvetésére vonatkozóan. Egyetlen kritériumnak kellett eleget tenniük, nevezetesen annak, hogy közleményüket a szerb nyelvből származó lexikai elemek, sajátosságok felhasználásával, vagyis vajdasági magyar kontaktusnyelven fogalmazzák meg. Tartalmi megkötés sem korlátozta őket. Az ily módon létrehozott promptok hiteles képet festenek az adott kontaktusnyelv jellegzetességeiről; természetesen nem reprezentálják az adott nyelvváltozatban használatban lévő, kölcsönzött lexikai és egyéb elemek széles skáláját, a karakterisztikát azonban mindenképpen érzékeltetik. A hitelesség érdekében az üzenetekben előforduló, szerb nyelvből származó szavak lejegyzését illető fonetikus írásmódot, illetve mellékjelhasználatot nem korrigáltuk, továbbá a nyelvjárási elemek (tudatos) tükrözését is érintetlenül hagytuk. Hasonlóképpen jártunk el a helyesírási hibák és elgépelések esetében is.

A felhasználói üzeneteket követően szintén változtatások nélkül közöljük azok ChatGPT általi értelmezését, majd pedig levonjuk a tudományos konzekvenciákat. Fontos kiemelni, hogy különösképpen szöveggenerálás során a chatbot működése részben probabilisztikus. A folyamatban tehát mindig szerepet játszik a véletlen, amelynek alapját a programban alkalmazott véletlenszám-generátor algoritmusai biztosítják. Irányított véletlenszerűségről van szó, amelyet a modell paraméterei szabályoznak.

1. számú prompt

Értelmezd az alábbi szöveget: Képzeld voltam egy bazenos buliban múlt szombaton. A piát magunknak kellett vinni, és nem tudtam eldönteni, hogy limenkás sört vagy pelinkovácot vigyek, ezért az utóbbinál döntöttem, hogy jól teljen a zsurka. És nem hiszed el, de az eladónő elkérte tőlem a licsnámat, pedig szerintem olyan fiatalnak nem is nézek ki. Le voltam dőbbenve, mondom ajde nézze meg, hogy hány éves vagyok. Aztán rájött, hogy nem kiskorúval van dolga. Végül megvettük a piát, aztán a buliban nagyon jól éreztük magunkat.

MI általi értelmezés

Képzeld, múlt szombaton voltam egy medencés buliban. A piát magunknak kellett vinni, és nem tudtam eldönteni, hogy limonádés sört vagy pelinkovácot vigyek, végül az utóbbit választottam, hogy jól induljon a buli. És el se hiszed, de az eladónő elkérte a személyimét, pedig szerintem nem nézek ki olyan fiatalnak. Teljesen ledöbbsentem, mondtam is neki, hogy nyugodtan nézze meg, hány éves vagyok. Aztán rájött, hogy nem vagyok kiskorú. Végül megvettük a piát, és a buliban nagyon jól éreztük magunkat.

2. számú prompt

Értelmezd az alábbi szöveget: Szia, képzeld ma elmentem patikát venni, és nem volt nálam a kártyám ezért nem tudtam fizetni. Mivel készpénz sem igazán volt nálam ezért elmentem a bankba. Ott nyilván kellett a licsnám meg valami potvrdát alá írni, hogy tudják pénzt kapni. Szóval már ez a cirkusz után nem csak patikát vettem, hanem hozzá egy trénerkát is meg egy majicát. Aztán mentem haza közben voltam boltban, vettem egy limenkás szokot meg kajának a péknél burekot. Kajáltam majd mentem edzeni, de ott meg lejárt a cslanirám. Azt meg nem tudtam már ki fizetni mert apámnak vettem szülinapjára egy slagot. Szóval tiszta káosz ez a mai nap, még majdnem fel buktam egy bankinába, de estére zsurka remélem jössz te is. Veled mizu?

MI általi értelmezés

Szia, képzeld el, ma elmentem gyógyszert venni, de nem volt nálam a bankkártyám, úgyhogy nem tudtam fizetni. Készpénzem se nagyon volt, ezért elmentem a bankba. Ott persze kellett a személyim, meg alá kellett írnom valami papírt, hogy tudjak pénzt felvenni. Na, miután ezen túl voltam, nem csak gyógyszert vettem, hanem egy melegítőnadrágot meg egy pólót is. Aztán hazafelé még beugrottam a boltba, vettem egy citromos üdítőt meg kajának a péknél bureket. Otthon ettem, utána mentem edzeni, de ott meg kiderült, hogy lejárt a bérletem. Azt már nem tudtam befizetni, mert apámnak vettem szülinapjára egy locsolócsövet. Szóval teljes káosz volt a mai nap, még majdnem elestem egy padkában is. De estére buli lesz, remélem te is jössz! Veled mi újság?

3. számú prompt

Értelmezd az alábbi szöveget: Lejárt a licsnám, és be kellett mennem az opstinára, hogy meghosszabbítsák, meg ha már ott voltam, úgy gondoltam

zsalizok a salteron, mer kaptam azt a papírt az odvodnjavanjeről, meg azt írja valami adósságom is van. Bemegyek a pizsárnicába, ott a goszpogya aszongya nekem, menjek a 62-es szobába. Kérdezi, kaptam e termint, mer ha nem, fognom kell a sort, mer sokan várnak a licsnára. Kérdezem tőle, hol kell termint foglalni, azt mondja az eupraván. Akkor elmentem az adósalterra, hogy miért kell az odvodnjavanjet befizetnem, erre kiderült, hogy tartozok is vastagon, mert azt írja nekik, hogy csak a glavnicát fizettem be multkor, de a kamatot nem, amire ők aztán még tettek kamatot.

MI általi értelmezés

Lejárt a személyim, és be kellett mennem az önkormányzathoz, hogy meghosszabbítsák. Ha már ott voltam, gondoltam, utánanézek a pénzügyeknek is, mert kaptam egy papírt a vízelvezetési díjról, és azt is írták, hogy van valamilyen tartozásom. Bementem az ügyfélszolgálatra, ott a hölgy azt mondta, menjek a 62-es szobába. Megkérdezte, hogy van-e időpontom, mert ha nincs, akkor ki kell várnom a sort, mivel sokan várnak a személyi igazolványra. Megkérdeztem tőle, hol lehet időpontot kérni, azt mondta, az adminisztráción. Ezután elmentem a pénzügyi ügyintézőhöz, hogy megtudjam, miért kell befizetnem a vízelvezetési díjat. Kiderült, hogy tényleg van tartozásom, mert szerintük legutóbb csak a tőkerészt fizettem be, a kamatot nem, és arra még további kamatot is rászámoltak.

4. Az MI által generált köznyelvi változatok értékelése

A köznyelvi üzenetek értékelése az eredeti, vagyis kontaktusnyelven megfogalmazott promptok komplex szemantikai horizontjának figyelembevételével történik. Amennyiben a chatbotnak sikerült a hibrid forrásnyelvi kölcsönszót vagy egyéb elemet a kontextuális vonatkozásokkal egyetemben lehetőség szerint azonos vagy legalább hasonló jelentésű célnyelvvél helyettesíteni, hitelesnek és precíznek minősül a transzformáció. Más esetben kifogásolható a „köznyelvesítés” eredménye. A döntő tényező tehát a jelentésbeli ekvivalencia vizsgálata, felmérése. Ennek megfelelően a jelen értékelés nem tárgyalja a ChatGPT által végrehajtott egyéb szintaktikai, stilisztikai módosításokat, azok ugyanis a sztenderd nyelvváltozat produktumainak finomhangolásának minősülnek. Az értékelés hasonlóképpen figyelmen kívül hagyja a helyesírási korrekciókat.

Az 1. számú promptban összesen 5, szerb nyelvből származó kölcsönszó fordul elő. Ezek a felhasználói üzenetek írásmódját használva és egyben megőrizve az eredeti grammatikai alakokat is a következők: *bazenos*, *limenkás*,

žurka, ličnámát, ajde. Az idézett 5, szerb nyelvből kölcsönzött komponensből a mesterséges intelligencia 4-et helyettesített megfelelő célnyelvi formával, vagyis fordított le hitelesen. A *bazenos medencés*, a *žurka buli*, a *ličnámátból személyim* lett. Az idézett célnyelvi formák hitelesek, pontosak, továbbá a kontextusnak megfelelő toldalékokkal ellátva jelennek meg a köznyelvi transzformáció eredményeképpen. A szerb nyelvben használatos a köznyelvi *hajde* és annak a promptban megjelenő informális, népies *ajde* változata. Mindkettő magyar nyelvű megfelelője, pontos fordítása a *gyerünk* lenne, már ami magát a szót illeti, ez azonban némiképpen mesterkélt hatást keltene a fiatalos hangvételű, stílusú és egyben tematikájú felhasználói üzenet MI általi értelmezésében. Ennek megfelelően a köznyelviesített mondatban szereplő *mondtam is neki* nem csupán elfogadható fordítás, hanem kimondottan megfelel a hibrid forrásnyelvi szöveg hangulatának. Az 1. számú promptban egyetlen kölcsönzött értelmezett tévesen a chatbot, ez pedig jelesül a *limenkás* forma. A felhasználói üzenet *limenkás sört* említ, aminek magyar nyelvű megfelelő és egyben pontos fordítása a *dobozos sör* lenne, ehelyett azonban az MI általi értelmezés a *limonádés sör* szókapcsolatot tartalmazza. A félrefordítás magyarázata a *limenka és a citrom* jelenésű *lemon* szó közötti hasonlóságban keresendő.

A 2. számú prompt a szerb nyelvből kölcsönzött elemek tekintetében a korábbinál tarkább képet mutat, ugyanis több mint 10 átvételt tartalmaz, és ezek között egy hibrid, magyar toldalékkal ellátott, szerb nyelvből származó szókapcsolat is előfordul. Az MI általi értelmezés színvonala pedig kimondottan hullámzó. A felhasználói üzenet két alkalommal is *patikát* említ, ami az adott összefüggésben (*elementem patikát venni*, majd *nem csak patikát vettem*) egyaránt tornacipő jelentésben szerepel, a chatbot interpretációjában viszont az adott szöveghelyeken a *patika* szó helyett rendre a *gyógyszer* fordul elő (*elementem gyógyszert venni*, majd pedig *nem csak gyógyszert vettem*). A félreértésre a gyógyszertár jelentésben használatos patika szó nyújt magyarázatot. Kategorikusan kijelenthető, hogy ez téves fordítás, a vajdasági magyar nyelvhasználók ugyanis sohasem használják a patika szót gyógyszer jelentésben. Alkalmazzák tornacipőre és gyógyszertárra vonatkoztatva, gyógyszerre viszont sohasem. Érdekes, hogy az MI miért nem talált rá a patika = edzőcipő, tornacipő, sportcipő (Vajdaság) jelentésre, hiszen számos szógyűjtemény, nyelvművelő írás foglalkozik vele.⁴ Valószínűsíthető, hogy a chatbot a gyógyszertár jelentésű patika szó megléte miatt ezekben az esetekben nem is kutatott megfelelő szó után a szerb nyelvben, a nagy nyelvi modellek működési elve éppen ez: a legnagyobb valószínűségű megoldást

⁴ Pl. Termini szótár. Letöltés: <https://termini.nytud.hu/htonline/htlista.php?action=search> (2025. 04. 25.)

kínálják, és nem keresnek tovább. A promptban szereplő *kárticám* forma a köznyelvi változatban *bankkártyaként* szerepel, ami hiteles és egyben pontos fordítás. A *potvrda* (*potvrdat*) esetében az interpretáció (*alá kellett írnom valami papírt*) csupán részben tekinthető pontosnak, ugyanis a *potvrda* igazolást jelent, tehát nem valamilyen papírt, vagyis nem bármilyen hivatalos dokumentumot. A felhasználói üzenetben előforduló *trénerka* és *majica* szó a köznyelvi verzióban helyesen *melegítőnadrág* és *póló*.

A 2. számú promptban is megjelenik a *limenka* szó, méghozzá az *üdítő* jelentésű *szok* kölcsönszó jelzőjeként (*vettem egy limenkás szokot*), és ez esetben is tévesen *citromosként* szerepel a fordításban (*vettem egy citromos üdítőt*). Ez arra enged következtetni, hogy szisztematikusan előforduló hiba jelentkezik a transzformációkban. A promptban megjelenő *slag* szó esete meglehetősen érdekes. A szerb nyelvű *slag* szó ugyanis tejszínhabot jelent, a felhasználói üzenetben azonban mellékjel, illetve hosszú ékezet nélkül szerepel az *s* betű, a magyarban pedig a *slag* szó locsolócső jelentésben használatos. Ennek megfelelően a köznyelvi változat (*apámnak vettem szülinapjára egy locsolócsövet*) elfogadható. Az pedig, hogy az üzenet megfogalmazója inkább tejszínhabot, mintsem locsolócsövet vásárolna édesapjának születésnapjára, legfeljebb hipotézisként, nem pedig tényállításként fogalmazható meg. A prompt tartalmazza továbbá az útpadka jelentésű, egyébként az olasz nyelvből származó, a vajdasági magyar nyelvhasználatba a szerbből kölcsönzött *bankina* szót, ami a köznyelvi változatban helyesen *padkaként* fordul elő. A 2. számú felhasználói üzenetben is megjelenik a *zsurka* szó, azonban immár fonetikus írásmódban, és annak MI általi értelmezése (*buli*) ez esetben is pontos és hiteles. Valószínűsíthető, hogy a szóban forgó prompt esetében az értelmezési teljesítmény egyenetlensége mögött nem a kölcsönszavak nagyobb száma, inkább azok karakterisztikája húzódik meg. Olyan lexikai elemekről van ugyanis szó, amelyeknek analógiás hálózati elrendeződése nagyobb valószínűséggel okoz félrefordítást.

A 3. számú promptban is megjelenik, méghozzá két alkalommal is a személyi igazolvány jelentésű *lična* szó fonetikus írásmódot követve (*Lejárt a licsnám, sokan várnak a licsnára*). A köznyelvi változatban ez alkalommal is a helyes fordítás szerepel (*Lejárt a személyim, sokan várnak a személyi igazolványra*). A felhasználói üzenetben feltűnő, szintén fonetikusan lejegyzett szerb nyelvű *opština* szó (*be kellett mennem az opstinára*) MI általi értelmezése önkormányzat (*be kellett mennem az önkormányzathoz*), ami szintén helytálló. A *zsalizok a salteron* hibrid, magyar toldalékokkal és egyben határozott névelővel ellátott szókölcsonzéseket tartalmazó megfogalmazásmódnak azonban hiányzik a helyes és hiteles köznyelvi változata. A kontaktusnyelvi részlet fordítása ugyanis a következő lenne: *panaszt teszek a tolóablaknál*

vagy az ügyintézőnél, ehelyett azonban az szerepel az értelmezésben, hogy utánanézek a pénzügyeknek, ami semleges jelentésű, ugyanis nem tartalmazza azt a vonatkozást, amelynek megfelelően van ok panaszt tenni. Ez tehát ismételten az MI félrefordításának minősül. A felhasználói üzenet kétszer említi az *odvodnjavanjet* (kaptam azt a papírt az odvodnjavanjeről, miért kell az odvodnjavanjet befizetnem), ami fordításban mindkét alkalommal hitelesen jelenik meg (kaptam egy papírt a vízelvezetési díjról, miért kell befizetnem a vízelvezetési [csatorna]díjat). A promptban feltűnő, alanyt nem tartalmazó, meglehetősen igénytelen *azt írja* megfogalmazásmódok a szerb nyelv hatására jelennek meg gyakran a vajdasági magyar kontaktusnyelvben, a chatbot a vonatkozó szöveghelyeket nyelvhelyességi szempontból javította (*azt is írták*). A felhasználói üzenet vízelvezetési adósságot említ, ami a köznyelvi változatban már korrigálva, *tartozásként* tűnik fel. A prompt említi továbbá *pisarnicát* (Bemegyek a piszárnicába), *goszpogyát* (ott a goszpogya aszongya nekem) és két alkalommal időpont jelentésű *termint* (kaptam e termint, és hol kell termint foglalni). Ezek a szókölcsonzések többé-kevésbé hiteles értelmezések formájában jelennek meg a köznyelvi változatban (Bementem az ügyfélszolgálatra, ott a hölgy azt mondta, van-e időpontom, hol lehet időpontot kérni), azt azonban érdemes kiemelni, hogy a *pisarnica* egészen pontosan *iktatót* jelent, nem pedig általánosságban ügyfélszolgálatot. A ChatGPT ez esetben tehát nem volt kellőképpen precíz. További hiba, illetve pontatlanság tapasztalható a promptban említett *uprava* tolmácsolásában, az ugyanis a köznyelvi változatban pusztán *adminisztrációként* jelenik meg, holott az elektronikus közigazgatási portál (erre utal a szó elején szereplő *e* betű). A felhasználói üzenetben *adósalter* szerepel (*elementem az adósalterra*), ami adóhivatal jelentésben használatos, azt azonban *pénzügyi ügyintézőként* fordította az MI (*elementem a pénzügyi ügyintézőhöz*). Elfogadható lenne az adózási vagy pedig adóhivatali, esetleg adóhatósági ügyintéző, azonban a pénzügyi ügyintéző túlon túl általános megfogalmazásmód, ennek megfelelően pontatlan. A promptban feltűnő *glavnica* szó (*csak a glavnica-t fizettem be multkor*) valóban *tőkerész*, miként az a vonatkozó szövegrészlet köznyelvi változatában megjelenik (*csak a tőkerészt fizettem be*).

A fentiekből kitűnik, hogy a három, vajdasági magyar kontaktusnyelven megfogalmazott prompt esetében a ChatGPT számára leginkább azoknak a szerb nyelvből származó komponenseknek a tolmácsolása jelentett kihívást, amelyek analógiás hálózatba rendeződnek, vagy pedig közigazgatási, államigazgatási jellegűek, vagyis amelyek országspecifikusak. Ez utóbb említett típusba tartozó lexikai elemek autentikus fordítása komoly körütekintést és mérlegelést igényel, továbbá a helyi viszonyok ismeretét előfeltételezi.

5. Összegzés és kitekintés

A jelen netnyelvészeti esettanulmány korpuszát három, vajdasági magyar kontaktusnyelven megfogalmazott prompt alkotja. A felhasználói üzenetek korlátolt számuk és terjedelmük ellenére az adott nyelvváltozatban használatban lévő szókölcsonzések és egyéb elemek főbb jellemvonásait tükrözik. A kutatási eredmények nem reprezentatív, inkább irányadó jellegűek a ChatGPT-nek a vajdasági magyar nyelvhasználat kurrens értelmezési potenciálját illetően. A vizsgálati tényezők és körülmények azonban megfelelnek a verifikálhatóság és megismételhetőség tudományos követelményeinek.

Az elkövetkező években a különféle generatív mesterségesintelligencia-alkalmazások, a nagy nyelvi modellek óhatatlanul részét fogják képezni a munkavégzésnek, a szórakozásnak és a tanulásnak (Mollick 2024). A jelen kutatási eredmények nem csupán ahhoz járulnak hozzá, hogy legalább megközelítőleg felmérjük a vajdasági magyar nyelvhasználat ChatGPT általi értelmezésének jelenlegi lehetőségeit, hanem egyúttal a promptolási technikák fejlesztésének, finomhangolásának alapjául szolgálhatnak (vö. Bognár 2023). Az esettanulmányban alkalmazott módszertani keretrendszer pedig a kutatás folytatását, egyéb nagy nyelvi modellek, valamint kontaktus- vagy akár keveréknyelvek középpontba állítását és egyúttal komparatív műveletek elvégzését is lehetővé teszi.

Szakirodalom

- Arató, Balázs – Balázs, Géza 2024b. The communicative-linguistic modes of artificial intelligence with a focus on justice. *Magyar Nyelvőr* 601–17.
<https://doi.org/10.38143/Nyr.2024.5.601>
- Assmann, Aleida 2009. Szövegek, nyomok, hulladékok: a kulturális emlékezet változó médiumai. In: Kisantal Tamás (szerk.): *Narratívák 8. Elbeszélés, kultúra, történelem*. Kijárat Kiadó. Budapest. 146–59.
- Balázs Géza 2003. Az új média retorikája. *Vigilia* 1: 12–8.
- Balázs Géza 2014. Hálózat kutatás, közösség, nyelv. *Századvég* 72: 23–42.
- Balázs Géza 2017. Hálózat kutatás és nyelvtudomány. *Magyar Nyelvőr* 20–32.
- Balázs Géza 2023. *Az internet népe. Internet – társadalom – kultúra – nyelv. A kulturális és a tervezett evolúció határán*. Ludovika Egyetemi Kiadó. Budapest.
- Balázs Géza 2024a. Harmadlagos beszéd és írás. A mesterséges értelem kommunikációs-nyelvi létmódja. *Szabad Piac* 2024/9–10: 38–47. https://uni-milton.hu/wp-content/uploads/2025/03/szabadpiac_kiadvany_2024_2szam_0307_javitott.pdf (Letöltés: 2025. 04. 25.)
- Balázs, Géza – Beke, Ottó – Szűts, Zoltán 2024. ChatGPT language recognition limitations Prompts in Hungarian in a network of Latin and Cyrillic letters

- (Netlinguistics case study. The “Chinese room” experiment). *Magyar Nyelvőr* 666–81. <https://doi.org/10.38143/Nyr.2024.5.666>
- Balázs László 2023. Az iskolai kommunikáció egy rejtett dimenziója – szalutogén pedagógiai kommunikáció. *Magyar Nyelvőr* 692–712. <https://doi.org/10.38143/Nyr.2023.5.692>
- Barabási Albert-László 2011. *Behálózva*. Helikon Kiadó. Budapest.
- Benczik Vilmos 2001. *Nyelv, írás, irodalom kommunikációelméleti megközelítésben*. Trezor Kiadó. Budapest.
- Benczik Vilmos 2006. *Jel, hang, írás. Adalékok a nyelv medialitásának kérdéséhez*. Trezor Kiadó. Budapest.
- Berger Viktor 2023. *Az életvilág mediatizálódása*. L’Harmattan Kiadó. Budapest. <https://doi.org/10.56037/978-963-646-019-8>
- Bognár Amália 2023. A generatív mesterséges intelligencia promptolása, egy új képesség fejlesztése az oktatásban. *Pannon Digitális Pedagógia* 2: 17–26. <https://doi.org/10.56665/PADIPE.2023.2.2>
- Chomsky, Noam 2018. *Miféle teremtmények vagyunk?* Kossuth Kiadó. Budapest.
- Csikszentmihályi Mihály 2016. *Flow – Az áramlat. A tökéletes élmény pszichológiája*. Akadémiai Kiadó. Budapest.
- Derrida, Jacques 1991. *Grammatológia: Első rész. Életünk – Magyar Műhely*. Budapest–Párizs.
- Ferraris, Maurizio 2008. *Hol vagy? A mobiltelefon ontológiája*. Európa Kiadó. Budapest.
- Harari, Yuval Noah 2018. *21 lecke a 21. századra*. Animus Kiadó. Budapest.
- Harari, Yuval Noah 2024. *Nexus. Az információs hálózatok rövid története a kőkorszaktól az MI-ig*. Animus Kiadó. Budapest.
- Hayles, N. Katherine 2004. A nyomtatvány sík, a kód mély: a médiaspecifikus kutatások jelentősége. *Filológiai Közöny* 3–4: 151–70.
- Kovács László 2010. Hálózatelmélet és nyelvészet. In: Balaskó Mária – Balázs Géza – Kovács László (szerk.): *Hálózat kutatás. Hálózatok a társadalomban és a nyelvben*. Tinta Kiadó. Budapest. 9–17.
- Kovács László 2011. Hálózat kutatás és szociolingvisztika. *Magyar Nyelvőr* 90–6.
- Kőkuti, Tamás – Balázs, László – András, István – Rajcsányi-Molnár, Mónika 2023. Collaborating with Artificial Intelligence – AI in Business Communication Education In: Óbudai Egyetem (szerk.) *IEEE 6th International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE 2023): Proceedings*. Danvers (MA). 287–94. <https://doi.org/10.1109/CANDO-EPE60507.2023.10417987>
- Ladányi Mária – Hrenek Éva (szerk.) 2019. *Alkalmazott nyelvészeti kasszótár*. ELTE Eötvös Kiadó. Budapest.
- Mollick, Ethan 2024. *Társintelligencia. Élet és munka a mesterséges intelligencia új korszakában*. HVG Könyvek. Budapest.
- Ong, Walter J. 2010. *Szóbeliség és írásbeliség. A szó technológizálása*. Alkalmazott Kommunikációtudományi Intézet – Gondolat Kiadó. Budapest.

- Orbán, Jolán 1994. *Derrida írás-fordulata*. Jelenkor Kiadó. Pécs.
- Orbán Jolán 1998. Nyelvjátékok, írásjátékok: Wittgenstein és Derrida. In: Kálmán C. György – Orbán Jolán (szerk.): *Irodalom, nyelv, kultúra: exkurzus*. Jelenkor Kiadó. Pécs. 65–98.
- Pariser, Eli 2012. *The Filter Bubble*. Penguin Books. London.
<https://doi.org/10.3139/9783446431164>
- Saussure, Ferdinand de 1997. *Bevezetés az általános nyelvészetbe*. Corvina Kiadó. Budapest.
- Szekfü András 2007. *Kommunikáció, nyilvánosság, esélyegyenlőség Magyarországon. A távirótól a Web 2.0-ig*. Gondolat Kiadó. Budapest.
- Szűts Zoltán 2018. *Online. Az internetes kommunikáció és média története, elmélete és jelenségei*. Wolters Kluwer Hungary. Budapest.
- Termini szótár (Termini kutatóhálózat). <https://termini.nytud.hu/htonline/htlista.php?action=search> (Letöltés: 2025. 04. 25.)
- Veszelszki Ágnes 2017. *Netnyelvészet. Bevezetés az internet nyelvhasználatába*. L'Harmattan Kiadó. Budapest.
- Vukov Raffai Éva 2019. *Nyelvi lokáldinamika. A magyar nyelv vajdasági kontextusai*. Verbi Nyelvi Kutatóműhely, Szabadka. <https://verbi.org/wp-content/uploads/2021/04/Nyelvi-lokaldinamika.pdf> (Letöltés: 2025. 04. 23.)

Balázs Géza

egyetemi tanár

Színház- és Filmművészeti Egyetem, Doktori Iskola,

Németh Antal Drámaelméleti Intézet, Budapest

E-mail: balazs.geza@szfe.hu

<https://orcid.org/0000-0003-3440-2959>

Beke Ottó

egyetemi docens

Újvidéki Egyetem Magyar Tannyelvű

Tanítóképző Kar, Szabadka

E-mail: bekeotto1@gmail.com

<https://orcid.org/0000-0003-3519-4901>

Szűts Zoltán

egyetemi tanár

Eszterházy Károly Egyetem Pedagógiai Kar

E-mail: szuts.zoltan@uni-eszterhazy.hu

<https://orcid.org/0000-0001-8539-670X>

Abstract

BALÁZS, GÉZA – BEKE, OTTÓ – SZŰTS, ZOLTÁN

CONTACT LANGUAGE AND CHATGPT

Current possibilities for interpreting Hungarian language use in Vojvodina using artificial intelligence (A case study in net linguistics)

Hungarian linguists have been researching the phenomena of Hungarian language use in Vojvodina for decades. However, with the widespread use of artificial intelligence, a knowledge gap has emerged, which this case study attempts to identify and address by assessing the current potential of ChatGPT for interpreting Hungarian language use in Vojvodina. The corpus of the study consists of three prompts formulated in the contact language. The paper presents the user prompts containing lexical elements and other characteristics originating from the Serbian language in their original form, followed by their colloquial interpretation by artificial intelligence. The interpretation of the colloquial versions is based on the assessment of semantic equivalence.

Keywords: Hungarian language use in Vojvodina, contact language, prompt, ChatGPT