

SUMMARY

Veres-Guśpiel, Agnieszka

The effect of contextual factors on the referential interpretation of virtually used first person plural forms

The paper focuses on the construal of WE in a specific virtual usage, reflecting the physical and social worlds of the intersubjective context (cf. Verschueren 1999). From the social cognitive point of view adopted here (Croft 2009), interpersonal relations are formed and re-negotiated dynamically. The utterances do not express *a priori* existing interpersonal relations, but rather construe them as discourse unfolds. In this context, WE has remarkable semantic potential and its use is strongly context-dependent (Björten 2010, Marmaridou 2000), often expressing the social attitude of the speaker. The empirical research presented here on the referential processes of WE in a specific virtual usage reveals its capacity to express social attitude deixis and its dependence on basic contextual components. The paper is based on experimental data gathered between 2012 and 2016 from Polish and Hungarian native speakers (86 responders in total), who were asked to mark on three different schemas the reference of WE in the following utterance: *It's cold in here. Why don't we close the window?* The schemas showed different spatial and interpersonal information (position of the participants, their age and sex, basic information about their relationship). After the test part the participants were asked to justify their choices. Based on the answers it was possible to identify if the responder interpreted the utterance as a directive or a commissive speech act, which also had an influence on the interpretation of WE. The findings show that virtual usage is influenced not only by interpersonal but also by spatial relations. The processing of spatial position plays a major role in the referential interpretation of WE when the participants come from similar socio-cultural backgrounds. By contrast, in cases where sex and age differences are pronounced, information pertaining to the social world of the discourse has a higher impact on referential interpretation; nevertheless, some significant differences can be observed regarding which spatial relations or social relations had a major impact.

Keywords: virtual use of WE, spatial relations, social relation, intersubjective context, referential interpretation, attitude deixis, social deixis, personal deixis, cross-categorization

Korpuszneveink helyesírásáról

1. Bevezetés

Manapság már magától értetődőnek számít az a gyakorlat, hogy a nyelvleírás, a különféle nyelvészeti kutatások nagy mennyiségű elektronikus szöveget tartalmazó adatbázisokra, korpuszokra támaszkodnak. Fontos felhívni a figyelmet arra a kettősségre, hogy a *korpusz* szó a nyelvtudományban kétféle értelemben is használatos. Az Idegen szavak szótára (Tolcsvai Nagy 2007) – számos más területen használt jelentés mellett – az alábbi két értelmezését adja meg:

1. 'meghatározott módszerrel és előismeretekkel összegyűjtött nyelvi vagy irodalmi adatmennyiség, amely a tudományos kutatás vagy vizsgálat alapja'
2. 'számítógépre vitt és elemző programokkal előzetesen feldolgozott, további kutatásokra (pl. gyakorisági vizsgálatokra) alkalmas különböző szövegtípusokból gyűjtött szövegmennyiség'

A számítógépes nyelvészet csak a 2. jelentésben meghatározott kritériumoknak megfelelő szöveggyűjteményeket nevezi korpusznak, vagyis (a Magyar nemzeti szövegtár korpuszdefiníciójára támaszkodva) olyan – ténylegesen előforduló – írott vagy lejegyzett beszélt nyelvi adatok gyűjteményét, ahol a szövegeket valamilyen szempont szerint válogatják és rendezik, feltüntetik a szövegek bibliográfiai adatait, és jelölik a szövegek szerkezeti egységeit (bekezdés, mondat), továbbá bármilyen egyéb annotációt is tartalmazhatnak (pl. szófaji címkézés, morfológiai elemzés stb.). Fontos különbség a kétféle értelmezés között, hogy míg a 'valamilyen kutatás alapjául szolgáló gyűjtött adatmennyiség' értelemben a korpusznak nem feltétlenül szükséges elektronikus formájának lennie, lehet például papíralapú, kéziratos is, addig a számítógépes nyelvészetben használt *korpusz* terminus magában foglalja azt, hogy elektronikus formájú adatmennyiségről van szó. Jelen dolgozat a továbbiakban kizárólag a második jelentésben használja a *korpusz* műszót.

A korpuszok természetesen nem összefüggő szövegeként olvasandók, hanem azt a célt szolgálják, hogy megadott kritériumoknak megfelelő szavakat vagy szókapcsolatokat keressünk bennük. A korpuszoknak fontos tartozéka az úgynevezett keresőfelület, lekérdezőfelület, amelynek segítségével beállíthatjuk azokat a feltételeket, amelyeknek megfelelő eredményeket szeretnénk kapni.

1.1. A Nemzeti korpuszportál

Az elmúlt két évtizedben több, különféle célból készült korpusz látott napvilágot, ezek (kevés kivétellel) publikusak, szabadon felhasználhatók (esetleg előzetes, ingyenes regisztrációt igényelnek). Annak érdekében, hogy minél több magyar nyelvű, szabadon hozzáférhető korpusz megtalálható legyen egy helyen, 2015 novemberében létrejött a Nemzeti korpuszportál, amely az MTA Nyelvtudományi Intézetének honlapjáról érhető el (Sass 2017). A Nemzeti korpuszportál mellett a Nyelvtudományi Intézet honlapjának Adatbázisok menüpontjából is számos elektronikus szövegtár, tágabb értelemben vett adatbázis is elérhető.

2. Problémafelvetés, célok

Mínthogy a korpuszok a nyelvészeti kutatásoknak manapság elengedhetetlen segédeszközei, a kutatási beszámolóknak is jelölni kell, mely korpuszok szolgálták az adott kutatás forrásául. Ezért célszerű, hogy a korpusznevek egységes (és normatív) helyesírással legyenek feltüntetve. A korpusznevek a tulajdonnevek egy sajátos alcsoportját alkotják, az akadémiai helyesírási szabályzat által felsorolt tulajdonnévtípusok közül a címekhez lehet őket besorolni, ilyenformán tehát (ha nem is explicit módon) létezik normatív helyesírási szabályozás a korpusznevek írásmódjára.

Hogy miért szükséges mégis foglalkozni a korpusznevek helyesírási kérdéseivel, annak az oka, hogy részint a korpusznevek írásmódja a gyakorlatban meglehetősen ingadozást, bizonytalanságot mutat; részint pedig mert az úzus erősen hajlik a többtagú korpusznevek csupa nagybetűs írásmódjára, ami azonban korántsem tekinthető normatívnak.

A dolgozatban bemutatom a létező korpusznelvezések lehetséges típusait a Nemzeti korpuszportál és egyéb források alapján, ismertetem az írásgyakorlatban megfigyelhető tendenciákat, végül javaslatot teszek a korpusznevek egységes, normatív írásmódjára. Mivel a több tagból álló korpusznelvezések meglehetősen hosszúak, ezért ezekre többnyire rövidített formában (többnyire betűszókkal, szóösszevonásokkal) szokás hivatkozni. A dolgozatban kitérek az elnevezések rövidített formáinak írásmódjára is.

3. A létező korpusznelvezések főbb típusai

A Nemzeti korpuszportál jelenleg tizenegy korpusznevet tartalmaz. A Nyelvtudományi Intézet Adatbázisok menüpontjából további korpuszok, adatbázisok érhetők el, felvettem továbbá adataim közé a nyilvánosan nem elérhető Pázmány-korpuszt is (Endrédi 2016), illetve a készülő Magyar orvosi nyelv korpuszát is (Kuna 2016; Kuna–Kocsis–Ludányi 2017). Az 1. táblázat összefoglalva

tartalmazza azokat a korpuszokat, amelyek vizsgálatom anyagát képezik. A táblázatban a korpuszokat adatolt írásmódjukkal tüntetem fel, vagyis abban a formában, ahogy a Nemzeti korpuszportálon, illetve az Adatbázisok oldalon előfordulnak. Ez többnyire azonos a korpuszok saját lekérdezőfelületén szereplő írásmóddal, de előfordulhat az is, hogy eltér a kettő. Ez az ingadozás jól mutatja a korpusznevek írásmódja körüli bizonytalanságot. A táblázat harmadik oszlopában az adott korpusz alternatív elnevezése olvasható.

1. táblázat. *A vizsgálat korpusznevek teljes és rövidített formáinak adatolt írásmódja (a Nemzeti korpuszportál és a Nyelvtudományi Intézet honlapjának Adatbázisok menüpontja alapján)*

A korpusz elnevezés adatolt írásmódja	Rövidített formában	Alternatív elnevezés
„új” Magyar történelmi szövegtár	MTSZ	
Mikes-szótár		
Magyar Nemzeti Szövegtár	MNSZ2	
Történelmi Magánéleti Korpusz		Ó- és középmagyar történelmi magánéleti korpusz
Ómagyar korpusz		Régi magyar konkordancia
Budapesti Szociolingvisztikai Adatbázis	BUSZI	
Hunglish Korpusz		
Mazsola		
Webkorpusz		
Magyar Nemzeti Szövegtár	MNSZ1	
Magyar történelmi szövegtár		A magyar nyelv nagyszótárának korpusza
Orwell: 1984		
Magyar spontán beszéd adatbázis	BEA	
E-Szókincs		
Uralonet		
Pázmány Korpusz		
Magyar orvosi nyelv korpusza		Magyar orvosi nyelvi korpusz

A korpusz elnevezéseket felépítésük szempontjából háromféle csoportba sorolhatjuk. Az egyes típusoknál tárgyalom továbbá a felmerülő helyesírási kérdéseket.

3.1. Egytagú korpusznevek

Kevés ilyen típusú korpusz név létezik. Ilyen például a *Mazsola*, a *Webkorpusz*, az *Uralonet* és az *E-Szókincs*, bár az utóbbi nem is nevezhető igazán korpusznak, hanem inkább egyfajta adatbázisnak. Mivel azonban vizsgálatom szempontjából e különbség nem releváns, a továbbiakban eltekintek a különbségtételtől. Az ilyen típusú korpusznevek helyesírása nem különösebben problémás, mivel a többtagú korpusz elnevezések írásgyakorlatában megfigyelhető, minden tagot nagy kezdőbetűvel író gyakorlat értelemszerűen itt nem fordulhat elő. Egyedül az *E-Szókincs* helyesírása vet föl kérdéseket, nem tűnik ugyanis indokoltnak az *e-* 'elektronikus' előtagot követő nagy *sz*. Mivel az *e-* előtag szorosan összekapcsolódik az azt követő utótaggal, összetételről van tehát szó, amely kötőjelet tartalmaz, az utótagot nem kell nagy kezdőbetűvel írni. (Az *e-* előtagos neologizmusok helyesírási kérdéseiről l. Súlyom 2012.)

3.2. Tulajdonnévvel alkotott összetételek mint korpusznevek

Ebbe a típusba sorolható a normatív helyesírású *Mikes-szótár*, valamint a *Pázmány-korpusz*. Ez utóbbi esetén a csupa nagybetűs *Pázmány Korpusz*, illetve a címszerű *Pázmány korpusz* írásmód váltakozik egymással, akár egy szövegen belül is (l. Endrédi 2016), amely egyértelműen jelzi az írásmód körüli bizonytalanságot. A minden tagot nagy kezdőbetűvel írás nem normatív, mivel a korpusz elnevezések nem állandó, hanem egyedi címek (hasonlóan a könyvek, cikkek, képzőművészeti alkotások, zeneművek címéhez). A *Pázmány* és a *korpusz* szavak közti kapcsolat többféleképpen értelmezhető: amennyiben minőségjelzős (kijelölő jelzős) szerkezetnek tekintjük az alakulatot, ahol a minőségjelző tulajdonnév (vö. *Pista bácsi*), úgy védhető a *Pázmány korpusz* írásmód (a jelzett szó, a korpusz fogalomkörét leszűkíti a Pázmány nevű egyedre, vagyis a korpuszok közül a Pázmány elnevezésű). Valószínűbbnek tűnik azonban, hogy jelentésszűkítő összetételről van szó: 'olyan korpusz, amelyet a Pázmány egyetemen hoztak létre, ehhez az egyetemhez köthető korpusz'), így célszerűbbnek tűnik a tulajdonnévvel alkotott összetételek szokásos írásmódja, a kötőjelezés: *Pázmány-korpusz*.

3.3. Többtagú korpusznevek

Korpuszneveink legnagyobb része több tagból áll. Grammatikai felépítésük alapján két csoportot különíthetünk el: nagyobb részük minőségjelzős szerkezet (többszörösen bővített jelzővel), kisebb részük birtokos jelzői alárendelő szintagma.

a) Minőségjelzős szerkezetek

Az általam vizsgált korpusznevek közül ebbe a csoportba tartozik (maradva az adatolt írásmódnál): *Magyar történeti szövegtár*, *Magyar Nemzeti Szövegtár*, *Történeti Magánéleti Korpusz*, *Ómagyar korpusz*, *Budapesti Szociolingvisztikai Adatbázis*, *Ó- és középmagyar történeti magánéleti korpusz*, *Régi magyar konkordancia*, *Magyar orvosi nyelvi korpusz*.

b) Birtokos jelzői alárendelő szerkezetek

Ide sorolható *A magyar nyelv nagyszótárának korpusza*, illetve a *Magyar orvosi nyelv korpusza* elnevezés (utóbbi esetben a határozott névelő nem része a korpusz nevének). Az orvosi nyelvi korpusz előfordul *Magyar Orvosi Nyelv Korpusza* formában is, l. Kuna 2016.

A többtagú korpusznevek esetén az a tendencia figyelhető meg, hogy a két-, de különösen a gyakori háromtagú alakulatok írásmódja általában állandócím-szerű, vagyis az úzus a korpusznév minden tagjának nagy kezdőbetűs írását részesíti előnyben: *Magyar Nemzeti Szövegtár*, *Történeti Magánéleti Korpusz*, *Budapesti Szociolingvisztikai Adatbázis*. Ennek oka valószínűleg az angol írásmód hatása, mivel az angol nyelvterületeken ez a forma a szokásos: *British National Corpus*, *Cambridge International Corpus*, *Brown University Standard Corpus of Present-Day American English*, bár ez a tendencia más nemzetek korpuszainál is megfigyelhető: *Tycho Brahe Parsed Corpus of Historical Portuguese*, *Das Deutsche Referenzkorpus*. A magyar korpusz elnevezések írásmódja is (nem biztos, hogy tudatosan) feltehetőleg a nemzetközi gyakorlathoz kíván igazodni.

Ha azonban a korpusznév háromnál több tagú, az írásgyakorlat inkább a (normatív) címszerű írásmód felé hajlik: *Ó- és középmagyar történeti magánéleti korpusz*.

A kezdőbetűket tekintve megfelelő a *Magyar spontán beszéd adatbázis* írásmódja, azzal a kiegészítéssel, hogy a (*spontán beszéd*) + *adatbázis* alakulat írásmódjakor az AkH. 141. b) pontját kell alkalmazni (a szabályzat korábbi kiadása ezt mozgószabálynak nevezte). A javasolt írásmód ennek megfelelően: *Magyar spontánbeszéd-adatbázis*. Megjegyzendő, hogy a teljes korpusznév ritkán használatos, általában a rövidített formában, *BEA*-ként hivatkoznak rá készítői és a használói. Ez a gondolat át is vezet a következő témához, a rövidítésekhez és a mozaikszókhoz.

4. A korpusznevek rövid formái

Miként a gyakrabban használt szótáraknak is megvannak a szakmában megszokott, közkeletű rövidítései (*Értsz.*, *ÉKSz.*, *TESz.* stb.), a korpuszelvevéseknek is léteznek rövidített formái: *MNSZI*, *MNSZ2*, *BUSZI* stb. A szótárrövidítésekkel szemben – amelyek valódi rövidítések, vagyis nem lexémaértékűek (felolvasáskor és toldalékoláskor mindig a teljes formát vesszük alapul), a korpuszelvevések rövidített formái mozaikszók. Többnyire betűszók: *MNSZ*, *MTSZ* (= Magyar Történeti Szövegtár), de szösszevonás is akad köztük: *BUSZI*. (Bár ez utóbbinak a rövidített formája nem a korpusz lekérdezőfelületének elkészítésekor keletkezett, hanem jóval korábban, még a Budapesti Szociolingvisztikai Interjú munkálatai során [Váradai 2003].) A korpuszelvevések rövid formáinak kiejtése is a mozaikszókra jellemző: [emenesz], [emtéesz], [buszi], vagyis ezek az elnevezések lexémaértékűek.

Ha a korpusz teljes formájának írásmódja címszerű, vagyis csak az első kezdőbetű a nagy, célszerűnek látszik a rövidített formát is ennek megfelelően jelölni: *MNSz.*, *MTSz.*, ebben a formában jobban beleillenek a szótárak közkeletű rövidítéseinek sorába. Ugyanakkor az is igaz, hogy az *MNSz.* forma nem valódi rövidítés abban az értelemben, hogy a kiejtése gyakran [emenesz], a toldalék kapcsolása sem a teljes formához történik: nem **MNSz.-ban*, hanem *MNSz.-ben*. Így hasonlít például a *kft.* szóhoz, amely formailag rövidítés, kiejtése viszont a betűszókéhoz hasonlatos: [káefté]. A rövidítések és a mozaikszók között azonban nem olyan éles a határ, az *MNSz.* írásmód (rövidítésszerű helyesírás, betűszószerű kiejtés) véleményem szerint így is célszerűbb, mivel jobban illeszkedik a filológiában szokásos rövidítések sorába, mint a mozaikszavas megoldás. (Bár az utóbbira sem mondhatjuk, hogy nem normatív, már csak azért sem, mert a rövidítések, mozaikszók terén egyébként sem beszélhetünk merev szabályokról, legfeljebb tendenciákról, kialakult szokásokról).

5. Összefoglalás:

helyesírási javaslatok a korpuszelvevések írásmódjához

A korpusznevek helyesírását nem tárgyalja külön a hatályos akadémiai helyesírási szabályzat, ám a címek helyesírásával foglalkozó fejezet elég támpontot nyújt a kérdésben (AkH. 196–198.). A szabályzat kétféle címtípust különböztet meg: az állandó címeket és az egyedi címeket. Az egyes típusok helyesírása tükrözi a két típus jellegbeli kettősségét. Az időbeli rendszerességgel megjelenő kiadványok (újságok, folyóiratok) címei az állandó címek, ezeknek minden tagját nagy kezdőbetűvel írjuk. A korpuszelvevések azonban nem ebbe a csoportba tartoznak, hanem egyedi címeknek minősülnek, akárcsak az irodalmi művek, a cikkek, a képzőművészeti alkotások, a zeneművek, a műsorszámok stb. Az állandó címet viselő kiadványok tartalma ugyanis időről időre változik, míg a korpuszokra ez nem áll – eltekintve attól, hogy tartalmukat időnként bővítik újabb feldolgozott szövegekkel, mint ahogy például az *MNSz.* esetén is történt, valamint attól, hogy léteznek úgynevezett monitorkorpuszok is, amikor a régebbi tartalmakat újabakra cserélik oly módon, hogy a korpusz mérete ne változzon (a monitorkorpuszokról bővebben l. Szirmai 2005: 32, 171; Pajzs 2004). Ha a korpuszelvevéseket tehát egyedi címeknek tekintjük, akkor az írásmódjukra vonatkozó helyesírási szabály az, hogy típustól függetlenül csak az első szót írjuk nagy kezdőbetűvel: *Mazsola*, *Pázmány-korpusz*, *Magyar nemzeti szövegtár*, *Magyar történeti szövegtár*, *A magyar orvosi nyelv korpusza*. A tudomány nemzetközi színterein, ahol a tudományos kommunikáció nyelve az angol, a magyar korpuszok angol elnevezései használatosak: *Hungarian National Corpus*, *Hungarian Historical Corpus*, ezek esetében természetesen várható – igazodva a nemzetközi gyakorlathoz – a csupa nagy kezdőbetűs írásmód. Fontosnak tartom azonban, hogy a magyar nyelvű tudományos életben a magyar helyesírás rendszeréhez igazodva az egyedi címek írásmódját kövessük. Javaslatomat megerősíteni látszik az a tény is, hogy az AkH.¹¹-hez készített helyesírási segédkönyv, az *Osiris Helyesírás (OH.)* is ezt az írásmódot alkalmazza. Nem tér ugyan ki explicit módon a korpusznevek írásmódjára sem a szabálymagyarázó, sem a szótári részében, de az előszóban, amikor a példaanyag forrásait ismertetik a szerzők, következetesen címszerű írásmódot alkalmaznak: *Magyar nemzeti szövegtár*, *Magyar történeti szövegtár* (Laczkó–Mártonfi 2004: 12).

Bizonyos korpuszok esetén fontos lehet a verziószám feltüntetése is, például a *Magyar nemzeti szövegtár* újabb változata (Oravecz–Váradí–Sass 2014) olyan nagy mértékű változásokat tartalmaz az elődjéhez képest (Váradí 2002), hogy a korpusz második verziójaként szokás rá hivatkozni, megkülönböztetendő a sokkal kisebb méretű, elavultabb keresőfelülettel rendelkező első változattól. A verziószám feltüntetése véleményem szerint abban a formában lenne célszerű, miként az az akadémiai helyesírási szabályzatok, értelmező szótárak kiadászámának megjelölésekor szokásos. Ha az akadémiai helyesírási szabályzat tizenegyedik kiadásának közkeletű jelölése *AkH.*^{1,2}, a Magyar értelmező kéziszótár 2. kiadásáé pedig *ÉKSz.*², akkor ennek analógiájára a Magyar nemzeti szövegtár második verziójának a rövidítése lehetne *MNSz.*².

Jelen tanulmányban áttekintettem a legismertebb magyar nyelvű korpuszok alapján a főbb korpusznév típusokat és az írásgyakorlatban megfigyelhető helyesírási tendenciákat. Beillesztve a korpuszneveket a hatályos akadémiai szabályozás által felsorolt tulajdonnévtípusba (címek), javaslatokat fogalmaztam meg a korpuszok helyesírásával kapcsolatban. Tettem mindezt annak reményében, hogy idővel az írásszokás is szentesíti majd azokat.

SZAKIRODALOM

- AkH. = Magyar Tudományos Akadémia 2015. *A magyar helyesírás szabályai*. Tizenkettedik kiadás. Akadémiai Kiadó, Budapest.
- Endrédy István 2016. *Nyelvtechnológiai algoritmusok korpuszok automatikus építéséhez és pontosabb feldolgozásukhoz*. PhD-értekezés. Pázmány Péter Katolikus Egyetem, Budapest. Elérhető: https://itk.ppk.hu/uploads/articles/163/file/doi_Disszert%C3%A1ci%C3%B3_EIG_pdf_A.pdf.
- Kuna Ágnes 2016. A Magyar Orvosi Nyelv Korpusza. *Magyar Orvosi Nyelv*, 27–31.
- Kuna Ágnes – Kocsis Zsuzsanna – Ludányi Zsófia 2017. A Magyar orvosi nyelv 16–17. századi alkorpusza. Tervezet, átirás, annotálás. In: Forgács Tamás – Németh Miklós – Sinkovics Balázs (szerk.): *A nyelvtörténeti kutatások újabb eredményei IX*. SZTE Magyar Nyelvészeti Tanszék, Szeged, 239–53.
- OH. = Laczkó Krisztina–Mártonfi Attila 2004. *Helyesírás*. Osiris Kiadó, Budapest.
- Oravecz, Csaba – Váradí, Tamás – Sass, Bálint. 2014. The Hungarian Gigaword Corpus. In: Calzolari, Nicoletta – Choukri, Khalid – Declerck, Thierry – Loftsson, Hrafn – Maegaard, Bente – Mariani, Joseph – Moreno, Asuncion – Odijk, Jan – Piperidis, Stelios (eds.): *Proceedings of Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, 1719–23.
- Pajzs Júlia 2004. A korpuszalapú szótárírás alternatívái. In: Tóth Szergej – Földes Csaba – Fóris Ágota (szerk.): *Lexikológiai és lexikográfiai látkép: problémák, paradigmák, perspektívák*. Fasciculi Linguistici Series Lexicographica, Szeged, Generália, 134–41.
- Sass Bálint 2016. Nyelvészeti szövegkeresők, Nemzeti Korpuszportál. *Magyar Tudomány*, 798–808.
- Sólyom Réka 2013. E-előtagú neologizmusaink szemantikájáról. In: Váradí Tamás (szerk.): *VI. Alkalmazott Nyelvészeti Doktoranduszkonferencia*. MTA Nyelvtudományi Intézet, Budapest, 156–64.
- Szirmai Monika 2005. *Bevezetés a korpusznyelvészetbe*. Tinta Könyvkiadó, Budapest.
- Tolcsvai Nagy Gábor 2007. *Idegen szavak szótára*. Osiris Kiadó, Budapest.
- Váradí Tamás 2002. The Hungarian National Corpus. In: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)* (Las Palmas, 2002. május 29–31.). ELRA, Párizs, 385–9.
- Váradí Tamás 2003. A Budapesti Szociolingvisztikai Interjú. In: Kiefer Ferenc – Siptár Péter (szerk.): *A magyar nyelv kézikönyve*. Akadémiai Kiadó, Budapest, 339–59.

Ludányi Zsófia
tudományos munkatárs
MTA Nyelvtudományi Intézet

SUMMARY

*Ludányi, Zsófia***On the spelling of names of corpora**

Linguistic research today is often based on electronic corpora containing large amounts of text (e.g. Hungarian National Corpus, Hungarian Historical Corpus). The spelling of names of corpora, however, exhibits quite some indeterminacy: for names consisting of several words, it is typical that each word is capitalised, but other irregular forms also occur. The paper investigates the names of the most widely known Hungarian corpora from the point of view of normative orthography. Suggestions are made concerning a uniform spelling of names of corpora that is in line with the rules of Hungarian orthography. Furthermore, orthographic issues concerning the abbreviations of corpus names are also discussed.

Keywords: corpora, orthography, proper names, spelling of titles, abbreviations