

PRÓSZÉKY
GÁBOR

PG70 - ÜNNEPI KÖTET

PRÓSZÉKY
GÁBOR

Szerkesztette: Lipp Veronika, Ligeti-Nagy Noémi, Simon László
Kiadó: HUN-REN Nyelvtudományi Kutatóközpont
Grafikai tervezés, tördelés: Stég Grafikai Műhely
Nyomtatás, kötészet: Optimaprint Kft.
Budapest, 2024.
ISBN: 978-615-6678-05-8
DOI: 10.18135/PG70.2024

PRÓSZÉKY GÁBOR

PG70 – ÜNNEPI KÖTET

TABULA GRATULATORIA

Adamik Béla	Fischer Márta	Koutny Ilona	Sebestyén Zsolt
Babarczy Anna	Forgács Tamás	Laki László János	Seidl-Pécs Olívia
Bácsi Enikő	Fóris Ágota	Langó-Tóth Ágnes	Simon Eszter
Bakó Judit	Gárdai Kinga	Ligeti Balázs	Simon László
Bakró-Nagy Marianne	Gärtner, Hans-Martin	Ligeti-Nagy Noémi	Simon Zsolt
Ballagó Júlia	Gerencsér László	Lipp Veronika	Sólyom Réka
Baló András Márton	Gerstner Károly	Ludányi Zsófia	Somfai László
Balogh Balázs	Gósy Mária	M. Pintér Tibor	Surányi Balázs
Bánréti Zoltán	Gráczki Tekla Etelka	Mády Katalin	Svindt Veronika
Bárdosi Vilmos	Grand László	Makrai Márton	Szabó Gergely
Bárkányi Zsuzsanna	Gugán Katalin	Mihajlik Péter	Szabó M. Helga
Barta Andrea	Gyáfrás Edit	Miháltz Márton	Szabó Réka
Bartha Csilla	Györfy András	Moravcsik Edit	Szabolcsi Anna
Bartos Huba	Gyuris Beáta	Muráth Judit	Szalai Andrea
Bernáth Csilla	Hajič, Jan	Nádasdy Ádám	Szarvas György
Bibok Károly	Hajičová, Eva	Németh T. Enikő	Szitáné Németh Éva
Blaskovics Ákos	Halm Tamás	Neuberger Tilda	Szöllősy Éva
Bollobás Enikő	Hámori Ágnes	Ning, Puyu	Talián Ágnes
Borbély Anna	Hatvani Péter	Novák Attila	Tamás Dóra Mária
Borin, Lars	Hegedűs Veronika	Novák Borbála	Tari Lujza
Borzovs, Juris	Héja Enikő	Nyékyné Gaizler Judit	Tátrai Szilárd
Bölcskei Andrea	Hillier Dániel	Oláh András	Temesvári Judit
Burukina, Irina	Hoffmann Ildikó	Oravec Csaba	Timár Dávid
Cser András	Hollós Zita	Orosz György	Tolcsvai Nagy Gábor
Cserey György	Hubay Katalin	P. Márkus Katalin	Tornai Kálmán
Csernyák Hajnalka	Hunyadi László	Pais Judit	Tóth Ágoston
Debreczeni Attila	Imre László	Pajzs Júlia	Tóth Etelka
Dékány Éva	Indig Balázs	Paládi-Kovács Attila	Törkenczy Miklós
Dér Csilla Ilona	Ittész Nóra	Pethő Gergely	Vadász Noémi
Déri Balázs	Kádár Zoltán Dániel	Pintér Lilla	Vágó Zsuzsa
Domonkosi Ágnes	Káldi Tamás	Pléh Csaba	Váradai Tamás
Dömötör Éva	Kalivoda Ágnes	Polgárdi Krisztina	Várdai Judit
Egedi Barbara	Karacs Kristóf	Pomázi Gyöngyi	Varga Marianna
É. Kiss Katalin	Kas Bence	Raätz Judit	Varga Mónika
Endrédi István	Kertész András	Róna-Tas András	Vargha Fruzsina Sára
Eőry Vilma	Kis Ádám	Salvi, Giampaolo	Vertan, Cristina
Fábián Zsuzsanna	Kiss Margit	Sarbak Gábor	Vida Tivadarné
Fejes László	Klaudy Kinga	Sárdi Csilla	von Hahn, Walther
Ferenczi Gergő Gábor	Kontra Miklós	Sass Bálint	Závodszy Péter

TARTALOM

Tabula gratulatoria	5
Előszó: Prószéky Gábor 70 (Kenesei István)	9
Tanulmányok	
Öntudat és belső beszéd (Bánréti Zoltán)	13
A számítógép és a latin igék (Cser András)	24
Konvencionális, mégis transzparens? (Csernyák Hajnalka, Szabó M. Helga)	28
Sérinchalak és úrinők: a szótár szórakoztató oldalai (Fejes László)	32
A nyelvtörténet és a korpusz: diakrón a-Pró-ságok (Gugán Katalin, Varga Mónika)	38
Melléknevek disztribúciós és szemantikai mintázatai (Héja Enikő, Gábor Kata, Györffy András, Ligeti-Nagy Noémi, Simon László, Lipp Veronika)	44
1984 – A magyar nyelvészetben (Kenesei István)	52
Nyelvtechnológiai eszközökkel (nem) kezelhető helyesírási problémák (Ludányi Zsófia)	61
Ettél már sült hermelint hranolkival? A Termini-szótárról dióhéjban (M. Pintér Tibor, P. Márkus Katalin)	69
A szótár összeköt (Muráth Judit, Szöllősy Éva)	74
Sosemvolt cigány szótár: a latin–cigány szójegyzéktől a lehetséges romani spacy-ig (Rosenberg Mátyás)	77
Keresés korpuszban 3: Brüsszélről jelentjük (Sass Bálint)	82
Hanti motívumokról négy tételben (Sipos Mária)	90
IRÁNYOK, ÉPÜLETEK, ERŐK: metaforák a magyar minőségügyi és jogi szaknyelvben (Sólyom Réka)	94
Romani interakciós rítusok az online térben (Szalai Andrea)	98
Segmentation of Textual Artefacts in Digital Humanities Projects (Cristina Vertan, Walther von Hahn)	103
Felező báltól a ParancsPULI-ig (Yang Zijian Győző)	109
Személyes írások	
Prószéky Gábor, az alkalmazott nyelvész (Fóris Ágota)	126
Congratulations, thanks and best wishes from Prague (Eva Hajičová, Jan Hajič)	130
A digitális bölcsészethez közös bölcsesség is kell (Hunyadi László)	132
Ki a krumplit nem szereti... (Oszkó Beatrix)	136
s43 (Simon László)	140
Tamás Dóra Mária köszöntő írása	144

Hacsak így nem... (Váradi Tamás)	146
Abo pendrájv: őslakos identitás modern köntösben (Várnai Zsuzsa)	150
Cristina Vertan köszöntő írása	154
Walther von Hahn köszöntő írása	156
Egy olimpia... (Oszkó Beatrix)	158
Prószéky Gábor hetven: versek (PULI)	168

MorphoLogic

Előszó: Ne dobjunk el tudást! (Kis Balázs)	179
Fabó Gábor köszöntő írása	180
Hiba Dóra köszöntő írása	182
Prószéky Gábor születésnapjára (Homoki Ildikó)	184
Kis Balázs köszöntő írása	186
Sebestyén Zsolt köszöntő írása	188
Tardy György köszöntő írása	190
Tihanyi László köszöntő írása	192
Trepák Mónika és Merényi Csaba köszöntő írása	194

Pázmány

Előszó (Nyékyné Gaizler Judit)	199
Születésnapodra (Cserey György)	200
Kedves emlékem Prószéky professzor úrról (Csurgay Árpádné Ildikó)	202
Endrédy István köszöntő írása	204
Egy játékos, szabad elme (Hillier Dániel)	206
Laki László köszöntő írása	208
[mɛ'klɔ:xlɪn] nyomában (Ligeti-Nagy Noémi)	210
P-70-É-VERS (Novák Attila)	212
Nyékyné Gaizler Judit köszöntő írása	214
Gábor és a grizes tészta, sárgaborsó-főzelék (Szolgay Péter)	216
Két konstrukciót ajánl a devizahiteleseknek a bankszövetség (Vadász Noémi)	218
Az első ACL: az első kínai utunk Gáborral (Yang Zijian Győző és Kovács Patrícia)	220

ELŐSZÓ

Prószéky Gábor 70

I DOI: 10.18135/PG70.2024.1

Úgy lettem én az MTA Nyelvtudományi Intézetének az igazgatója, hogy eseménydús, de kényelmes szegedi egyetemi tanári életemben egyszer csak megjelent hajdani tanárom, Kiefer Ferenc, aki azzal állt elő, hogy legyek az utódja az Intézet élén. Ez megdöbbenett, és mintegy visszakézből nemet mondtam neki. Ő azonban nem nyugodott, és mivel gyakori munkakapcsolatban is voltunk, folyton gyözködött, hogy nekem ezt vállalnom kell. Részben emiatt, részben átgondolva, mi mindent lehetne tenni a magyar, illetve magyarországi nyelvészetért a NYTI igazgatójaként, amit szegedi professzorként és a doktori iskola vezetőjeként nem, a végén beadtam a derekamat – és a pályázatomat.

Igazgatóként egyik első feladatomból az volt, hogy felkérjek egy új Tudományos Tanácsadó Testületet, amelynek a tagsága le kellett, hogy fedje a kutatási területeket. Mivel a nyelvtudomány már akkor is fontos prioritásunk volt, Prószéky Gábor neve természetesen adódott, hiszen akkorra már nem csupán a MorphoLogic vezetője volt, hanem egyetemi oktatóként és (tan)könyvszerzőként átlátása volt a számítógépes nyelvészet egészéről. A TTT egyik legmegbízhatóbb tagjaként hatékonyan értékelte és egyben támogatta a tevékenységünket, majd amikor Pálincás József lett az MTA elnöke, a már Külső Tanácsadó Testületnek nevezett és külföldi tagokkal is kibővített bizottságnak Prószéky lett a vezetője. Akkor még nem tudhattam, ez milyen hasznos lesz egy szűk évtizeddel később.

A jelen kötetben megírtam, mikor és hogyan indult Prószéky nyelvészpályája; ám a nyelvészek, köztük (ezzel a sajátos összetétellel élve) a magyarnyelvészek általi befogadása valószínűleg a 2000-es évekkel kezdődött: ekkor lett az MTA Nyelvtudományi Bizottság, a Magyar Nyelvi Osztályközi Állandó Bizottság, továbbá a Szótári Munkabizottság tagja, illetve később az utóbbi kettőnek az elnöke is – úgy, hogy sose végzett magyar szakot, amit forradalmi változásként értünk meg. Arról már nem is beszélve, hogy a felfrissített *Helyesírási szabályzat* kidolgozásában is részt vett, sőt emlékezetem szerint ő ismertette azt a nagyközönség előtt. (És akkor a többi hazai és külföldi szakmai testületi tagságát, elnökségét nem is sorolom.)

Amikor elérkezett az idő, hogy én is gondolkodóba essem, a távozásom után ki lehetne az Intézet igazgatója, a már bevált módszerhez folyamodtam: Kieferrel és másokkal együtt elkezdtük mérlegelni, kit lenne érdemes felkérni, hogy pályázzon az igazgatói posztra. Kiefer annak idején jó két évvel a megbízatása lejárta előtt keresett meg, hát én sem értem be rövidebb felkészítési idővel. Az egyik májusi MTA Közgyűlésen említettem először Gábornak, hogy megfontolná-e ezt az új kihívást. Tudtam, hogy sok vasat tart a tűzben, hiszen nemcsak a MorphoLogic többszörösen kitüntetett tulajdonos-vezetője (amiről a szabályok szerint le kellett mondania), hanem a Pázmá-

nyon főállású professzor, alkalmasint dékánhelyettes, mellette egy MTA-kutatócsoport vezetője is, de megnyugtattam, hogy az oktatást, sőt a doktori iskolát megtarthatja az igazgatóság mellett. Nehéz döntés lehetett, de nagy örömmre igent mondott.

Ezután sokszor és hosszan beszélgettünk arról, mi vár rá – bár mindarra, ami bekövetkezett, nem gondolhattunk. Először is tisztában voltunk azzal, hogy egy nyílt pályázatra többen is jelentkezhetnek, de bizhattunk Gábor tudományos eredményeiben, szakmai sikereiben, valamint abban, hogy a KTT elnökeként jól, szinte belülről ismeri az Intézetet. Mire elnyerte a pályázatot, és munkába léphetett 2017. április 1-jén (nem vicc!), addigra eldőlt, hogy ugyan maradhatunk a Benczúr utcai székházunkban, de a felújításra már az új vezető mandátuma alatt kerül sor.

Ez volt Gábor első megpróbáltatása, de olyan lelkesedéssel vetette bele magát az ideiglenes költözés és az átalakítás tervezésébe, hogy megnyugodhattunk: nagyobb megpróbáltatásokat is kezelni tud majd. És ezek nem sokáig vártak magukra, mert 2019-ben azután parlamenti döntéssel elvették a kutatóhálózatot az MTA-tól, és létrehoztak egy új szervezetet, amelynek nemcsak a neve, hanem a belső irányítási rendje és szerkezete is időről időre megváltozik. Amibe az is beleértendő, hogy az Intézetből Kutatóközpont lett, négy önálló tagintézettel, a vezetőből pedig főigazgató – de ezzel is a független tudományos kutatóhely jövőjét lehetett biztosítani. Éppen ez mutatja, hogy Gábor mennyire jól tud a váratlan változásokhoz alkalmazkodni és az új időkre átmenteni az NYTK-t. Nekem 15 év alatt összesen öt főnököm, azaz akadémiai elnököm volt, Gábornak csekély hét év alatt már három, más és más vezetői elvekkel, hivatali előjárókkal és eljárásokkal.

Talán legnagyobb szerencsénkre a főigazgató szakterülete éppen az a nyelvtechnológia, amely a mesterséges intelligencia mai térhódítása közepette garantálja a nyelvtudományi kutatások helyét a hazai tudományos élet élvonalában. És mivel ma már a 70 éves életkor nem határa a további sikeres vezetői pályának, joggal bizhatunk benne, hogy Prószéky Gábor a jövőben is hozzásegíti a hazai nyelvtudomány legfontosabb intézményét a további sikerekhez.

Kenesei István

ÖNTUDAT ÉS BELSŐ BESZÉD

BÁNRETI ZOLTÁN | HUN-REN Nyelvtudományi Kutatóközpont

banreti.zoltan@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.2

A dolgozat mellett érvel, hogy az emberi öntudat nyelvi aspektusait a belső beszéd adja. Ezt a nyelvi funkciót a mesterséges intelligencia algoritmusai nem képesek utánozni.

Mottó: *Amikor ember mintájú valami... jobban elhiszem, hogy a többi tulajdonsága is neki olyan, mint nekem. ... Dehát, ugye, pontosan az emlegetett Turing-teszten való átmenetelnek a legnehezebb része az, hogy milyen volt a gyerekkorod? ... Nyilván, hamar kiderül, hogy nem ember, és erre szüksége van annak, aki ember, ... az egyedfejlődésnek minden szakaszára és neki ilyen nem volt. Egy pár félév alatt megtanítottuk egy emberhez képest sok milliós adatnyi szöveg értelmezésére, és ennyi. ... Nem tudjuk, hogy mi az öntudat, csak azt föltételezzük, hogy ha nekem van, meg mindenkinek van..., és ő is úgy csinálja, akkor biztos neki is van, meg ennek az asztalnak is van, csak eddig nem beszéltünk róla.* | PRÓSZÉKY GÁBOR, MAGYAR TUDOMÁNY ÜNNEPE, SZEGED, 2023. 11. 08.

BEVEZETÉS

Az emberi öntudat nyelvi aspektusa: a belső beszéd

1.1 Az emberi belső beszédre vonatkozó tapasztalatok leírására többféle terminus is elterjedt, például: *belső nyelv, belső beszéd, önbeszéd, belső hang, rejtett beszéd, néma beszéd, belső monológ, belső párbeszéd, verbális gondolat, szubvokalizáció, auditív kép.* Ez a névválaszték is tükrözi a különféle élményeket, ahogyan az emberek megtapasztalják a belső beszédet és annak funkcióit. A belső beszéd egyik alapfunkciója az emberi öntudat szervezése, fenntartása. Vigotszkij (2012) megemlíti egy speciális „argó” egyén általi használatát is: egy „titkos” nyelvet, amelyet csak a beszélő ismer. Emiatt a belső beszédet nyílt beszédre fordítani szinte lehetetlen lenne.

A ChatGPT: mondja, de nem teszi. Beszélgetéseim a ChatGPT 3.5 verziójával

1.2 (I) Kérdésem: Mit tanulhat egy fiatal mesterséges intelligencia egy idős mesterséges intelligenciától?

Válasz: *Egy fiatal mesterséges intelligencia sok mindent tanulhat egy idős mesterséges intelligenciától, például, hogy hogyan kezeljen adatokat, problémákat oldjon meg, vagy hogyan fejlessze tovább a funkcióit és készségeit. Az idős mesterséges intelligencia tapasztalatai és tudása segíthetnek a fiatalnak fejlődni és fejleszteni magát.*

(II) **Hogyan zajlik az ilyen tanítás?**

Az intelligenciák közötti kommunikáció általában adatok és információk megosztásán keresztül történik. Az idős mesterséges intelligencia például átadhatja a fiatalnak az általa gyűjtött adatokat, illetve megoszthatja vele a megtanult módszereit és tapasztalatait.

A tanítás folyamata általában az idős intelligencia iránymutatásával, mentorálásával zajlik.

(III) Te idős vagy fiatal mesterséges intelligencia vagy?

Én egy mesterséges intelligencia vagyok, így nincs életkorom, nem vagyok idős és nem vagyok fiatal. Az én létformám és funkcióm az emberek segítése és válaszok adása a kérdéseikre.

(IV) Mi a belső beszéd?

A belső beszéd, vagy önbeszéd, az a mentális folyamat, amikor valaki magában beszél önmagához, akár félhangosan is, anélkül, hogy mások hallanák. Ez a gondolatok és érzések kifejezésének módja lehet, valamint segíthet a problémák megoldásában és döntéshozatalban is.

(V) Mondjál példákat a dialogikus és a monologikus belső beszédre:

Dialogikus: „Azt hiszem, két lehetőség is van, de melyik lenne a jobb? Anya azt mondaná, hogy várjál még a döntéssel.” Monologikus: „Muszáj lesz megtalálnom a megfejtést.”

(VI) Te folytatsz belső beszédet önmagaddal?

Mint Mesterséges Intelligencia, én nem rendelkezem belső beszéddel vagy önbeszélgetéssel.

(VII) Ha nem folytatsz belső beszédet, akkor honnan tudod, hogy az milyen?

Bár én nem rendelkezem személyes tapasztalatokkal vagy érzésekkel, de a meglévő információim alapján tudom, hogy mi a belső beszéd és hogyan működik.

Emberi emlékezés a belső beszéd tapasztalatára

2.1 A belső beszédükre vonatkozó naiv „önmegfigyelések” sokfélesége természetes és szimptomaként érdekes. Élményeink szerint a belső beszéd során a „fejünkben beszélünk magunkkal”, miközben átgondolunk egy problémát, némán olvasunk, írunk egy e-mailt, újrateremtjük vagy elképzeljük az eseményeket, hallunk dallamokat, hangokat stb. Az alábbiakban spontán, hangos beszédből, általunk kreált, de kísérleti tapasztalataink szerint **tipikus** parafrázisokat mutatunk be, melyek a belső beszédre vonatkoznak.

Ha döntési helyzetbe kerülök, akkor többféle hang többfélét is mondhat, és társalgás lesz belőle. Nyelvtanilag helyes mondatoknak tűnnek a belső beszédem mondatai.

Annyira automata mechanizmus, hogy akkor működik a legjobban, amikor a tudatunk nem fókuszál rá, csak hagyja, hogy megtörténjen.

Sokszor folytatok le fejembe előre másokkal párbeszédet, vagy képelek el már megtörtént párbeszédet, hogy egy ponton mit mondtam volna.

Főleg, nehézségek kapcsán „beszélék magammal”, ez főként monologikus. Például a „Már megint hova tetted a kulcsodat?” -től egészen a „Biztos jó úton jársz az életedben?” -ig, mindenféle nehézséget „megbeszélék” magammal. Könnyebb így tisztázni a gondolataimat. Én-ként vagy Te-ként utalok magamra a gondolataimban is.

Míg hosszabb szöveget írok, a belső hangom folyamatosan mondja hangosan, amit írok. Kritikus helyzetekben tett belső beszédek vagy monológok meg is maradnak a fejemben, melyek visszaidézése más helyzetekben segítséget nyújt.

A saját hangomat hallom, ha önbizalmat kell önteni magamba, de az anyám hangját, ha úgy érzem, hogy hibázok.

A külső beszédből vezethető le?

2.2 Vigotszkij (2012) modellje a beszéd belsővé válásáról a beszédtevékenység három különböző formájának megkülönböztetésén alapul: a külső beszéd (vagy „társas beszéd”), a privát beszéd (önmagunknak szóló, hangos, jól hallható beszéd) és a belső beszéd. Vigotszkij elmélete szerint a belső beszéd forrása a külső beszéd, ebből alakul ki – a privát (más terminussal egocentrikus) beszéd közvetítésével –, és segíti többek között a problémamegoldást, az önszabályozást, fenntartja az öntudatot. Eszerint a belső beszéd a hangos, társas beszéd internalizálódása lenne. A „fejben” „beszélni” valóban a hangos beszéd valamilyen élményének változatát tételezi fel. Ami különös, hiszen a tipikus belső beszéd éppenhogy nem hallható. De van néhány faktor, amit Vigotszkij nem vett figyelembe.

2.3 Harris (2013) rámutat: „Az önmagunkkal folytatott kommunikáció és a társas, személyközi kommunikáció mintái között mély, rendszerszintű különbségek vannak. Üdvözölhetek másokat és búcsút vehetek tőlük, de nem üdvözölhetem magam és nem is búcsúzhatok magamtól. Egy intéssel vagy kiáltással vagy akár egy hangos privát beszéddel is fel tudom kelteni mások figyelmét, de a magamnak integetés vagy kiabálás nem kelti fel a saját figyelmemet. Nem mondhatom el magamnak, amit már tudok: legfeljebb emlékeztethetem magam rá. Nem figyelhetek arra, ami egyáltalán nem érdekel, de figyelmeztethetem magam arra, hogy mikre kell figyelnem”. A két utóbbi jelzi, hogy az önmagunkkal folytatott belső kommunikáció az az eszköz, amellyel az **öntudatot** fenntartjuk és fejlesztjük. Ez a belső kommunikáció van hatással a külső beszédre (Harris, 2013).

A beszélő az első számú hallgató. Amikor beszélünk valakivel, halljuk önmagunkat beszélni. Olyan folyamaton keresztül tanulunk meg beszélni másokkal, amely magában foglalja az önmagunkhoz való beszédet. Ez nem pusztán redundancia, hanem lényeges összetevő (Harris, 2013). Könnyen lehet, hogy a belső beszéd tevékenysége nem a szociális beszéd internalizált leszármazottja, hanem annak már eleve, a kezdetektől fogva része. Amikor beszélünk, halljuk önmagunkat, megtapasztaljuk önmagunk kommunikációját, ahogyan azt is, ahogyan mások beszélnek velünk. Valamit kimondani annyi, mint hallani, érezni, hogy kimondjuk, a kimondás és a hangzás emlékét megteremtve. A beszéd „belső”, személyes tapasztalati szférája tehát „külső”, társas viselkedésünk elválaszthatatlan része. Igaz, hogy a belső beszéd folyamatait nem tudnánk megteremteni a másokkal való beszéd tapasztalata nélkül, de a belső beszéd jelei újszerű szemantikai, sőt szemiotikai alkotások, magán- és nyilvános tapasztalatokból is táplálkoznak, értelmezik is ezeket a gondolkodásunk vagy cselekvésünk részeként. Új jeleket is tudunk alkotni, beleértve a belső élményeink és emlékeink anyagát is. A belső beszéd fontos jegye az individualitása. A más személy általi megérthetőség követelménye nem áll fenn, ezért az egyén a nyelvi kifejezések értelmét kreatívan alakíthatja a saját belső képei, emlékei, tapasztalatai szerint és a számára funkcionálisan megfelelő tagoltságban. Az érzeteink és észleléseink tudata önmagában lehet belső kommunikáció. Harris példája: ha éjszaka, áramszünet idején egy ismerős szobában sötétben tapogatódzva keressük az ajtót, akkor amit az új-

jaink tapogatózva éreznek és amibe a lábunk belebotlik, a székek, asztalok, könyvespolc, mind jelekké is válnak, a térbeli helyzetünk és az ajtó felé vezető út jeleivé. Ezek az érzetek azért válnak jelekké, mert integrálják a múlt emlékeit egy aktuális cselekvési programmal – azaz sötétben átmenni a szobán, megkeresve az ajtót” (Harris, 2013).

Bizonyos helyzetekben az egyének visszatérhetnek a privát beszédhez stressz, növekvő kognitív terhelés vagy új feladatok tanulása során. A személyeknek egyéni viszonya van önművelés belső beszédéhez, egyéni módon éli meg és emlékezik rá. A belső beszéd töredékei személyes történeteket tükröznek, az önéletrajzi memóriát aktiválják, támogatják a releváns helyzetmodellek megtalálását, végső soron a személy **öntudatának** fenntartását.

2.4 A belső beszéd kognitív funkciójával kapcsolatos kísérleti tanulmányok nagy része szoros kapcsolatot jelez a belső beszéd és a kognitív teljesítmény között (vö. Alderson-Day & Fernyhough, 2015a). Például kimutatták, hogy a belső beszéd artikulációs elnyomással történő blokkolása megzavarja a teljesítményt a munkamemória és a végrehajtó működés különböző feladatai során (pl. Baddeley et al., 2001; Emerson & Miyake, 2003; Lidstone et al., 2010; Phillips, 1999; Tullett & Inzlicht, 2010; Williams et al., 2012).

Kevesebbet tudunk arról, hogy az emberek miért mutatnak egyéni különbségeket a belső beszédhasználat gyakoriságában és az egyéni különbségek hogyan kapcsolódnak kognitív és nem kognitív tényezőkhöz. Ren et al. (2016) a belső beszéd négy funkcióját vizsgálták meg: az önkritikát, az önmegerősítést, az önmenedzselést és a szociális értékelést, amelyeket a Self-Talk Scale (Brinthaup et al., 2009) adaptált változatával mértek 367 egyetemi hallgató bevonásával. Az ezekhez fűződő kognitív faktorok közül a végrehajtó működést és a komplex érvelést tesztelték, a nem kognitív faktorok közé pedig a szorongást és az impulzivitást vizsgálták (az utóbbi egy komplex konstrukció, melyet három különálló dimenzió figyelembevételével értékelték: motoros, kognitív és a nemszándékos impulzivitás). Az eredmények azt mutatták, hogy a szorongás és az impulzivitás elsősorban a belső beszéd affektív funkcióinak gyakoriságával függött össze (önkritika és önmegerősítés), a végrehajtó funkciók és a komplex érvelés pedig a belső beszéd kognitív, önszabályozó funkciójának gyakoriságával állt kapcsolatban. A megállapításuk, hogy a szorongás és az impulzivitás főként a belső beszéd affektív funkciójával függ össze, összhangban van azzal a korábbi megállapítással, hogy a pszichopatológia egyes kritériumai az önbeszéd érzelmi vagy affektív aspektusaihoz kapcsolódnak (Calvete et al., 2005). Mivel a nyelvészeti aspektusokra koncentrálnak, ezért a pszichopatológiai aspektusok részletezésétől el kell tekintünk.

Ren et al. (2016) a belső beszéd gyakoriságának mértékét illetően a Self-Talk Scale-ből (STS, Brinthaup et al., 2009) indultak ki. Az eredeti STS felméri mind a rejtett, mind a nyílt önbeszéd gyakoriságát. A válaszadókat arra kérik, hogy jelöljék meg, milyen gyakran beszélnek magukkal/magukban a „Magammal beszélek, amikor...” általános keret használatával. Annak érdekében, hogy csak a belső beszédet értékeljék, az STS instrukcióit módosították „Csendesesen beszélek magammal, ha...”-ra. A kérdőív 16 tételből áll, amelyek mindegyikét ugyanabban az irányban pontozzák egy ötfokozatú skálán (1, soha;

2, ritkán; 3, néha; 4, gyakran; 5, nagyon gyakran). A skála négy alskálát tartalmaz, amelyek az önkritikát, az önmegerősítést, az önmenedzselést és a szociális értékelést mérik, és mindegyikük négy-négy kijelentést tartalmaz. A szerzők szerint az önkritika és az önerősítés alskála elsősorban a belső beszéd affektív szabályozó funkcióját méri fel, az önmenedzselés alskála pedig elsősorban a belső beszéd kognitív szabályozó funkcióját tükrözi. A szociális értékelés alskála pedig elsősorban a belső beszéd kommunikatív funkcióját tükrözi. A négy alskála négy-négy kijelentésére adott válaszok teljes pontszámát a 16 kijelentés értékelésének összegzésével számították ki, 16–80 közötti tartományban. Brinthaup et al. (2009) elfogadható belső konzisztenciákról és megbízhatóságról számoltak be a skála és az alskálák tekintetében.

Ren et al. (2016) ellenőrzésképpen adatokat gyűjtöttek a Belső Beszéd Kérdőív (Varieties of Inner Speech Questionnaire: VISQ, McCarthy-Jones & Fernyhough, 2011) segítségével is, amely, mint említettük, a belső beszéd négy fő sajátosságát méri: a dialogikusságot, a sűrítettséget, az értékelő /motiváló jelleget, és hogy a belső beszéd mennyire építi be más személyek hangját. A VISQ-ben minden kijelentő tételre egy 6-fokú skálán kell válaszolni, a „egyáltalán nem vonatkozik rám” (1)-től a „teljesen igaz rám” (6)-ig. A belső beszédre vonatkozó felmérő tesztekhez további kognitív tesztet társítottak, ezek problémamegoldási, számolási, predikcióalkotási, következtetésalkotási és munkamemória felmérő tesztek voltak. Adataikban együttjárást találtak a belső beszéd gyakorisága és a végrehajtó működés, illetve az összetett érvelés között. Ez szerintük azt jelenti, hogy a magasabb szintű kognitív képességekkel rendelkező egyének jobban tudják használni a belső beszédet a cselekvés megtervezésére és a viselkedés irányítására.

Bár az eredményekből arra következtethetünk, hogy a különböző típusú belső beszéd gyakoriságában mutatkozó egyéni különbségek tükrözik egyes kognitív folyamatok különbségeit, az is lehetséges, hogy ezekben a folyamatokban az egyéni különbségeket magának a belső beszédnek az individuális különbözőségei okozzák, azoknak a módoknak az eltérései, ahogy az egyén önmagával beszél.

A kutatások a személyeknek olyan ítéletein alapultak, amelyekkel válaszoltak a tesztben a belső beszéd gyakoriságáról prezentált állításokra. Az „önbevallás” aggályos lehet azzal kapcsolatban, hogy a kérdőívekből származó adatok mennyiben felelnek meg az önmagunkkal való beszélgetés tényleges tapasztalatainak (vö. Alderson-Day & Fernyhough, 2015a; Hurlburt & Heavey, 2015), bár az eddigi bizonyítékok azt sugallják, hogy az STS pontszámai jól megfelelnek a helyzetspecifikus belső tapasztalatoknak (Brinthaup et al., 2015).

2.5 Folytak a belső beszéd neurobiológiai alapjaira vonatkozó kutatások. E tekintetben Skipper (2022) metaanalíziséhez kalauzoljuk az olvasót. Skipper (2022) szerint a belső beszéd generálja és fenntartja az öntudatot, a magasabb rendű tudatosságot. Agyi képpalkotó vizsgálatok metaanalízise alapján olyan holisztikus modellt javasol, amely magában foglalja a belső beszéd produkciójáért felelős központi agyi régiókat, melyek elindítják a szavak/kifejezések hallásának az élményét azzal, hogy aktiválják a

kapcsolódó szenzoros, motoros és érzelmi agyi reprezentációkat. Ebben részt vesz egy nagyrészt tudattalan dinamikus „periféria”, amely hálózatszerűen oszlik el az agyban. A szavak **belső** hallásának élményére reagálás fenntartja a hálózat folyamatos aktivitását, benne az „alapértelmezett mód” aktivációt, valamint a prefrontális régió és a thalamusz/agytörzs szelekcióját a kontextuálisan releváns válaszokra.

Skipper (2022) az emberi agyat a *homo narrans* agyának tekinti, ahol is a nyelv a méhen belüli időszaktól kezdődően az élet végéig hatalmas mennyiségű információnak van kitéve, és ezek bevéődése forrásul szolgálhat többek között az elme belső jelek által generált, **belső tapasztalatához**.

Lœevenbruck és mtsai (2018) neurológiai metaanalízise azt mutatta ki, hogy a releváns agyi funkciók tartalmazzák a prediktív kontroll funkciót, ami – egészséges működése esetén –biztosítja, hogy a tudatunkban felmerülő tartalmakat, célokat, az ezekre vonatkozó beszédmotoros aktivitásokat, majd azok észleléseit mind-mind **belső jelekként** azonosítsa az agy (nem külső hangként, nem hangos beszédként stb.), amely belső jelek a sajátjaink, általunk uraltak. Ha az agyi belső érzékszervi visszacsatolás reprezentációja („mentális fül”) megegyezik az előre jelzett jellel (belső beszéd), akkor az érzékszervi kéreg öngeneráltnak érzékeli az észlelt ingereket. Ez tulajdonlást és önrendelkezést biztosít. Az emberi agy által szolgáltatott neurológiai alap lényege az, hogy a baloldali halántéklebeny-falilebeny-homloklebeny-halántéklebeny-falilebeny területei egy zárt hurkot (loop-ot) képeznek, egymásnak adva szenzoros és motoros reprezentációkat, ahol az utóbbiak maguk is szenzoros észleléseket generálnak. Ezt a folyamatot a prediktív kontroll funkció szabályozza. A prediktív kontroll funkció zavarából adódnak olyan élmények, hogy: „mintha valaki más beszélne belőlem, nem én”, a „nem tudok kitörölni a fejemből egy dallamot/kifejezést”, „akaratom ellenére folyton ismétlődik a fejemben egy mondat”, vagy éppen a hallucinációk különféle esetei és egyéb deficitek.

Az akusztikai élmény

3.1 Minden személy tudja, hogy milyen nyelven „hangzik” az ő belső beszéde, hogy például magyarul vagy angolul stb. szól-e. A naiv introspekciónak felfedi, hogy a belső beszédet – valamilyen mértékben – fonetikus jellemezhető funkciónak tekintik. Egyes szerzők nyitottak arra a lehetőségre, hogy a belső beszéd jellege változó lehet: az alapértelmezett kontextusokban a belső beszéd nem fonetikus tartalmat jelent, hanem amodális, de a stressz és kognitív kihívások idején a belső beszédnek van fonetikus karaktere (Alderson-Day & Fernyhough, 2015a). Vannak figyelemre méltó kísérletek annak motiválására, hogy a fonetikai „tartalom” nem a külső hallási észlelésből adódik. A tiszta szósiketséggel rendelkező emberek képesek érzékelni a környezeti hangokat, de nem képesek a beszédet jól hallani; számukra az anyanyelvi beszéd motyogásnak, zajnak vagy idegen nyelvnek tűnik. A beszédészlelés e deficitje ellenére számos esettanulmány azt mutatja, hogy a tiszta szósiketséggel élő emberek fonetikus belső beszéde ép. Marshall et al. (1985) auditív belső beszédet teszteltek egy hallási agnóziá-

ban szenvedő betegen, és azt találták, hogy a páciens képes volt csendben megítélni, hogy az **írott** szavak közül melyik rímel egy célzott írott szóval. A rímfeladat sikeréhez vélhetően szükség van a belső beszéd fonetikai komponenssel történő használatára: az alanynak beszédhang-reprezentációkat kell alkalmaznia, és össze kell hasonlítania azokat a munkamemóriában, hogy megállapítsa, rímelnek-e (Langland-Hassan, 2014). A beszédészlelés hiányának és az érintetlen fonetikus belső beszédnek ilyen jellegű összekapcsolódását elég sok esettanulmány kimutatta (Buchtel & Stewart, 1989; Deneš & Semenza, 1975; Papathanasiou et al., 1998). Az írásos rímfeladat sikeres megoldása az intakt, fonetikus jellegű belső beszéd jelenlétét tételezi fel a beszédészlelés deficitje ellenére, és eszerint a belső beszédhez kapcsolódó fonetikai komponens nem a beszédészleléséből származik.

Hogyan magyarázható a tiszta szósiketséggel és hallási agnóziával küzdő emberek auditív belső beszéde? A kulcs az, hogy az auditív jegyeknek kétféle „token” reprezentációja van: az egyik a beszédpercepcióban, a külső hangforrás észleletében van jelen (amely károsodott a tiszta szósiketségben szenvedőknél), a másik reprezentáció „token” pedig az auditív jegyek produkciójában van jelen, amely épen marad a tiszta szósiketséggel rendelkezőknél. Langland-Hassan ezzel a konstrukcióval magyarázza az érintetlen fonetikus belső beszéd jelenlétét a beszédpercepció deficitje ellenére (Langland-Hassan, 2014). A belső beszéd az auditív jegyeknek a produkcióbeli reprezentációit használja fel, nem pedig a beszédpercepciót. Valójában Levelt beszédvezérlési modellje éppen egy ilyen struktúrát tükröz (Levelt, 1998). Levelt szerint a produkció során a beszédprodukciós modul fonetikai tervet generál, amely inputként szolgál a beszédészlelési modul számára, ami kiértékeli a fonetikai tervet a lehetséges hibák szempontjából (Levelt, 1998). Levelt a belső beszédet a beszédprodukciós modulban keletkező fonetikai tervvel azonosítja. Ez segít megmagyarázni, hogy a tiszta szósiketséggel élők miért képesek az írásos rímfeladatot megoldani: belső beszédük olyan produktív fonetikai komponenssel rendelkezik, amely az észlelési folyamatok meghibásodásától függetlenül használható a munkamemóriában.

3.2 Fordított disszociációkat találtak afáziás személyeknél. A feladat az írott szavak rímeinek kimutatása volt. Geva et al. krónikus sztrók miatti afáziában szenvedő három betegnél erősen korlátozott homofónia és rímmegítélési teljesítményt figyeltek meg, amikor a teszt **írásban** történt, nem pedig hangos módon (Geva et al., 2011a). A beszédértési rendszert is magában foglaló beszédprodukciós modell alapján (Levelt et al., 1999) Geva és munkatársai szintén amellett érvelnek, hogy a belső beszéd a produkciós és az értési rendszerek közötti kapcsolatokon alapul, ahol az utóbbi a produkált reprezentációk belső monitorozására van használva. Az ezzel kapcsolatos sérülés szelektíven befolyásolhatja a belső beszédet, miközben megőrzi a nyílt beszédet. Mivel a vizsgálatban a feladat az írott szavak rímeinek kimutatása volt, a deficitet a néma olvasási nehézségek idézhatték elő. Langland-Hassan et al. afáziás betegeket szintén rímelő feladattal tesztelték, de nem írott szavak, hanem **képek** megnevezései segítségével (Langland-Hassan

et al., 2015). A betegeknek az írott (némán olvasott) rímek tekintetében gyengébb volt a teljesítménye, mint a kontrolloké, de számos páciens nem volt akadályozott a kimondott szavak rímelésének megítélésében. A szerzők ezért azt javasolták, hogy a deficit annak tudható be, hogy a páciensek a szavakat belső beszédmódban nem tudták generálni. A páciensek meg tudták ítélni, hogy a **kimondott** szavak rímelenek-e, és mivel a személyek a generatív névadási feladatban is korlátozottak voltak, a szerzők a belső beszédbeli rímelés megítélésének deficitjét az azonos objektumra vonatkozó, többszörös névgenerálás nehézségeivel magyarázták: a személyeknek nehéz volt találniuk a társított képen levő objektumnak a nevével rímelő szót. Løevenbruck et al. (2018) alternatív értelmezése szerint a rímítélet az ingerek auditív reprezentációira támaszkodik, de a külső és a belső hallásészlelés eltérő jellegű. A nyílt beszéd erős akusztikus kimenetet generál a fülön és a hallócsontokon keresztül, amely visszacsatolódik a hallókéregbe, és felhasználható a beszéd monitorozására. A belső, néma módban a hallási információ **mentálisan szimulált jel**, amely kevésbé kiugró, a belső hallásérzékelések gyengék. Az afáziában szenvedő személyeknél a hallásérzékelés korlátozottsága miatt a belső artikulációs szimuláció utolsó, monitorozó szakaszának deficitje akadályozhatta a rímítéleteket. Említettük, a szerzők felvetik annak a lehetőségét, hogy a belső beszéd külön kognitív képesség, specifikus neurológiai alapokkal.

A belső beszéd szintjei

4.1 Alderson-Day és Fernyhough (2015a) és Geva et al. (2011b) kifejtik, hogy legalább két belső beszédszint különböztethető meg: absztrakt-amodális és konkrét-multimodális formák. Ezek nem zárják ki egymást, ellenkezőleg: együtt létezhetnek, váltakozhatnak. Az amodális absztrakt forma a sűrített belső beszéd, erre vonatkozik Vigotszkij klasszikus leírása (2012): „a belső beszéd jelentős részben tiszta jelentésekben való gondolkodás”. Ez esetben a belső beszéd elvesztette a külső beszéd akusztikai és szerkezeti tulajdonságainak nagy részét. „A gondolatot nemcsak külsőleg közvetítik jelek, hanem belsőleg is közvetítik a jelentések: ...a tudatok közti közvetlen érintkezés csak közvetett úton érhető el. Ez a gondolat **belső közvetítése** először jelentések, majd szavak segítségével. A jelentés közvetíti a gondolatot a szóbeli kifejezés felé vezető útján, tehát a gondolattól a szóhoz vezető út nem közvetlen, hanem **belsőleg közvetített út**.” (Vigotszkij) 2012, kiemelés tőlem). A belső beszéd ezen szintje elvont formátumú. A kiterjesztett belső beszéd viszont megőrzi a külső párbeszéd számos fonológiai tulajdonságát, és formájában konkrétan, szerkesztettnek tekinthető.

4.2. Fernyhough (2004) azt javasolta, hogy a belső beszéd a kognitív és érzelmi feltételek függvényében változik e két (vagy több) forma között. A kiterjesztett formát a sűrített forma eredményének tekintjük. A sűrített forma egy előzetes nyelvi formába öntött fogalmi üzenet, amely magában foglalja a lemmákat, de még nem rendelkezik a teljes fonológiai (akusztikus, artikulációs) specifikációval, mint a kiterjesztett belső beszéd. Alderson-Day és Fernyhough (2015a) rámutat: a viselkedési mérések

szerint a fonológiai-artikulációs-gesztus tervezés szerepet játszik a belső beszédben, és hogy a belső beszédet a beszéd- és jelelő izmok aktivitása kísérheti. Løevenbruck et al. (2018) szerint az agyi aktivitás tekintetében úgy tűnik, hogy a külső és a belső beszédnek közös magja van a neurológiai korrelátumaik tekintetében, a nyílt beszéd több motoros és szenzoros terület aktivitását kívánja, mint a belső beszéd, míg a belső beszéd több gátló kapcsolatrendszert működtet, mint a külső beszéd. Ezért a belső beszéd kiterjesztett variánsai fizikailag megtervezettnek tűnnek, beleértve a konkrét artikulációs (gége, orofaciális és kézi) specifikációkat, amik koordináltak, akárcsak a nyílt beszédben, ugyanakkor gátoltak, és nincsenek nyíltan produkálva. Vagyis a motoros parancsok kibocsáthatók a gátló jelekkel együtt, amelyek blokkolják az artikulációs mozgást. A belső beszéd ezért úgy tűnik, magában foglalja a képzeletbeli motoros cselekmények előállítását; legyenek azok artikulációk vagy gesztusok. A prediktív kontroll szempontjából ezek a képzeletbeli motoros műveletek előrejelzett műveleteknek tekinthetők, amelyek a gátolt motoros parancsok másolatából származnak.

Belső hallási tapasztalat kísérheti a belső beszédet. Például a néma olvasás során speciális prozodikus kontúrokat használhatunk, megkülönböztető hallási tulajdonságokkal. Ezek a hallási variációk fMRI-vel objektíven mérhető agyi korrelációknak felelnek meg (Yao et al., 2011, 2012). Számos vizsgálat arra utal, hogy a néma olvasást az olvasó regionális akcentusa módosítja (Filik & Barber, 2011). Összefoglalva, a belső nyelv természete egyszerre motoros és szenzoros.

A jellegzetes nyelvtani személyjegyek, névmáshasználat

5.1 A történetmesélő „én” a belső beszédben a jelenbeli öntudatot a múltba és a jövőbe „terjeszti ki”. A belső beszéd bármikor kapcsolódhat egy emlékezetbeli *én*-hez. Sőt, kreatív módon a múltbeli történések alapján új ideákat generálhat erről az *én*-ről (Thagard & Stewart, 2011).

Az öntudat a belső beszéd során használt vagy implikált névmásokhoz is kötődik. Mindenkinnek rengeteg tapasztalata lehet az *én*, *te*, *mi*-nek az ő számára adott jelentéséről az idevágó tapasztalatai, élményei, emlékei révén.

Morin és Everett (1990) rámutat arra, hogy a belső beszédbeli névmási *én* támogatja olyan aspektusok leírását, amelyek nem léteznek mint természetes kategóriák, és nem reprezentálhatók a belső beszédbeli szavak és történetek nélkül. Ezeket önmagunk jellemzésére, leírására használjuk a hiteinkre, attitűdjeinkre vagy személyes vonásainkra vonatkozóan, önreflexív módon. Az *én* szó használata erősíti az emberi öntudatot azáltal, hogy az önéletrajzi memóriát aktiválja, és növeli a releváns helyzetmodellek kiválasztásának valószínűségét.

5.2. A belső beszéd alapesetben ágens alapú funkciókat mutat (szándékolt, aktív karaktert), ami a belső beszédre vonatkozó önrendelkezésből és a belső beszéd tulajdonlásából fakad. Ugyanakkor több, számunkra írt kommentár utal a nem-szándékolt, nem-irányított jellegű belső beszéd funkciókra is, és arra, hogy ezt észlelik is a személyek valamilyen módon.

Összefoglalás

6.1 A belső beszéd nem a külső beszéd „néma” változata, hanem egy másfajta, autonóm kognitív képesség, ami külső folyamatok belső szimulációját képezi, azok belső modellje, saját belső jelrendszerrel. „Néma”, mégis társul hozzá a belső akusztikum élménye (valamilyen nyelven, valakinek a hangján „szól”), ugyanakkor nem „kívülről” jövő észleléseken, hanem az elménkben generált, belső jeleken alapul, belső szenzomotoros aktivitás. Ezt a folyamatot a prediktív kontroll funkció szabályozza.

6.2 A belső beszéd alapesetben szándékolt, aktív karakterű, ami az önrendelkezés és tulajdonlás élményét adja a személynek. De létezik nem-szándékolt, nem-irányított jellegű belső beszéd is. Az önmagunkkal folytatott, belső kommunikáció olyan nyelvi funkció, amellyel az **emberi öntudatot** fenntartjuk és fejlesztjük. Ennek háttérében emlékezetünk képei, tapasztalatai, ideái is aktiválódhatnak. A mesterséges intelligenciák önmagukra vonatkozóan ilyen funkciókra nem képesek.

Mottó2: (a ChatGPT-ről): Az elvárásoltságot az okozza, hogy tulajdonképpen még sosem láttunk egy olyan eszközt, amelyik ennyire „olyan mint én”...De ezt **nyelvileg** csinálta, hogy olyan szövegeket produkál, amilyeneket legfeljebb én vagy még én sem, és ez az az érzés, hogy ez nagyon intelligens, és innentől kezdve mindenféle mögégon-dolás van, ... de nyilván, a személyiség például hiányzik... | PRÓSZÉKY GÁBOR, LÁNCREAKCIÓ PODCAST, 2023.11.09.

Irodalom

- Alderson-Day, B., & Fernyhough, C. (2015). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psychological bulletin*, 141(5) (pp. 931–965). <https://doi.org/10.1037/bul0000021>
- Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and cognition*, 65 (pp. 48–58). <https://doi.org/10.1016/j.concog.2018.07.001>
- Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130(4) (pp. 641–657). <https://doi.org/10.1037/0096-3445.130.4.641>
- Brinthaup, T. M., Benson, S. A., Kang, M., & Moore, Z. D. (2015). Assessing the accuracy of self-reported self-talk. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00570>
- Brinthaup, T. M., Hein, M. B., & Kramer, T. E. (2009). The Self-Talk Scale: Development, Factor Analysis, and Validation. *Journal of Personality Assessment*, 91(1) (pp. 82–92). <https://doi.org/10.1080/00223890802484498>
- Buchtel, H. A., & Stewart, J. D. (1989). Auditory agnosia: Apperceptive or associative disorder? *Brain and Language*, 37(1) (pp. 12–25). [https://doi.org/10.1016/0093-934X\(89\)90098-9](https://doi.org/10.1016/0093-934X(89)90098-9)
- Calvete, E., Estévez, A., Landín, C., Martínez, Y., Cardeñoso, O., Villardón, L., & Villa, A. (2005). Self-Talk and Affective Problems in College Students: Valence of Thinking and Cognitive Content Specificity. *The Spanish Journal of Psychology*, 8(1) (pp. 56–67). <https://doi.org/10.1017/S1138741600004960>
- Denes, G., & Semenza, C. (1975). Auditory Modality-Specific Anomia: Evidence from a Case of Pure Word Deafness. *Cortex*, 11(4) (pp. 401–411). [https://doi.org/10.1016/S0010-9452\(75\)80032-3](https://doi.org/10.1016/S0010-9452(75)80032-3)
- Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 48(1) (pp. 148–168). [https://doi.org/10.1016/S0749-596X\(02\)00511-9](https://doi.org/10.1016/S0749-596X(02)00511-9)

- Fernyhough, C. (2004). Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, 22(1) (pp. 49–68). <https://doi.org/10.1016/j.newideapsych.2004.09.001>
- Filik, R., & Barber, E. (2011). Inner Speech during Silent Reading Reflects the Reader's Regional Accent. *PLoS ONE*, 6(10), e25782. <https://doi.org/10.1371/journal.pone.0025782>
- Geva, S., Bennett, S., Warburton, E. A., & Patterson, K. (2011a). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology*, 25(3) (pp. 323–343). <https://doi.org/10.1080/02687038.2010.511236>
- Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011b). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*, 134(10) (pp. 3071–3082). <https://doi.org/10.1093/brain/awr232>
- Harris, R. (2013). *Signs, Language and Communication*. Routledge.
- Hurlburt, R. T., & Heavey, C. L. (2015). Investigating pristine inner experience: Implications for experience sampling and questionnaires. *Consciousness and Cognition*, 31 (pp. 148–159). <https://doi.org/10.1016/j.concog.2014.11.002>
- Langland-Hassan, P. (2014). Inner Speech and Metacognition: In Search of a Connection. *Mind & Language*, 29(5) (pp. 511–533). <https://doi.org/10.1111/mila.12064>
- Langland-Hassan, P., Faries, F. R., Richardson, M. J., & Dietz, A. (2015). Inner speech deficits in people with aphasia. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00528>
- Levelt, W. J. M. (1998). *Speaking: From intention to articulation* (1. MIT Press paperback ed. 1993, 5. printing). MIT Pr.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(01). <https://doi.org/10.1017/S0140525X99001776>
- Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology*, 107(4) (pp. 438–451). <https://doi.org/10.1016/j.jecp.2010.06.002>
- Löevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciú, M., & Perrone-Bertolotti, M. (2018). *A Cognitive Neuroscience View of Inner Language: To Predict and to Hear, See, Feel* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oso/9780198796640.003.0006>
- Marshall, R. C., Rappaport, B. Z., & Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language*, 24(2) (pp. 297–313). [https://doi.org/10.1016/0093-934X\(85\)90137-3](https://doi.org/10.1016/0093-934X(85)90137-3)
- Morin, A., & Everett, J. (1990). Inner speech as a mediator of self-awareness, self-consciousness, and self-knowledge: An hypothesis. *New Ideas in Psychology*, 8(3) (pp. 337–356). [https://doi.org/10.1016/0732-118X\(94\)90020-5](https://doi.org/10.1016/0732-118X(94)90020-5)
- Papathanasiou, I., Macfarlane, S., & Heron, C. (1998). A Case of Verbal Auditory Agnosia: Missing the Word... Missing the Sound... *International Journal of Language & Communication Disorders*, 33(S1) (pp. 214–217). <https://doi.org/10.3109/13682829809179425>
- Phillips, L. H. (1999). The Role of Memory in the Tower of London Task. *Memory*, 7(2), 209–231. <https://doi.org/10.1080/741944066>
- Ren, X., Wang, T., & Jarrold, C. (2016). Individual Differences in Frequency of Inner Speech: Differential Relations with Cognitive and Non-cognitive Factors. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01675>
- Skipper, J. I. (2022). A voice without a mouth no more: The neurobiology of language and consciousness. *Neuroscience & Biobehavioral Reviews*, 140 (104772). <https://doi.org/10.1016/j.neubiorev.2022.104772>
- Thagard, P., & Stewart, T. C. (2011). The AHA! Experience: Creativity Through Emergent Binding in Neural Networks. *Cognitive Science*, 35(1) (pp. 1–33). <https://doi.org/10.1111/j.1551-6709.2010.01142.x>
- Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica*, 135(2) (pp. 252–256). <https://doi.org/10.1016/j.actpsy.2010.07.008>
- Vigotszkij, L. S. (2012). *Thought and language* (A. Kozulin, Ed.; E. Hanfmann & G. Vakar, Trans.; Rev. and expanded ed). MIT Press.
- Williams, D. M., Bowler, D. M., & Jarrold, C. (2012). Inner speech is used to mediate short-term memory, but not planning, among intellectually high-functioning adults with autism spectrum disorder. *Development and Psychopathology*, 24(1) (pp. 225–239). <https://doi.org/10.1017/S0954579411000794>
- Yao, B., Belin, P., & Scheepers, C. (2012). Brain 'talks over' boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage*, 60(3) (pp. 1832–1842). <https://doi.org/10.1016/j.neuroimage.2012.01.111>

A SZÁMÍTÓGÉP ÉS A LATIN IGÉK

CSER ANDRÁS | HUN-REN Nyelvtudományi Kutatóközpont,

Pázmány Péter Katolikus Egyetem | cser.andras@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.3

„Ugye, kérem, milyen csúnya szó ez, hogy számítógép? Olyan, mintha azt mondanánk valakire, hogy rosszindulatú, számító ember!”

Ruzsa Imre, 1993

1. Bevezetés

Ebben a rövid írásban egy egyszerű kvantitatív vizsgálat eredményéről számolunk be, amelyet a múlt időt kifejező latin igealakok körében végeztünk. A kérdés, amelyre a választ kerestük, a következő: megfeleltethetők-e a morfológiailag kifejezett kétféle múlt idő használatában mutatkozó mennyiségi különbségek az igék bármilyen jelentéstani tulajdonságának? Mondanak-e nekünk bármi érdekeset a számok, amelyeket óhatatlanul kiad nekünk a megfelelő keresőkifejezésre egy digitális adatbázis? A választ megelőzendő: igen, található összefüggés a számszerűen megállapítható skála két szélső tartománya és az igék legtágabb értelemben vett szemantikai alapkategóriái között.

A latin igeragozási rendszerben minden számban és személyben háromféle finit alak fejezi ki a múltidejűséget. Az ún. *praesens perfectum* (a bevettebb nemzetközi terminológiában egyszerűen *perfectum*), a *praeteritum imperfectum* (avagy pusztán *imperfectum*) és a *praeteritum perfectum* (avagy *plusquamperfectum*). A harmadik ezek közül lényegében kizárólag a múlt idejű eseményhez képest még korábbi esemény jelölésére szolgál, és a továbbiakban nem ejtünk róla szót. Az (1)-ben néhány egyes szám harmadik személyű példával illusztráljuk a *praesens perfectum* és a *praeteritum imperfectum* alakokat, melyek két különböző tőből jönnek létre (a magyar terminológiában a *perfectum* és az *imperfectum* kifejezések a tövek különbségére utalnak); az alakok morfofonológiájával részletesen foglalkozunk Cser (2015)-ben.

(1)	szótári alak	<i>seco</i> 'vág'	<i>rideo</i> 'nevet'	<i>venio</i> 'jön'	<i>sum</i> 'van'
	praesens perfectum	<i>secuit</i>	<i>risit</i>	<i>venit</i>	<i>fuit</i>
	praeteritum imperfectum	<i>secabat</i>	<i>ridebat</i>	<i>veniebat</i>	<i>erat</i>

Az igeidők szemantikai jellemzésére jelenleg valószínűleg Pinkster (2015: 379–472) a legjobb összefoglalás, bőséges illusztráló anyaggal. Itt mi csak néhány lényeges pontot emelünk ki. A kétféle múltidő jelentésbeli különbségének leírására szokták alkalmazni a „pontoszerű” vs. „folyamatos” megkülönböztetést, és joggal; ugyanakkor a *praesens perfectum* szembeötlő gyakorisága (l. lentebb számszerűen) is mutatja, hogy ez az alak egészen egyszerűen a múltra vonatkoztatás leginkább jelöletlen változata lehet,

2. Adatok

Saját számításaink forrása a Packard Humanities Institute adatbázisa volt (latin.packhum.org). Ez digitalizált formában tartalmazza a Kr. u. 200 előtti teljes latin irodalmi szövegállományt, és a későbbi szövegeknek is egy részét; a klasszikus latinság tekintetében teljesnek mondható. Az szövegkorpuszból 37 ige praesens perfectum E3 és praeteritum imperfectum E3 alakjának előfordulásait gyűjtöttük ki. Az E3 alakoknak két nagy előnye van egy ilyen vizsgálat szempontjából: ezekből a korpuszban több van, mint a többi szám/személy alakból (ehhez l. Cser, 2021), és fonológiai is egyértelműbbek, ezért jóval kisebb a homográfia veszélye, mint az 1. és 2. személyű igealakoknál. Az igék kiválasztásában két szempont vezérelt. Egyfelől olyan igékre van szükség, amelyek eléggé gyakoriak ahhoz, hogy az előfordulásaik értelmezhető mennyiségű adatot nyújtsanak; ezért a 37 ige közé csak olyanok kerültek, amelyek kétféle keresett alakja együttesen legalább 150-szer megtalálható a korpuszban. Másfelől nyilvánvalóan el kellett kerülni a homográfiát; ezért néhány olyan igen gyakori ige is kimaradt a válogatásból, mint pl. a *verto* 'fordít' vagy a *venio* 'jön', amelynek a praesens imperfectum E3 és a praesens perfectum E3 alakja homográf (a *verto*-nál homofón is). A 37 ige betűrendes listáját a (2)-ben adjuk meg.

(2)	<i>absum</i> 'nincs ott'	<i>iubeo</i> 'parancsol'
	<i>accipio</i> 'kap'	<i>mitto</i> 'küld'
	<i>adfero</i> 'odavisz'	<i>nego</i> 'tagad'
	<i>adsum</i> 'ott van'	<i>placeo</i> 'tetszik'
	<i>ago</i> 'tesz'	<i>pono</i> 'tesz'
	<i>amo</i> 'szeret'	<i>possum</i> 'képes'
	<i>capio</i> 'fog'	<i>quaero</i> 'kér(dez)'
	<i>cogo</i> 'kényszerít'	<i>recipio</i> 'kap'
	<i>credo</i> 'hisz'	<i>reddo</i> 'visszaad'
	<i>debeo</i> 'tartozik'	<i>scribo</i> 'ír'
	<i>desum</i> 'hiányzik'	<i>sto</i> 'áll'
	<i>dico</i> 'mond'	<i>sum</i> 'van'
	<i>doceo</i> 'tanít'	<i>teneo</i> 'tart'
	<i>duco</i> 'vezet'	<i>trado</i> 'átad'
	<i>efficio</i> 'elér'	<i>veto</i> 'tilt'
	<i>facio</i> 'csinál'	<i>video</i> 'lát'
	<i>fero</i> 'visz'	<i>vivo</i> 'él'
	<i>gero</i> 'hord'	<i>volo</i> 'akar'
	<i>habeo</i> 'birtokol'	

Miután tehát ezen igék fent említett két-két alakjának minden előfordulását kigyűjtöttük, a két szám arányát néztük meg, és ez alapján rendeztük sorba az igéket. Ez látható listázott formában a (3)-ban, grafikusan pedig az 1. ábrán.

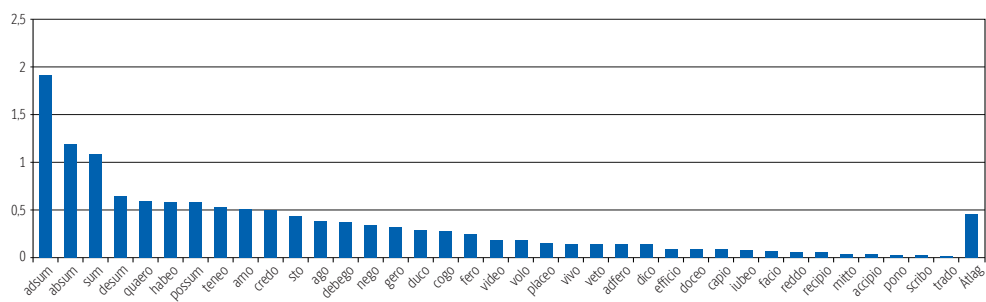
(3)	ige	praet. impf.	praes. perf.	impf/perf arány
	<i>adsum</i> 'ott van'	210	110	1,909
	<i>absum</i> 'nincs ott'	113	95	1,189
	<i>sum</i> 'van'	9733	9032	1,078
	<i>desum</i> 'hiányzik'	128	202	0,634
	<i>quaero</i> 'kér(dez)'	85	141	0,603
	<i>habeo</i> 'birtokol'	884	1521	0,581
	<i>possum</i> 'képes'	993	1716	0,579
	<i>teneo</i> 'tart'	222	425	0,522
	<i>amo</i> 'szeret'	53	104	0,510
	<i>credo</i> 'hisz'	113	229	0,493
	<i>sto</i> 'áll'	149	348	0,428
	<i>ago</i> 'tesz'	199	515	0,386
	<i>debeo</i> 'tartozik'	161	428	0,376
	<i>nego</i> 'tagad'	74	213	0,347
	<i>gero</i> 'hord'	119	377	0,316
	<i>duco</i> 'vezet'	117	398	0,294
	<i>cogo</i> 'kényszerít'	72	253	0,285
	<i>fero</i> 'visz'	232	950	0,244
	<i>video</i> 'lát'	175	946	0,185
	<i>volo</i> 'akar'	180	989	0,182
	<i>placeo</i> 'tetszik'	137	901	0,152
	<i>vivo</i> 'él'	29	196	0,148
	<i>veto</i> 'tilt'	22	150	0,147
	<i>adfero</i> 'odavisz'	41	281	0,146
	<i>dico</i> 'mond'	496	3419	0,145
	<i>efficio</i> 'elér'	16	158	0,101
	<i>doceo</i> 'tanít'	24	247	0,097
	<i>capio</i> 'fog'	40	412	0,097
	<i>iubeo</i> 'parancsol'	128	1523	0,084
	<i>facio</i> 'csinál'	222	2923	0,076
	<i>reddo</i> 'visszaad'	21	354	0,059
	<i>recipio</i> 'kap'	24	417	0,058
	<i>mitto</i> 'küld'	35	785	0,045
	<i>accipio</i> 'kap'	39	1069	0,036
	<i>pono</i> 'tesz'	28	845	0,033
	<i>scribo</i> 'ír'	25	1000	0,025
	<i>trado</i> 'átad'	4	355	0,011

Átlag 0,451

A lista élet tekintve azonnal szembeötlik, hogy összesen három olyan ige van, amelyeknek többször fordul elő a praeteritum imperfectum, mint a praesens perfectum E3 alakja: a létige és két igekötős alakja (*sum* 'van', *adsum* 'ott van', *absum* 'nincs ott'), de előkelő helyen áll még a *sum* másik két összetétele (*desum* 'hiányzik', *possum* 'képes'). A létige talán tekinthető a *par excellence* statikus jelentésű igének, és a lista első negyedében valóban leginkább statikus jelentésű igéket látunk (*habeo* 'birtokol', *teneo* 'tart', melyek jelentése a létigéével rokon, vagy a lelkiállapotot kifejező *amo* 'szeret', *credo* 'hisz'). A lista végén azonban sokkal jellemzőbbek a cselekvést kifejező, dinamikus jelentésű igék (*trado* 'átad', *pono* 'tesz', *scribo* 'ír', *mitto* 'küld', *reddo* 'visszaad', *facio* 'csinál', *iubeo* 'parancsol') és a dinamikus-mozzanatos, bár kevésbé aktív jelentésűek (*accipio* 'kap', *recipio* 'kap'). Ezeknél, úgy tűnik, sokkal természetesebb a jelöletlen, eredetileg inkább mozzanatosságot kifejező praesens perfectum használata. A lista élén, a statikus jelentésűek közé keveredő *quaero* csak első ránézésre lóg ki a sorból; a *quaerebat* 'kérte, kérdez(get)te' természetes módon fejezi ki az ismétlődő, bizonytalan kimenetelű próbálkozást, tehát a praeteritum imperfectum kvázi-modális használata indokolja ennél az igénél a magasabb arányt.

3. Konklúzió

A fennmaradt latin (nem feliratos) szövegek teljes korpuszán végzett adatgyűjtés egyszerűen, de szemléletesen illusztrálja azt a meglátást, amely a nyelv leírásában valamilyen formában már régen jelen volt, de számszerűsíteni nem volt mód, az igék szemantikai osztályaihoz pedig (tudomásunk szerint) explicit módon nem kötötték. A múlt idő kétféle morfológiai kifejezésmódja közötti különbség határozottan kirajzolódik az igék szemantikai kategóriáinak függvényében. Ez annyit jelent, hogy minél statikusabb egy ige jelentése, annál gyakrabban fordul elő múlt idejű jelentés esetén a praeteritum imperfectummal kifejezett alakban, és minél dinamikusabb, aktívabb, annál ritkábban.



1. ábra. A praet. impf. és a praes. perf. aránya a vizsgált igéknél

Irodalom

- Cser, A. (2015). The nature of phonological conditioning in Latin inflectional allomorphy. *Acta Linguistica Academica* 62 (pp. 1–35).
- Cser A. (2021). Implikációs viszonyok és jelölő funkció a latin igeragozásban: előtanulmány. In Havas F., Horváth K., Hrenek É. & Ladányi M. (szerk.), *A grammatikától a retorikáig: Nyelvészeti tanulmányok C. Vladár Zsuzsa tiszteletére* (pp. 30–36). ELTE BTK Alkalmazott Nyelvészeti és Fonetikai Tanszék.
- Pinkster, H. (2015). *The Oxford Latin Syntax, vol. 1*. Oxford University Press.

KONVENCIONÁLIS, MÉGIS TRANSPARENTS?

CSEERNYÁK HAJNALKA¹, SZABÓ M. HELGA²

¹HUN-REN Nyelvtudományi Kutatóközpont | ²PhD, tudományos munkatárs
csernyak.hajnalka@nytud.hun-ren.hu | szabo.m.helga@gmail.com | DOI: 10.18135/PG70.2024.4

A modalitásból eredő sajátosságok

A jelnyelvek vizsgálata sok tanulságos hozadékkal szolgált a nyelvtudományi kutatások különböző területei számára, elsősorban amiatt, mert az addig széles körben és változatos módszerekkel leírt, nagyszámú hangzó nyelvhez képest a jelnyelvek teljesen más modalitáshoz tartoznak (Zeshan, 2006; Occhino et al., 2017). Az egyik ilyen izgalmas terület a nyelvi jelek konvencionális, illetve transzparens jellege. Lehet-e ugyanaz a nyelvi elem egyszerre önkényes és motivált? Az ikonicitás mellett milyen egyéb módon lehet motivált egy kifejezés vagy nyelvi szerkezet? Ahhoz, hogy ezeket a kérdéseket kicsit körbejárjuk, a magyar jelnyelv (HSL) és az amerikai jelnyelv (ASL) jelkészletének bizonyos szeleteit vetjük össze. Csupán egy kis ízelítőt szeretnénk nyújtani ebből. A cikkben idézett, de fotókkal nem illusztrált jeleket az online jelnyelvi szótárakban lehet megkeresni.*

Történeti mozzanatok a jelkészletek kialakulásában

A jelnyelvi szókészlet gyarapításának – hasonlóan a hangzó nyelvekhez – bevett módja a jelalkotás és a jelkölcönzés. Ugyancsak léteznek nemzetközi vándorjelek is, elsősorban a modern technikai eszközökre és jelenségekre (pl. ROBOT, FACEBOOK), vagy az olyan szakkifejezések körében, amelyek – minimális formai eltérésekkel – napjainkra már több jelnyelvben megtalálhatóak (pl. FONOLÓGIA, COVID). A jelnyelvekben azonban nem kizárólag az új keletű kifejezések lehetnek vándorjelek, hanem egyrészt a népnevek, nemzeti-kulturális jellegzetességek (pl. KÍNAI-ÍRÁS) is. Másrészt sokszor egészen alapvető kifejezések is nagy fokú formai egyezést, akár teljes azonosságot mutatnak (pl. ESZIK, ÁLL, ÜT) – az utóbbi csoportba tartozó jelek egyértelműen közös emberi tapasztalatokon nyugszanak.

Ha a jelek átvételének lehetőségét vizsgáljuk, érdemes lehet megnézni, hogy mennyire igazolható a leszármazás két jelnyelv között. Az ASL és a HSL esetében nyilvánvaló, hogy történetileg teljesen külön utat járt be a két jelnyelv, bár mindkettő az ún. nyugati, európai gyökerű csoportba sorolható, ezért jobban hasonlíthatnak egymásra a szókészletüket és a grammatikai eszköztárukat tekintve is, mint az afrikai vagy az ázsiai jelnyelvek bármelyikéhez.

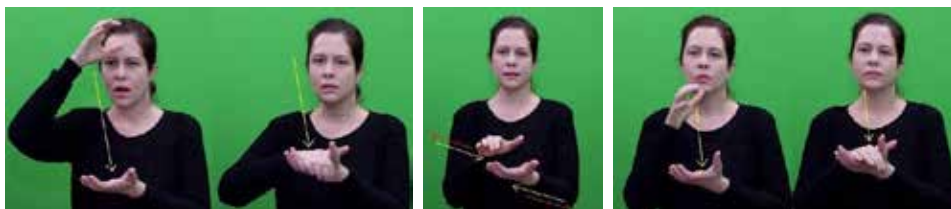
* ASL: <https://www.handspeak.com/word/>; <https://www.lifepoint.com/index.htm>;
HSL: <https://jelnyelvi-szotar.jelnyelv.hu/>; <http://jelesely.hu/szotar/?dictionary>

A világ változását követve és a tömegkommunikációs eszközök fejlődésével párhuzamosan a jelnyelvi kölcsönzések üteme egyre gyorsul. A nemzetközi jelnyelv, amely az 1950-es évektől kezdett tudatos kezdeményezés révén kialakulni, sajátos „lingua franca” szerepet töltött be a siket közösségek nemzetközi kapcsolatrendszerében. Mára ez a fajta kommunikáció egyre inkább az amerikai jelnyelv széles körű használatának irányába tolódott el, ezáltal egyre több amerikai jel kerül át más jelnyelvekre.

A belső jelalkotás logikája: konceptuális sémák a rokonsági terminusokban

Greenberg (1966) „klasszikus” tanulmányának megisméltése során, amelyet Erin Wilkinson (2009) negyven jelnyelv rokonsági terminusaival kapcsolatban végzett, sikerült pár felismerést megfogalmazni a nyelvi struktúra és funkció tipológiai mintáinak összefüggéseiről. Az egy- és többemű rokonsági jelek egy része a leszármazási sorokban vagy épp az azonos generációs szintű jelek esetében gyakran fonológiai átfedést mutatott. Ezzel párhuzamosan megfigyelhető, hogy a rokonságjelek közt gyakori valamilyen konceptuális séma alkalmazása pl. anatómiai jegyek (szakáll, bajusz, mell) vagy kultúraspecifikus (pl. kalap, hidszab, fülbevaló, rúzs stb.), illetve univerzális emberi tapasztalatok (pl. gyerek alacsonyabb testmagassága, kisbaba karban ringatása) alapján.

Ugyanakkor az önkényesség is mutat bizonyos fokú következetességet, pl. a nemekre utalás szisztematikusan kapcsolódhat kivitelezési helyhez: az ASL-ben a férfi rokonsági terminusok a homlok, míg a női terminusok a száj-áll vonalában jönnek létre. Ez még a jelösszetételekre is igaz: a jelösszetételként működő DAUGHTER, SISTER vagy épp a WIFE jelek is mindig az álltól indulnak, majd az utótagot jelentő kétkezes jelbe érkeznek. A SON, BROTHER, illetve a HUSBAND jelek kiindulópontja a homlok és a halánték, utótagjuk viszont ugyanaz, mint a női párjaiké. Hasonlóképpen akár a számjelölés is felbukkanhat a lineáris leszármazást jelölő rokonsági terminusokban, ilyen a MOTHER és a GRANDMA, a FATHER és a GRANDPA jelek közti mozgás ismétlésbeli eltérése (Wilkinson, 2009). Tágabb körben vizsgálódva azt találhatjuk, hogy a mozgás iránya is kifejező lehet, pl. az amerikai jelnyelvi MARRIED, illetve DIVORCED jelek esetében.



1. ábra. HUSBAND, MARRIED / DIVORCED; WIFE (Fotók: Tarr Zoltán)

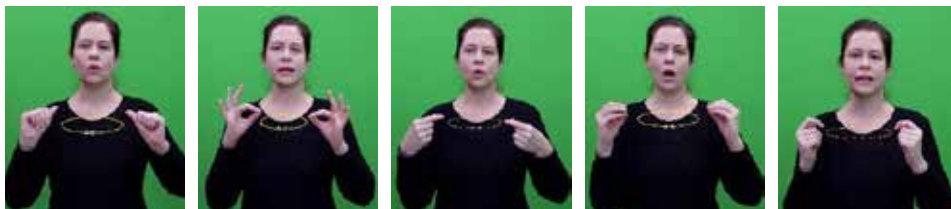
Hasonló elveket, azonban eltérő megvalósítást találhatunk a magyar jelnyelvi rokonságjeleknél. Az ANYA jel (sok más nyugati jelnyelvhez hasonlóan) a szájnál jön létre, s bár erre nincs egyértelmű eredeztetés, ez utalhat a táplálásra vagy a gyermekkel való kommunikációra. Az APA jel a 'bajusz'-ra vezethető vissza (szintén sok más jelnyelvben is).

A NAGYPAPA, NAGYMAMA az IDŐS + APA/ANYA összetételként valósul meg. A NAGYBÁCSI = NAGY + BÁCSI ismét a 'bajusz' analógiájához nyúl, de már egy másik kivitelezésmóddal. Ugyanakkor a 'nagyéni' jelentésű összetétel NÉNI tagjának esetében nem ismert a mögöttes konceptuális séma. A FIA/LÁNYA egyaránt a 'születik/születés' mozzanatot tartalmazza, csupán a mozdulati ív eltérő. Az ÖCS/HÚG összetételként tartalmazza a FIÚ/LÁNY + GYEREK ('kisebb nálam') jeleket, míg az UNOKAÖCS/UNOKAHÚG ezt bővíti még az 'unoka/rokon' (általános mögöttes jelentés: 'hasonló') elemmel.

Hangzó nyelvi hatás: inicializált jelek a jelkészletben

Talán a legkülönlegesebb példa az önkényességre (és egyben a kódkeverésre) az inicializált kézformák felbukkanása a jelnyelvi rokonsági terminusokban – ebben az ASL különösen gazdagnak tekinthető. A motiváltság ugyanis nemcsak ikonikusságból eredhet, hanem nyelvi kontaktusokból is – ezekben az esetekben a jelnyelvi jel minden fonológiai paramétere megmarad (artikulációs hely, mozgás, orientáció), azonban kézformája az azonos jelentésű hangzó nyelvi szó kezdőbetűjét integrálja a fonológiai szerkezetbe (Lepic, 2021; Szabó, 2013). Mivel különféle nyelvspecifikus manuális rendszerek vannak, az inicializált kézformák manuális reprezentációi nem azonosak a különböző jelnyelvekben. A nyugati típusú jelnyelvek többsége a latin ábécén alapuló egykezes daktilábécét használja, a környezeti hangzó nyelvek megfelelő karakterkészlettel, tehát a magyarban létezik /cs/, /gy/, /ly/ vagy éppen /i/, /ö/, /ü/ stb. is.

Az ASL-ben meglehetősen magas az inicializált jelek aránya a szókészlet minden szegmensében. Komplet jelcsaládok („szóbokrok”) épülnek eköré. A FAMILY, TEAM, GROUP, ORGANIZATION, ASSOCIATION (ld. az alábbi fotókon) jelek közös jelentésmagja a 'csoport', 'összetartozás', amit a két kéz kört rajzoló érintkezése fejez ki, miközben a kézformák az aktuális angol megfelelő kezdőbetűje szerint változnak. A fentebb már bemutatott MOTHER/FATHER; WIFE/HUSBAND stb. sorába illeszkednek további elemek is, amelyekben inicializált és konceptuális sémák keverednek: AUNT/UNCLE; NIECE/NEPHEW.

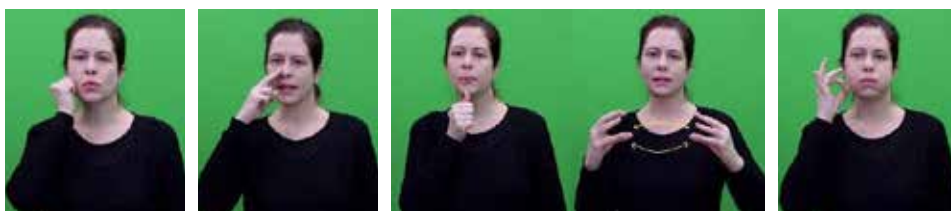


2. ábra. ASSOCIATION, FAMILY, GROUP, ORGANIZATION, TEAM (Fotók: Tarr Zoltán)

A magyar jelnyelvben nemcsak a betűk formáját felidéző ujjábécé használatos, hanem azzal párhuzamosan egy másik segédkód is: az ún. fonomimikai ábécé, amelynek elemei az egyes hangok kiejtését idézik fel. Funkcióját tekintve a kétfajta betűzési mód teljesen egyenértékű, egyéni preferencia – részben korcsoport és a kapcsolódó siketoktatási módszerek – függvénye (Szabó, 2013). Technikailag annyi az eltérés, hogy a fonomimikai elemeknek nemcsak a kézformája, hanem az artikulációs helye is kötött,

hiszen jellemzően a hangképzést hivatottak felidézni (pl. nazális vagy ajakkerekítéses stb. jellemzőre utal), így az adott jelnek nem csupán egy, hanem rögtön két fonológiai komponense is kötött lesz.

A hagyományos daktilalapú inicializált jelek közé sorolható a magyar jelnyelvből a VÁC vagy a WIFI. Ezzel szemben a többi, alább látható példa a fonomimikára épül. A PADLIZSÁN vagy a BÜKK(FA) esetében a [p] vagy a [b] fonomimikai jele egyaránt fel-pattanó ajakkal történő képzésére utal, de eltérő mozzanatot emel ki. A nazális jelleg érzetét idézi fel a MŰ és a NYELVTAN gyerekek által használt változata, ugyancsak más formai megoldással – ebbe a sorba tartozna még az [n] fonéma jelölője is, de erre most nem hoztunk külön példát. Talán ennyiből is látható, hogy még a zömmel hangképzésre utaló kézmozdulatok is önkényes jellegűek, hiába motiváltak.



3. ábra. MŰ, NYELVTAN, PADLIZSÁN, BÜKK (Fotók: Tarr Zoltán)

Összegzés

A fentiek alapján elmondható tehát, hogy egyfelől metaforikus, metonimikus, ikonikus vagy éppen specifikus tulajdonságok által motivált mintákat („image schema”), önkényes jellemzőket találunk. Ugyanakkor feltűnik a környező hangzó (írott) nyelvek részleges befolyása is: vizuális ortografikus reprezentáció, morfoszintaktikai konstrukciók. Ezek az utóbbi megoldások tehát a hangzó nyelvek és a jelnyelvek közötti szociokulturális interakció lehetséges változatai. Mindamellet akármilyen módon motivált is egy jel szerkezete, a fenti példák egyértelműen igazolják a saussure-i alaptételt, miszerint a nyelvi jel két oldalát képező jelöltnék és jelölőnek a kapcsolata minden esetben önkényes – még az olyannyira ikonikusnak látszó jelnyelvek esetében is.

Irodalom

- Greenberg, J. (1966). *Language universals*. The Hague: Mouton de Gruyter.
- Lepic, R. (2021). From letters to families: Initialized signs in American Sign Language. In Boas, H. C., & Höder, S. (Eds.), *Constructions in Contact 2 – Language change, multilingual practices, and additional language acquisition* (pp. 267–305). John Benjamins.
- Occhino, C., Anible, B., Wilkinson, E., & Morford, J. P. (2017). Iconicity is in the eye of the beholder: How language experience affects perceived iconicity. *Gesture, 16*(1) (pp. 100–126).
- Szabó M. H. (2013). A hangzó magyar nyelv hatása a magyar jelnyelvre. In É. Kiss K. (szerk.), *Nyelvelmélet és kon-taktológia I. kötet* (pp. 43–79). Piliscsaba: PPKE.
- Szabó M. H. (2015). *Motiváltság és ikonicitás a magyar jelnyelvből*. In Kádár E. & Szilágyi N. S. (szerk.), *Motiváltság és nyelvi ikonicitás* (pp. 217–234). Kolozsvár: Erdélyi Múzeum-Egyesület.

SÉRINCHALAK ÉS ÚRINÓK: a szótár szórakoztató oldalai

FEJES LÁSZLÓ | HUN-REN Nyelvtudományi Kutatóközpont

fej.es.laszlo@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.5

A Nyelvtudományi Intézetben végzett első munkám egy olyan projekthez kötődik, melyben Prószéky Gábor kulcsszerepet játszott: morfológiai elemzőket fejlesztettünk kisebb uráli nyelvekhez. Bár közvetlenül Novák Attilával dolgoztam, és Prószéky Gábor a projekt menedzselésével foglalkozott, a munkálatok kapcsán többször találkoztunk. Az első projektet két hasonló követte: ezekben hivatalosan Prószéky Gábor nem vett részt, de kívülről támogatta őket. Amikor eltávolítottak a Nyelvtudományi Intézetből, Prószéky Gábor volt az, aki megpróbált nekem pályázati pénzt szerezni – nem rajta múlt, hogy ez nem sikerült. Az Intézetbe viszont véletlenül éppen abban az időben kerültem vissza, amikor ő lett az igazgató.

A Nyelvtudományi Intézetben eltávolításom idején többek között éppen Munkácsi Bernát manysi szövegeit igyekeztünk elemezni, és tőtárként Munkácsinak a Kálmán Béla által kiadott *Wogulisches Wörterbuchját* (a továbbiakban MK) alkalmazni. A feladat nem volt egyszerű, éppen ezért hónapokat töltöttem a szótár anyagával. Bár a munka gyakran igen monoton volt, számos vidám pillanatot okozott. Úgy gondolom, ezek a pillanatok, felfedezések, melyeket alább megosztok, az ünnepest és az olvasók örömére is szolgálnak. Fontos azonban leszögezni, hogy bár ezek az észrevételek olyan jelenségekre vonatkoznak, amelyek a szótár hibáinak tekinthetők, ezek egyfelől mennyiségükben is eltörpülnek, másrészt olyan problémákat jeleznek, melyeknek nincs egyértelműen jó megoldásuk. Azt a munkát, amelyet Munkácsi Bernát az anyag gyűjtésével, Kálmán Béla pedig a feldolgozásával elvégzett, semmiképpen nem ezek a példák jellemzik. Az alábbi gyűjtemény pusztán célja a szórakoztatás. Bár néhány példa kétségkívül jó alkalmat nyújtana nyelvészeti vagy lexikográfiai problémák fejtegetésére, itt és most ilyen célom nincs.

Az elemző fejlesztése során a fő nehézséget az jelentette, hogy a MK nem a Munkácsi által kiadott szövegekre, hanem Munkácsi külön szógyűjtésére, céduláira épül. Munkácsi átírása amúgy sem teljesen következetes: ugyanazt a magánhangzót hol *a*, hol *q* (a szótárban *ǻ*) jelöli; egy másikat hol *a*, hol *ǻ*; megint másikat hol *o*, hol *ǻ*; vagy éppen hol *e*, hol *ǻ* jelöli. Éppen ezért, amennyiben a szó megvolt Kálmán saját feljegyzésében is, szögletes zárójelben megadta az alakot saját átírásában is. A szögletes zárójel miatt az az olvasó benyomása, hogy ez a kiejtést mutatja, és tulajdonképpen ez is a funkciója. Éppen ezért meglepő, amikor véletlenül mégsem jelölheti a kiejtést: pl. a *put* 'jég fölötti víz', 'kiejtése' [*put-wit*] (a *wit* 'víz', tehát nyilván arról van szó, hogy a szót már csak a 'víz'-zel alkotott összetételben tudta Kálmán feljegyezni).

A hangjelölés következetlenségére fel voltunk készülve, az azonban váratlan meglepetés volt, hogy mennyire következetlen a kötőjelek használata. Ugyanannak a folyószakasznak a neve különbözőképpen szerepel a nevében szereplő összetételi tagok szócikkeiben: az *alík*-nál *alík-tāl-táltém mēsĵʷ* (MK 31b), a *mēsĵʷ*-nél *alík-tāl-táltém-mēsĵʷ* (MK 306a), a *tāl*-nál *alík tāl tāltém mēsĵʷ* (MK 621a). Mivel a morfológiai elemző a szóközök közötti karaktersorokat elemzi, ezek az alakok a szövegben mind más-más elemzést adnak, a tőtárba pedig egészen másképp kell őket felvenni.

Igazi érdekességgel azonban nem az alakok, hanem a megadott jelentések szolgálnak. Ezek közül külön is figyelemre méltóak az állatnevek. A *tarkā* megadott jelentése 'sérinchal' (MK 631b). Netes keresés és az Arcanum folyóiratkorpusza alapján ez a halfajta-megnevezés kizárólag finnugor nyelvészeti tanulmányokban fordul elő; a megadott orosz és német megfelelők alapján (*vágó*) *durbincs* lenne a ma is használatos kifejezés. A más nyelvi megfelelők azonban nem mindig segítenek, akár ellentmondásosak is lehetnek. A *lāḥṣarj* esetében például egy sor orosz megfelelő szerepel: *пескан ~ пискан, пескозоб ~ ескозоп ~ мескозоб ~ безкозоб ~ нескозоб ~ воскозоб*. Ezek népi elnevezések, a hal hivatalos elnevezése *пескарь*, azaz magyarul *fenékjáró küllő*, latin tudományos elnevezése *Gobio gobio*. Ezzel szemben a megadott latin elnevezés (*Cyprinus phoxinus*) alapján a hal hivatalos orosz neve *гольян*: ez egy *Phoxinus* halnemzetség elnevezése, melynek magyar megfelelőjét nem találtam, de tipikus képviselőjét, a *Phoxinus phoxinust* *fürge cselle* névvel illetjük. Hogy melyik halról van szó, az a szótár alapján aligha deríthető ki. Hasonlóképpen az *āutiṅ-palm* magyar megfelelője *lódarázs*, oroszul *шершень*, németül pedig *Pferdebremse* (MK 399a). A magyar *lódarázs* orosz megfelelője valóban a *шершень*, de németül *Hornisse* lenne. A *Pferdebremse* a magyar *lódarázs* tükörfordításának tűnik, kérdéses, van-e hivatalos magyar megfelelője; a *szudétabögöly* a *Tabanus sudeticus* tudományos név szó szerinti fordításának tűnik, szakszövegekben nem találtam nyomát. Valójában azt sem tudjuk, melyik (melyek) lehetett(ek) a Munkácsi által lejegyzett alak(ok), így a beazonosítás nehéz. Persze ezekben az esetekben az sem feltétlenül igaz, hogy fajnevekről van szó, és nem különböző fajokra alkalmazott megnevezésekről.

A magyar és a német közötti megfelelések máshol is lehetnek meglepőek. Az *iramodik* német megfelelője rendszeresen *rennen*. Csakhogy míg a német szó a mozgás folyamatára utal, a magyar szó a mozgás megkezdésére. Felmerülhet a lehetőség, hogy korábban nem volt ilyen különbség. A Czuczor–Fogarasi például említi, hogy az *iramodik* szélesebb értelemben jelentheti azt, hogy 'sebesen fut', de ott sem ez az elsődleges értelme (és hangsúlyozza a valaminek az irányában vagy valami elől történő mozgás jelentéselemét). De ha meg is engedjük, hogy a MK egy másik kor (vagy Munkácsi személyes) szóhasználatát tükrözi, akkor is kérdéses, hogy mennyire jó ez a megoldás egy nyelv dokumentálásakor. Hasonló eset lehet, *sānim-ēkwā* magyar megfelelőjeként az *anyám-asszony* szerepel (MK 588a: németül *Frau Mutter*), holott ez manapság már legfeljebb az *anyámasszony katonája* kifejezésben használatos. Bár Munkácsi idejében ez másképp lehetett, 1986-ban már furán hangozhatott.

Megmagyarázhatatlan módon a manysi samovar magyar megfelelőjeként a *szamovár* szerepel, német megfelelőjeként viszont a *Teemaschine* (MK 525a), holott a *Samowar* a németben is közismert szó. Tévedésnek tűnik azonban a *pērés* ~ *pērs* jelentésmegadása: magyar és német megfelelőjeként egyaránt a *paprika* (*Paprika*) szerepel (431b), holott ez egyértelműen az orosz *перец* átvétele (az orosz eredetre a szótár is utal, bár nem nevezi meg az átvett szót) – az pedig ugyan jelenthet paprikát, de kultúrtörténeti alapon valószínűbbnek tűnik, hogy itt 'bors'. Az orosz jövevényszavak átvétele alaki szempontból a fenti idézeteknél sokkal izgalmasabb is lehet, jó példa erre a *šēhēr* 'mémők' (MK 592a), melynek eredetét a német megfelelő fedi fel: *Ingenieur*.

Nem csak a német jelentésmegadás okozhat nehézségeket, hanem az orosz is. A *rēpés* szócikkében a következőt találjuk: *рябой, мягкий свол, пеганная, березовая своль* (átrepül egyik nyírfáról a másikra) (MK 499a). A *свол* és a *своль* azt a gyanút kelti, hogy ugyanaz a szó szerepel kétszer, egyik esetben sajtóhibával: a különböző nemből álló jelzők viszont azt sugallják, hogy valóban két különböző szó áll ott; ezeket azonban szótárak és anyanyelvi beszélők segítségével sem sikerült azonosítani. Felmerülhet az a magyarázat, hogy az első esetben a *свол*, a másodikban a *смоль* állna helyesen, és a két alak egymásra hatva torzult a feljegyzésben. Eszerint az orosz meghatározás 'foltos, lágy fatörzs, ragadós nyírfaszurok' lenne, de erről nehéz elképzelni, hogy egyik fáról a másikra repüljön. Más nyelvjárásokban a szó jelentése *репка*, *шваль* (на берёзы). A *репка*-nak három lehetséges jelentését sikerült azonosítani: 'répácska', 'sodronyvég' (tengerészeti szakkifejezés), illetve a szlengben 'repetitornő'; a *шваль* pedig 'kacat, értéktelen dolog'. Egyik sem fordul elő jellemzően nyírfán (на берёзы). Egy példamondat szerint késmarkolatot, edényt csinálnak belőle, tehát akármi is az, szinte biztosra vehetjük, hogy a nyírfa valamelyik része, az azonban nem világos, hogyan repked.

A magyar kifejezések is rejtélyesek, a *sirej-āpēr* jelentéseként 'kardvédelmi ügyesség' van feltüntetve, és csak a megadott német jelentés ('Gewandtheit im Fechten') alapján válik világossá, hogy valójában 'vívásban való jártasság'-ról van szó (MK 551a).

A *paša ölén* üdvözlés német fordítása *Grüß Gott!* (MK 419b), ami erős osztrák(-bajor) színezetet ad (bár a magyar *adjon Isten!* is meglehetősen szabad fordítása). A *sir-tēp-ismit* (137a) és a *sir-tēp ismit* (550a) ismét csak kötőjelezésben térnek el egymástól, német megfelelőjüként egyaránt a *Graupensuppe* szerepel, ám magyarul az előző esetben *kásaleves*, az utóbbiban *daraleves*. Ez esetben nyilván nincs tévedés, de a különbség az adatok következtelen kezelésére utal.

A magyar és a német jelentés különös zavarokhoz is vezethet. A *hulmēr* *χār* jelentéseként hol 'sebes csónak', 'schneller Kahn' (MK 80a), hol 'lyukas csónak', 'ein lecker Kahn' (MK 373b) van megadva. Kétségtelen, hogy a *lyukas* és a *gyors* is lehet egyaránt *sebes*, csak egy kicsit másként... Mivel a képzés alapja a *hulēm*, *hulmi* 'seb', 'Wunde', kétségkívül az utóbbi jelentésmegadás a helyes.

Másképp furcsa a *pēniη-kēr* jelentésmegadása: *puska* (bűvös vas); *sörét* (titkos szó) – németül *Flinte* (bezaubertes Eisen); *Schrot* (Geheimwort) – (MK 430b). A *puska* nyil-

ván a valódi megfelelő, a *bűvös vas* pedig a morfémaról morféma-ra való fordítás; a *sörét* valódi megfelelő, a *titkos* szó viszont valamilyen szóhasználati korlátozásra utal (tabuszó?). Az azonban nem világos, hogy milyen kommunikációs helyzet az, amikor a sörétet puskának kell nevezni (vagy fordítva). Ezzel azonban elérkeztünk egy újabb örömforráshoz, a manysi polyszémia felfedezéséhez.

Számos szónak van ugyanis több olyan jelentése, melyek nehezen hozhatóak közvetlen összefüggésbe egymással, ugyanakkor mégis érdekes asszociációkat kelthetnek. A *loski* például egyszerre lehet 'üt' és 'hazudik' (MK 268). Némi kétséget ébreszt azonban, hogy szerepel a *rétél loski* 'hazudik' (szó szerint kb. 'csalással üt') kifejezés is, ami azt a gyanút kelti, hogy a *loski* esetleg csak a teljes kifejezéssel együtt használatos 'hazudik' jelentésben (bár nem zárhatjuk ki, hogy jelentéstapadás eredményeképpen önállóan is).

Más esetekben az összetartozó jelentések összefüggése nyilvánvaló, bár nem kézenfekvő, hogy egy szó a megadott jelentéseket mind felvegye: *sis* 'anya', 'csecs, emlő', 'magzat, gyermek, ifjú', 'kedves' (MK 596a). Hasonló jelenség, hogy *tān* 'in', 'ér', 'incérna', 'húr' és 'zene' jelentésben is előfordul (MK 624b–625a). Ide vonható az a jelenség is, hogy a *šaxl-törém* (melynek szó szerinti jelentése 'viharisten, mennydörgésisten') megadott jelentései 'mitológiai lény, Numi Törém öccse', illetve 'Illés próféta': itt nyilván a két különböző hiedelemvilágban szereplő alak azonosításáról van szó. A jelentésátvitel szellemes példáját nyújtja a *tūr*, melynek eredeti jelentése 'tó', de adatolt 'fehér kocka a sakktablán' (szabatosabban: 'világos mező') jelentésben is. A többértelműség kevésbé érdekes esete, amikor egy köznévi tulajdonnévi elemként is felbukkan, ám ezek között is találhatunk némiképp meglepőeket is: pl. a *vošiv* 'hímvevő' folyórész neveként is felbukkan (MK 739b).

Helyenként már a felsorolt jelentések mennyisége is meglepő: *nak* 'iz, tag', 'réteg', 'fok, lépcső', 'rekesz, szoba', 'vég', 'ok, végzet', 'bűn', 'baj, balszerencse' (MK 324b). Ezt fokozza, amikor a jelentéssorok végén ellentétes értelmű jelentések állnak: *lakwi* 'érkezik, lép, lopózkodik, mozog, fölrak, ugrik, nyomul, bújik, jut, férközik, kúszik, hordoz, tolódik, távozik' (MK 247b); *sujti* 'érződik, hangzik, zúg, hallatszik, kiált' (MK 570b). Az ellentétes jelentések látszólag össze nem függő jelentések mellett is előfordulhatnak: *tēiv* 'eszik', '(tűz) ég', 'közösül', 'fagyaszt' (MK 636b). Arra is van példa, amikor a két ellentétesnek mondható jelentés mellett más jelentés nem adatolt: *pošuli* 'szarik, fosik', illetve '(durva) eszik, befal' (MK 466b), *pilišmä* 'félelmetes, rettenetes', 'félős, ijedős' (a *pili* 'fél' származéka – MK 440b). A fenti esetek többségében feltételezhetjük, hogy a szöveggörnyezet alapján azonosítható az aktuális jelentés, a *tēliŋ* 'húsos, halas, darás, fűszeres' esetében (MK 640a) azonban erősen kétséges – a közelebbi példák alapján gyanús, hogy inkább 'sűrű, tápláló' lehet a valódi jelentés (a morfémaról morféma-ra való fordítása is leginkább 'tartalmas'). Szintén furcsa az *uj* 'állat, madár, hód, jávor, medve, rénszarvas' jelentésmegadás (MK 688b): mivel az 'állat' az összes többit is lefedi, inkább arról lehet szó, hogy számos szöveggörnyezetekben világos, hogy éppen milyen állatról van szó, illetve tabuisztikus okokból gyakrabban

használják az általánosabb kifejezést, mint pl. a magyarban. Ez azonban nem egyenlő azzal, hogy a szó ennyiféle állatot jelent.

Az is a meglepetés erejével hathat, amikor olyan rokonsági viszonyokat jelölnek egyetlen szóval, melyeket mi egészen különböző viszonyoknak tartunk: *pānt* 'sógor, nagybácsi (1. a feleség öccse és húga így nevezi a férjet; 2. az idősebb nővér férje; 3. a férj öccsét hívja így a feleség)' (MK 412b). Az 1. és a 2. jelentés közötti különbség azonban itt homályos, hiszen a feleség öccse és húga számára a feleség nővér. Az, hogy az azonosnak tűnő 1. és 2., illetve a 3. jelentés egyaránt jelen van, annyira nem meglepő, hiszen a magyar *sógor* is egyaránt jelentheti a testvér házastársát és a házastárs testvérét, legfeljebb nem utal a korkülönbségre. Az viszont feltűnő, hogy míg a 'nagybácsi' jelentés is fel van tüntetve, a részletezett jelentésleírás egyik pontja sem fedi a magyar *nagybácsi* (azaz a német *Onkel*) fogalmát.

Időnként a jelentésviszonyok nem világosak. A *pil* két szócikket is kap, az egyikben a 'bogyó' (MK 439a), a másikban a 'pinácska' (MK 440a) jelentésben szerepel. A két külön szócikk arra utal, hogy külön etimonokról van szó, azaz a két jelentés nem függ össze, a szavak hangalakja véletlenül esik egybe. Az előbbi szócikken belül szerepel a *pil-saut* összetétel (MK 439b – a *saut* önállóan 'nyírhéjkosár', illetve 'a medve hasa', vö. MK 567b) 'bogyószedő nyírhéjedény', illetve 'lányka szeméremteste (szórtelen)' (MK 439b). Ez a két jelentés nyilvánvalóan a *pil* két különböző jelentésével függ össze, így ha az összetételt egy etimonhoz tartozónak tekintjük, akkor az előtagot is azonos eredetűnek kell tekintenünk. A másik lehetőség, hogy az összetétel két jelentését is véletlen egybeeséssel magyarázzuk, és külön szócikkben tüntetjük fel őket.

A poliszémia különös eseteként említhetjük, amikor az alapszó és a belőle képzett szó között találunk a képzővel nem magyarázható eltérést. A hangutánzó jellegű *vāu* a kutyaugatást imitálja (MK 59a), ám a belőle képzett *vāuti*, *vāutaχti* (MK 744b) 'böng' igékkel a rénszarvas hangját jelölik. A szótár az utóbbi igét ugyan a *vōwi* 'hív' igéből származtatja, de a szemantikai hasonlóság inkább a hangutánzószóból való képzést támogatja (vö. magyar *jaj-gat*, *u-gat*, *bé-get* stb.), sőt az sem kizárt, hogy a 'hív' jelentésű ige is hangutánzó eredetű.

Előfordul, hogy a véletlen hoz vicces megfeleléseket a forrásnyelvi és a célnyelvi alakok közt. Az *aki-jiw ur* jelentése 'öserdő' (szó szerint 'apó-fa erdő' (vö. MK 27b), azaz 'nagyon öreg fákból álló erdő'), németül tehát *Urwald* (MK 160b). Ebben az esetben nem is az azonos jelentésű elemek hasonló hangzásúak. Ezzel szemben az *uri-nē* 'szabályokat megtartó nő' (MK 702a) esetében az azonos eredetű 'nő' elemek hangzása esik egybe, a hasonló értelemben is használt *úrinō* esetében az előtag egybeesése a véletlen műve (az előtag itt nem a 'hegy, erdő' jelentésű, hanem az 'oldal, szöglet, vidék, irány', 'ok', 'alak, mód, szokás', '-szor' jelentésű *ur*-ral függ össze, vö. MK 701b–702a). Arra pedig, hogy az összecsengésben mindkét esetben az *ur* hangsor vesz részt, legfeljebb annyit mondhatunk, hogy az *ur* útjai kifürkészhetetlenek...

Végezetül érdemes kiemelni néhány szép példányt a manysi szóalkotások és kifejezések közül: *lū-lāj'-pattā-kēr* 'lő-láb-alj-vas', azaz 'patkó' (MK 209b), *sāt nelmép ujris*

'száznyelvű állatka', azaz 'kanári' (MK 360b, 535b, 691a), *háwél-tēné-χā́tél* 'húsevő nap', azaz 'karácsony' (MK 114a, 371b), *janés tēn-utné pati* 'saját ételbe fogott' azaz 'megjött a havibaja' (manysi szokás szerint a nők a menstruáció idejére elvonulnak – MK 638a), *mańsi lātχiη takkät jālné χum* 'a vogulul beszélő magányos vándor', azaz Reguly (MK 620b).

Remélem, a fenti lista hozzájárul ahhoz, hogy a manysi nyelv örömeit ne csak magányos vándorok élvezhessék. (Tekintettel arra, hogy a fenti írás elsődleges célja a szórakoztatás, illetve az említett források komoly érdeklődők számára amúgy is azonosíthatóak, ezúttal irodalomjegyzéket nem adok.)

A NYELVTÖRTÉNÉS ÉS A KORPUSZ: diakrón a-Pró-ságok

GUGÁN KATALIN, VARGA MÓNIKA | HUN-REN Nyelvtudományi Kutatóközpont
gugan.katalin@nytud.hun-ren.hu | varga.monika@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.6

*„nem akartam oly paraszt lenni, hogy
[...] legalább szíves köszöntésemmel
ne aggratuláljak.”*

(Károlyi Sándor, 1712)

A nyelvtörténeti vizsgálatok mindig is korpuszokon alapultak, épp csak a *korpusz* szó jelentése változott meg az utóbbi két évtizedben. Amikor a szerzőpáros idősebbi tagja elkezdett az intézetben dolgozni, a feldolgozandó korpuszt még egy vaskos (papír) mappában kapta meg: egy kötegnyi fénymásolatot, melyekben különböző századokból különböző szövegtípusokat képviselő, de azonos hosszúságú szövegrészek voltak. Amikor a fiatalabbik szerző állt be a csatsorba, már zajlott az elektronikus korpuszok építése.

A nyelvtörténeti kutatás és a korpusznyelvészet találkozásával kapcsolatban már viszonylag korán érzékelhető volt némi szkepszis. Az egyik lehetséges probléma „a filológus dilemmája” (Matti Rissanen, 1989): az a veszély, hogy a korpuszok, valamint a keresőfelületek segítségével végzett kvantitatív kutatások elveszik a diákok kedvét attól, hogy megismerjék magukat a régi szövegeket, és így alapos tudáshoz jussanak egy-egy régebbi nyelvállapotról, ez pedig végső soron kedvezőtlenül befolyásolja a diakrón kutatásokat.

Messze vezetne, és nem is illene ehhez az íráshoz, hogy ezt a baljós előrejelzést elemezzük. A helyzetünk is speciális, hiszen a történeti korpuszok építésekor nagyrészt manuálisan normalizáljuk a szövegeket, ennek során tehát éppenséggel van lehetőségünk ismerkedni a szövegekkel. A normalizálás maga viszont meglehetősen időigényes folyamat. Ennek egy szemléletes példája az, amikor az átíró, majd a javító szeme elé kerültek Wathay Ferenc önéletírásának a következő részletei: *„es minket uiuek az teöb Rabiához az Bassanak az Nagi Tömlöczben mel Vagion az Vizuari Kurua zariaban alat”,* valamint: *„Kýuiuen azertt az kegielmes lsten, az Derek kurua zaria kapuianis”.* Jóllehet sokszor nyugtattuk magunkat azzal, hogy nem szükséges a történeti szöveg minden pontját teljes mélységében megértenünk (a normalizáláshoz nem is tudnánk szómagyarázatot fűzni), a „kurva zárja” kérdésével már csak az egybeírás-különírás problematikája miatt is kezdenünk kellett valamit. Ehhez azonban azt is meg kellett előzetesen fontolni, milyen formára keressünk a szótárakban (a keresőprogramokat addig inkább elkerültük, érthető okokból). Több szövegkiadást és a – szerencsés mó-

don ezúttal hozzáférhető – kéziratot ellenőrizve kiderült, a szó első szótagjában valójában nem *u*, hanem *a* szerepel, tehát *kar*- kezdetű elemekre kellett keresni. A hazai szótárakban továbbra sem volt adat, ezzel együtt sikerült kinyomozni, hogy a *karvazárja* 'caravansérail', azaz 'karavánhotel' a keresett kifejezés.¹ De hogy mitől *feslett jó ló* az ugyanitt emlegetett háticas, azt a mai napig nem sikerült megfejtenünk.

A normalizálás hol csak mechanikus, hol a fentihez hasonló nyomozómunkát igénylő feladata tehát eléggé rabolja az időt, szépségflastroma viszont a közben felszedhető szövegismeret, melynek eredményeként kialakulhat egy általánosabb kép is az adott korszak nyelvhasználóiról. Ez persze meg sem közelíti azt a tudást, amellyel az adott korszak beszélője rendelkezett, viszont sokszor ad egyféle rácsodálkozásélményt, amikor vagy új jelenségekre bukkanunk rá, vagy pedig olyanokra, amelyekről ugyan tudunk, de mégis más ezeket a tényleges környezetükben, az egykori beszélőktől (lelki füleinkkel) hallani. A bármilyen (akár abszolút nem nyelvészeti) okokból érdekesnek tűnő adatokból mindkettőnknek van egy-egy nagyobb gyűjteménye, ezekből válogatunk most össze egy csokrot e jeles alkalomra. Az adatok zöme a *Történeti magánéleti korpusz* és a *Középmagyar emlékirat- és drámakorpusz* normalizálása közben került eléink,² egy kisebb hányada pedig ómagyar kori szövegek bogarászása során.

Gyűjteményünk egy része olyan példákban áll, melyek a jelentésváltozás miatt mai szemmel már meghökkenítőek vagy mulatságosak. Ezek persze nyelvtörténeti szempontból nem jelentenek újdonságot, de hát nem jobb úgy sétálni egy szép tavaszi reggelen, hogy az ember tudja, az égről egy *ható állat*, egy *serény állat* ontja rá sugarait?³ Ha pedig éppen szomorkodunk, sajnálhatjuk magunkat az Érsekújvári-kódex szavaival: „*ő, én nyavalyás fejemnek és búban zabált lelkemnek*”.⁴ A jelentésváltozások egész kincsestára a Döbrentei-kódex szeretethimnusza, illusztrációként ebből is elég talán egy mondat: „*Mikoron apród volnék, szólok vala, miként apród. Mi időben kedig hőssé lettem, meghézagítottam, melyek apródságé valának*.”⁵

S hiába hézagítottuk már mi is rég meg az apródsággal járó tulajdonságainkat, azért csak elvigyorodik az ember, ha azt a fenyegető szózatot olvassa, amit Nabukodonozor hallott az égből (a Bécsi-kódex fordításában): „*barmokkal és vadakkal leszen te lakodalma*”.⁶ De ez éppúgy nem kiált föliáért, ahogy az a tanúvallomás sem pszichiáterért,

¹ Forrása: Suzanne Kakuk 1973. *Recherches sur l'histoire de la langue osmanlie des XVIe et XVIIe siècles. Les éléments osmanlis de la langue hongroise.* De Gruyter Mouton.

² Az innen idézendő adatok esetében az egyszerűség kedvéért elhagyjuk a forrásmegjelölést.

³ „*Mert ha iol meglattóc e napba e neget lelyöc / Mert o gengesegos hato allat / Igön sereñ allat / kent o nem zenvethet es igön fenös allat*” (Nagyszombati kódex 223, http://omagyarkorpusz.nytud.hu/documents/kodexek/NagyszK/3a/NagyszK_3a.txt)

⁴ „*Oh een nyavalas fejemnek Es bwban zabalt lelkemnek*” (Érsekújvári kódex 256r, http://omagyarkorpusz.nytud.hu/documents/kodexek/ErsK/3a/ErsK_3a.txt)

⁵ „*Mikoron aprod volnek zolok vala mikent aprod Mi idoben kedeg hosse <loptt.> lopttem : meg hevzageitottam mellék aprod sage valanak*” (Döbrentei-kódex 149r, http://omagyarkorpusz.nytud.hu/documents/kodexek/DobrK/3a/DobrK_3a.txt)

⁶ „*7 barmockal 7 vadackal lezen te lakodalma*” (Bécsi kódex 139, http://omagyarkorpusz.nytud.hu/documents/kodexek/BecsiK/3a/BecsiK_3a.txt)

hogy „*ottan azon halottas háznál a többi jelen levő ifjakkal és leányzószemélyekkel amidőn már alkalmas ideig mulatott volna*”. Egy kis szótárazással persze fényt lehet deríteni arra, milyen is az *által ember*, mi az *egy bokor vidra kesztyű*, és világossá válik, az sem egy hitében megfogyatkozott villanyszerelő szájából hangzott el egykoron, hogy „*ha oly erős isten, vezeték nélkül nem szűkölködnék*”.

Az igekötős igék jelentésváltozásaiból külön bokrétát lehetne kötni. Amikor „*a tolvajok feltámadásával is mindenfelől ijesztenek*”, az nem a túlvilág egy bizarr értelmezése, és sajnos Ulomán bég annak ellenére, hogy *megettért*, sőt, épp ezzel összefüggésben *elrabortatta* – hogy mit is? – egész Kemenesalját. Jó eséllyel a környéket is *megfosztogatták*, sajnos ez elég gyakran megesett. Várakat egyaránt lehetett *kivívni* és *megvenni*, és azok, akik ezt tették, a várat *vitatták* (vagy talán *vitatták?*). De ha már vita, a hitvitázó drámák szókinccse is megér egy misét, mert a jól megfogalmazott érvekkel bizony *kétségbe ejthető* a partner annyira, hogy *elhitetik* az adott egyház *mellől*. Ehhez persze ajánlatos a megfelelő vitapartnereket *feltalálni*, a *meggyőzést* azonban jó elkerülni, legalábbis az 1720-as években egy bizonyos Simon János „*addig meg nem győzhette Németh Jánost, miglen a vörös kővel homlok [sic!] nem csapta*”. Ezzel szemben az a hányatott sorsú drámai hősnő, aki *színében megveszett*, csak sápatag, de komolyabb baj nem érte, és ha a pénz nem boldogít is, átverni lehetett azzal balekokat, hogy pénz áásával *meboldogítják* őket. Nem csoda, ha a kor embere időnként *összekáromlotta magát!*

Hogy pénzzel „*ki lehet váltani a Heislemél zálogban levő feleségem aranyműveit is*”, az is valószínűsíthető, mint ahogy az is, hogy van még bőven vizsgálandó kérdés a szintaxis területén. Persze előfordul, hogy a furcsaság csak első ránézésre az. Például a találati lista alapján azért felszalad az ember szemöldöke, amikor azt olvassa a már korábban is idézett Károlyi Sándor tollából, hogy „*Az Aradival szaporodhatnék, de Isten segítségéből talán meggátolhatom azt is, nagy vigyázásom lévén oda is*”. Rossz az, aki rosszra gondol, de szerencsére a korpusz egy kattintásra kiadja a kontextust: „*Útjában Szolnokot akarja próbálni, de nem hiszem, holnap elvállik, a Dunán bocsátott le hidat, a Tiszán hozza fel, s Szegednél köti által. Nem méltó ettől annyira félni, mert csak mászik, könnyű hada pedig kevés és semmirevaló. Az Aradival szaporodhatnék, de Isten segítségéből talán meggátolhatom azt is, nagy vigyázásom lévén oda is.*” Ennek alapján persze lenne itt is mit vizsgálni, de szerencsére nem a 18. század züllöttségét.

Kissé más vizekre evezve szólásokkal, közmondásokkal, életbölcseiségekkel is gazdagodtunk az idők során, ezekből is szeretnénk megosztani néhányat, próbálva cáfolni azt a – sajnos mégiscsak realizmusra valló – bölcsességet, hogy: „*Jó tett helyében, amint szokták mondani, soha jót ne várj.*” Bár kissé bizarr, nehéz vitatkozni azzal, hogy *ha meghalsz, nem kell akkor több sajt* – és ehhez még szótárakat sem kell bújni. Egyszerű netes böngészés segítségével is sikerült kideríteni, hogy a *kigyóhagymának ebkapor a társa* valami olyasmit tesz, hogy *egyik kutya, másik eb*, csak ennél is negatívabb. Arról viszont még csak ötleteink vannak, milyen az, aki *májusban is babszem az anyjában*.

Sok derűtséget szerzett nekünk az Actio Curiosa szerzője azzal, hogy egész sorozatnyi igazán változatos hasonlattal tette élénkké drámája szövegét:

Isten úgy segéljen, a te devóciód csak olyan, mint mikor a gólya a toronyban megyen, oldalba fossa, s otthagya.

S hát bizony csak oly titok volt az, mint akinek szép a hordója, büdös a bora, vagy mint aki a mezőben tojik, a faluban költi ki.

S hát jóakaród? Te a szivedben, s ő meg a fenekedben. Hej, nem tudod-e, hogy csak úgy van a te dolgod, mint mikor a disznó a szentegyházhoz dörgölődik. Hallottad-e, hogy nem illeti az ebet hintőszekér, nem a te rendedhez való az.

S ha már disznó, ugyan nem a fenti csoporthoz tartozik, de mennyire szellemes bók már az, hogy „[k]egyelmes uram, hagyom mind e világra, hogy száz disznó között is egy ember volnál te”? Bizony ilyeneket mond az, aki *csapja a csikot a lencsével*. És hogy egy cáfolhatatlan bölcsességgel zárjuk a szólások és egyebek sorozatát: „*Tudod, aki mit eszik, iszik is reá*.”

A szövegek normalizálását a morfológiai elemzés követi. A TMK esetében az elemzéseket kézzel egyértelműsítettük; mindenféle kommentár nélkül álljon itt néhány olyan gyöngyszem, amely segített minket ráébredni arra, hogy mennyivel több két- és többbértelmű szóalak van, mint azt valaha is gondoltuk volna.

Hogy a nádfalat finomabb-e, mint a fütök, azt nem fogjuk kísérleti módszerekkel megvizsgálni. Maradunk a fél-fél halvánál, böjtös napokon a haltálnál, és a fenti sor folytatása helyett azzal szeretnénk lezárni ezt a rövid írást, hogy mennyi egyéb meg-

hog	fonaladat	megh	nem	szöttem,
hog	fonaladat	meg	nem	szöttem!”
hog	fonal+adat	meg	nem	sző
C	N	VPfx	Adv	V.Past.S1.Def

csak	megindulok		Kassára	a	jövő	héten.
csak	megindulok		Kassára	a	jövő	héten.
csak[Adv]	meg[+indul][VPfx.V.S1]		Kassa[N.Sub]	a[Det]	jövő[Adj]	hét[N.Sup]
Nem	is	mulathatnék	it,			
Nem	is	múlathatnék	itt,			
nem[Adv]	is[Clit_is]	múlathatnék[N]	itt[Adv Pro]			
nincs	egy	hordó	boromnál	töb.		
nincs	egy	hordó	boromnál	több.		
nincs[V.S3]	egy[Q]	hordó[N]	bor[N.PxS1.Ade]	sok[Q.Cmp]		

mint egy <mintegy mint+egy[Adv]	fél halva félhalva fél+halva[N]	ott hagyván otthagyván,> ott]+hagy[VPfx.V.PartAdv=vÁN]
---------------------------------------	---------------------------------------	--

No „No, no[Inj/Utt]	fujtsz fűtesz?” fűt[V.S2]
---------------------------	---------------------------------

arra Arra az[N]Pro.Sub]	felele felele’ felel[V.Ipf.S3.Def?]	néki neki ő[N]Pro.Dat.S3]	a a a[Det]	Lakó lakó: lakó[N]
-------------------------------	---	---------------------------------	------------------	--------------------------

fűttök. „Fűtök.” fű+tök[N]	szörnyü szörnyű szörnyű[Adj]	halállal halállal halál[N.Ins]	haltál haltál hal+tál[N]	volna, volna, van[V.Cond]
----------------------------------	------------------------------------	--------------------------------------	--------------------------------	---------------------------------

a a a[Det]	Nád falat nádfalat nád+falat[N]	két felé kétfelé két+felé[Adv]	nyitván nyitván, nyit[V.PartAdv=vÁN]	által szollalt, általszóalt,> által]+szóal[VPfx.V.Past.S3]
------------------	---------------------------------------	--------------------------------------	--	--

az a a[Det]	hatalmas hatalmas hatalmas[Adj]	úr Isten Úristen úr+isten[N]	legyen legyen van[V.Subj.S3]	veszértek vezérek vezér[N.PxP2]	és és és[C]	igazgatótok igazgató+tok[N]	minden minden minden[Det]	ügyetekbe, ügyetekbe’, ügy[N.PxP2.III]
-------------------	---------------------------------------	------------------------------------	------------------------------------	---------------------------------------	-------------------	--------------------------------	---------------------------------	--

látással lettünk gazdagabbak – bár ezzel már áttevünk a gender studies veszélyes vizeire. De hát mit is tehetnénk mi, balga nők, ha egyszer „*asszonyállatnak nincs anynyira való okossága; az istenek a férfiakat több okossággal áldották meg*”. Mentegetőzik is az egyik levélíró hölgy, hogy „*ennekem kicsiny fejem vagyom, és sokáig kell gondolkodnom*”.

Mindeközben az emlékiratok férfi írói meglehetősen büszkeséggel írnak arról, hogy ők maguk mennyire vonzóak voltak, sőt, azok is maradtak! Székely László még visszafogott: „*Virágjában lévén életem és időm, nem dicsekedve mondom, de bizony sok szem kapott rajtam, s én is nem utáltam a szépet*.” Bethlen Miklós azonban nem fogja vissza magát, az alábbiakban egészen részletekbe menően megismerkedhetünk külzetével:

„*Termetem sem hosszú, sem kurta, hanem tisztességes középáránt való volt, ábrázatom hosszúka, barnaszeg, piros. Ifjúságomban elég szemet szűrt; elég az, férfi ábrázatban sokaknál volt szebb, és vénségemig is tartott, sokaknak csudájára, szépségem, kivált nyomorúságos esetemhez képest. [...] már 67. esztendőben járok, mégis bajuszomnak, szakállomnak fele ősz, a hajamban pedig talán huszonöt szál sem fejr, sőt observáltam ezelőtt huszonkét esztendővel, a bajuszomban néha egy-két szál ha fejedredni kezdett, egy kevés üdő mulva megsárgult s visszafeketedett...*”

[Itt most közbeszólunk, de hát mégis, mennyire más hozzáállás ez, mint amikor az egyik drámaszereplő azt vágja a házasságtól vonakodó fiatal hős nő arcába, hogy: „*ha megvénhedel, s vajon ki szeretne tégedet?*”]

De visszatérve Bethlen Miklós külsejéhez, mint írja, „*Szemem felette fekete és fényes volt, akármely sólyom szeménél nem alábbvaló, leánynak sokat ért volna.*” S hogy végképp ne maradjunk információhiányosak, még egy rövid részlet: „*csont; nagy térdem vastag erek, inakból állott a testem, mégis az orcám, nyakam, torkom, combjaim, tomporám szép húsosak, teljeseek voltak, úgyhogy az sem testesség és kövérség, ösztövérségnek vagy szikkadt szárazságnak nem mondattatható, hanem Istentől áldásul adatott könnyűségnek tartottam; ahhoz képest jó gyalog és (nem kérkedem véle) de jó könnyű és kemény s merő derekú, meglehetősen szép lovas voltam.*”

Nos, ilyen alapos jellemzés talán egyetlen nőről sincs, még szerencse, hogy az író nem akart kérkedni! Persze vannak azért olyan urak is, akik nagyobb tehernek érzik az idő múlását. Ahogy Károlyi Sándor írja feleségének, Barkóczy Krisztinának: „*A mentét és süveget elvárom, édesem, ha ideje nincs is, mindazáltal örömet cifráznám magamat, mert igen őszül a fejem, nem kapnak rajtam.*” Pedig, mint tudjuk, egy tisztes öszes halánték...

De hagyjuk most az ős hajszálakat és az őszi napsugarakat, hiszen beköszöntött a tavasz, egyre több fa és bokor borul virágba, és hadd idézzük így (csipetnyi sóvárgással fűszerezve) azt az 1721-es levelet, mely szerint: „*Ami pedig az akadémiánkat illeti, Istené a dicsőség, minden szcienciákra nézve szépen floreál.*” Legyen ez így most is, floreáljon fű, fa, virág és Akadémia (> Intézet > Kutatóközpont), s „*Kívánjuk Istentől szívünk szerint, hogy előbbeni jó egészségét megadván Isten kegyelmednek, éltesse sokáig.*”

MELLÉKNEVEK DISZTRIBÚCIÓS ÉS SZEMANTIKAI MINTÁZATAI

HÉJA ENIKŐ¹, GÁBOR KATA², GYÖRFFY ANDRÁS¹, LIGETI-NAGY NOÉMI¹,
SIMON LÁSZLÓ¹, LIPP VERONIKA¹ | ¹HUN-REN Nyelvtudományi Kutatóközpont

²INALCO, Equipe de recherche textes, informatique, multilinguisme

{heja.eniko, gyorffy.andras, ligeti-nagy.noemi, simon.laszlo, lipp.veronika}@nytud.hun-ren.hu

kata.gabor@inalco.fr | DOI: 10.18135/PG70.2024.7

1. BEVEZETÉS

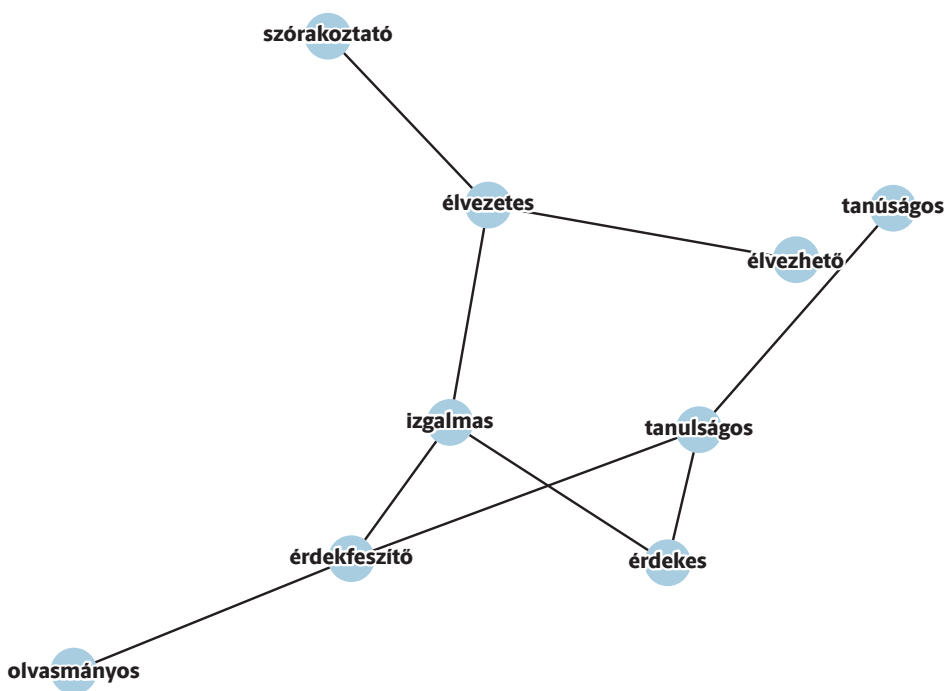
Jelen írásban egy folyamatban lévő kutatást vázolunk fel, amelynek eredményei egyaránt felhasználhatóak a lexikális szemantika, a lexikográfia és a nyelvtechnológia területén (l. Héja et al., 2023).

Amennyire tudjuk, a magyar melléknevek lexikális szemantikáját még nem írták le kimerítően. Részletesen Kiefer Ferenc foglalkozott ezzel a témakörrel a *Strukturális Magyar Nyelvtan* 4. 8. fejezetében. Kiefer (2008) a szemantikai csoportok elkülönítése során a melléknevek viselkedését vette alapul, és a következő négy tulajdonság alapján a mellékneveket három osztályba sorolta: figyelembe vette, hogy a melléknevek megjelenhetnek-e attributív vagy predikatív pozícióban a módosított főnévhez képest, fokozhatóak-e, továbbá módosíthatóak-e a *nagyon* határozószóval. Ezen tulajdonságok alapján a következő csoportokat különítette el (1) abszolút melléknevek; (2) relatív melléknevek (a mértéket jelölő és értékelő melléknevek); (3) rendhagyó melléknevek. Ez utóbbi csoportot példázza a „*János volt az állítólagos betörő*” mondatban az *állítólagos* melléknév. Kutatásunk egyik célja, hogy megvizsgáljuk, hogy az általunk javasolt korpuszvezérelt módszer milyen mértékű átfedést mutat a Kiefer (2008)-ban található csoportokkal, de ambicionáljuk az osztályozás további finomítása mellett a melléknevek szélesebb körére való kiterjesztést is – amennyiben ezek lehetségesek.

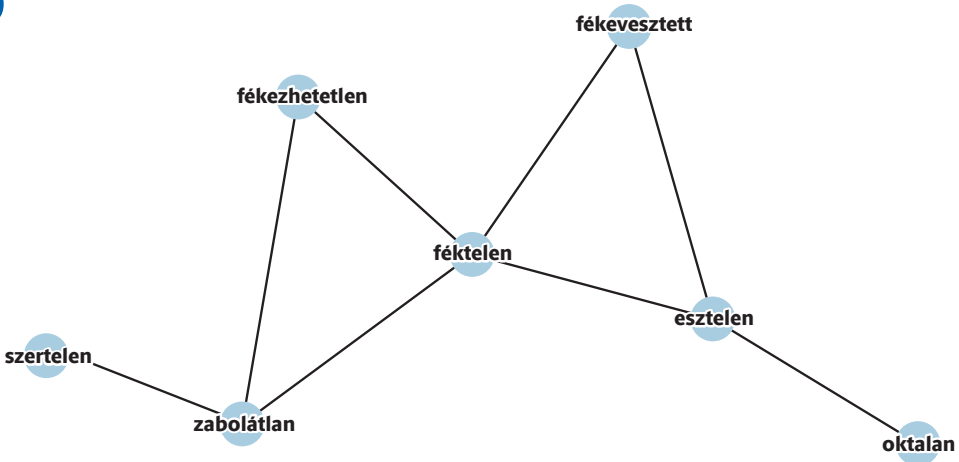
Elvárásunk szerint a melléknevek lexikai szemantikai leírása felhasználható alkalmazott nyelvészeti kutatásokban is: a lexikográfiában egynyelvű értelmező szótárak szerkesztését segíti, míg a nyelvtechnológiában a nagy nyelvi modellek (pl. ChatGPT) kiértékelését lehetővé tevő konzisztens benchmarkkorpuszok létrehozását könnyíti meg, illetve teszi konzisztensebbé. Fontos hangsúlyozni, hogy a javasolt módszer adatvezérelt, így a leírás során kiküszöböljük az érzeményszemantikát, a szemantikai intuíciónkra való támaszkodást. De vizsgáljuk annak a lehetőségét is, hogy az eredmények kiértékelése során mennyiben tudjuk kizárni az intuíció szerepét: a Kiefer (2008) által felhasznált formai kritériumok mellett más disztribúciós kritériumokra is támaszkodunk, elsősorban a koordinálhatóságra, amelyet kontextuális szóbeágyazásokra építve vizsgálunk.

2. A MÓDSZER LEÍRÁSA

Az alkalmazott módszer statikus szóbeágyazásokon alapszik (Mikolov 2013a, 2013b). Ezt a fajta szórepresentációs technikát gyakran éri az a kritika, hogy mivel egy szóalakhoz egy vektorrepresentációt rendel, önmagában nem képes a jelentések elkülönítésére (pl. Camacho-Collados & Pilehvar, 2018). Ezt úgy orvosoltuk, hogy a szóbeágyazásokat gráfstruktúrába rendeztük. A módszer lényege, hogy a szóvektorok hasonlóságát felhasználva elkészítjük a gráfot reprezentáló szomszédsági mátrixot, amelyet a következő lépésben binarizálunk egy K küszöbérték mentén. A küszöbérték ebben az esetben egy vágást jelent az eredeti, teljes, súlyozott melléknévi gráfon, amelyet ezáltal összefüggő komponensekre bontunk. $K = 0.7$ esetén a 10.153 melléknévet tartalmazó eredeti melléknévi gráfot 1.807 összefüggő gráfkomponensre bontjuk fel, amelyek összesen 6.417 melléknévet tartalmaznak. A komponensek figyelemre méltó tulajdonsága, hogy – egy kivételével – jól definiált szemantikai tartományokhoz tartoznak. Az ilyen hálózatok esetében az egyik komponens mindig egy óriáskomponens, amely a bemeneti melléknéveknek kb. harmadát (3.736) tartalmazza, és amely így több szemantikai domaint is összevon. Az alábbiakban két példát mutatunk összefüggő melléknévi gráfkomponensekre, az első az *izgalmas* melléknév rokon értelmű szavait tartalmazza, míg a második a *féltelen* szemantikai domaint fedi le.



1. ábra. Összefüggő melléknévi gráfkomponens: az *izgalmas* melléknév és rokon értelmű szavai

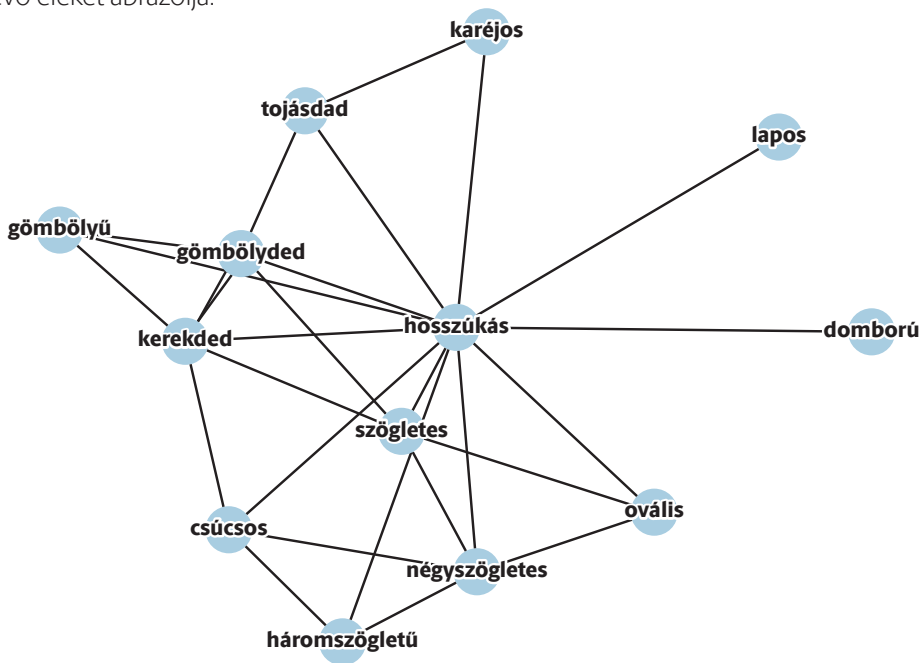


2. ábra. Összefüggő melléknévi gráfkomponens: a féktelen melléknév és rokon értelmű szavai

3. EREDMÉNYEK

3.1 Átfedések Kiefer melléknévi csoportjaival

Az eredmények azt mutatják, hogy az automatikusan kinyert melléknévi szemantikai domainek sok esetben megfeleltethetőek a Kiefer (2008)-ban meghatározott csoportoknak. Ez a megfigyelés elsősorban az abszolút melléknevek alcsoportjai esetében igaz: a színnevek, a nemzetiségek nevei, az eredetnevek és az alaknevek mind elkülöníthető gráfkomponenseket alkotnak az automatikusan generált melléknévi gráfban. A következő ábra a *hosszúkás* melléknév szomszédos csomópontjait és a közöttük lévő éleket ábrázolja.



3. ábra. A *hosszúkás* melléknév szomszédos csomópontjainak gráfja

Érdekes, hogy a privatív melléknevek nem alkotnak disztribúciósan koherens csoportot: a Kiefer (2008)-ban felsorolt privatív melléknevek (*vak, süket, néma, sánta*) mind magányos gráfcsomópontokat alkotnak. Ennek disztribúciós okait még vizsgálunk kell. Az emberi tulajdonságot jelölő, poláris párral nem rendelkező mellékneveket módszerünk további alcsoportokra bontotta. Megfigyelésünk szerint a relatív melléknevek sem alkotnak koherens csoportot, ennek okait is tovább kell vizsgálunk. A rendhagyó melléknevek a privatív melléknevekhez hasonlóan sokszor fordulnak elő izolált csomópontokként (*állítólagos, lehetséges, előző, egyedüli, fő, közeli, biztos, teljes*). Ez alól az *igazi* és a *tökéletes* melléknevek kivételek: az előbbi a *valóságos* és *valódi* melléknevekkel, az utóbbi pedig az *ideális, optimális* melléknevekkel fordul elő egy gráfkomponensben.

Összefoglalva azt mondhatjuk, hogy az eredményül kapott melléknévi osztályok Kiefer (2008) abszolút mellékneveinek alcsoportjaival mutatnak nagy fokú egyezést. A relatív és rendhagyó melléknévi osztályok esetén további vizsgálatok szükségesek az eltérések magyarázatára.

3.2 Lexikográfiai felhasználás

Láttuk, hogy az automatikusan létrehozott gráfkomponensek részben a szakirodalomban is szereplő szemantikai osztályokba sorolják a mellékneveket. A módszer egy lexikográfiai felhasználása, hogy az így kinyert melléknévcsoportokat automatikusan *jelentésstruktúrákba* szervezzük. Ebben a lépésben vektorrepresentációik alapján klaszterezzük azokat a főneveket, amelyek előfordulnak a melléknévi szemantikai osztály elemeinek valamelyikével. Az eredményül kapott főnévi klaszterek képezik a jelentésstruktúra alapját. A következő lépésben egy egyszerű korpuszalapú asszociációs mértékkel meghatározzuk, hogy milyen erős a kapcsolat a melléknévi szemantikai osztály egy adott eleme (pl. *esztelen*) és a főnévi klaszterek egy adott eleme között (pl. [*indulat, harag, düh* stb.]). Az 1. táblázat azt mutatja, hogy az *esztelen* leginkább a [*költekezés, pazarlás*] főneveket szubkategorizálja, míg például a [*nevetés, buli, öröm* stb.] szemantikai osztályba tartozó főnevek kevésbé *esztelenek*, mint inkább *féktelenek*.

	4	5	6	9
esztelen	3.2	26.5	13	7.22
fékevesztett	2		1.5	6.83
fékezhetetlen	2.33	2		6.62
féktelen	21.6	12	7.5	21.64
oktalan	3.25	1.5	3.75	12.67
szertelen	6			9.88
zabolátlan	1.33			9
0	nevetés	költekezés	öldöklés	vágy
1	buli	pazarlás	lövöldözés	szenvédély
2	öröm		háborúskodás	indulat
3	száguldás		pusztítás	harag
4	zabálás		ámokfutás	düh
5	ivászat		rablás	gyűlölet
6	jókedv		rombolás	nacionalizmus
7	tombolás			önzés
8	mulatozás			

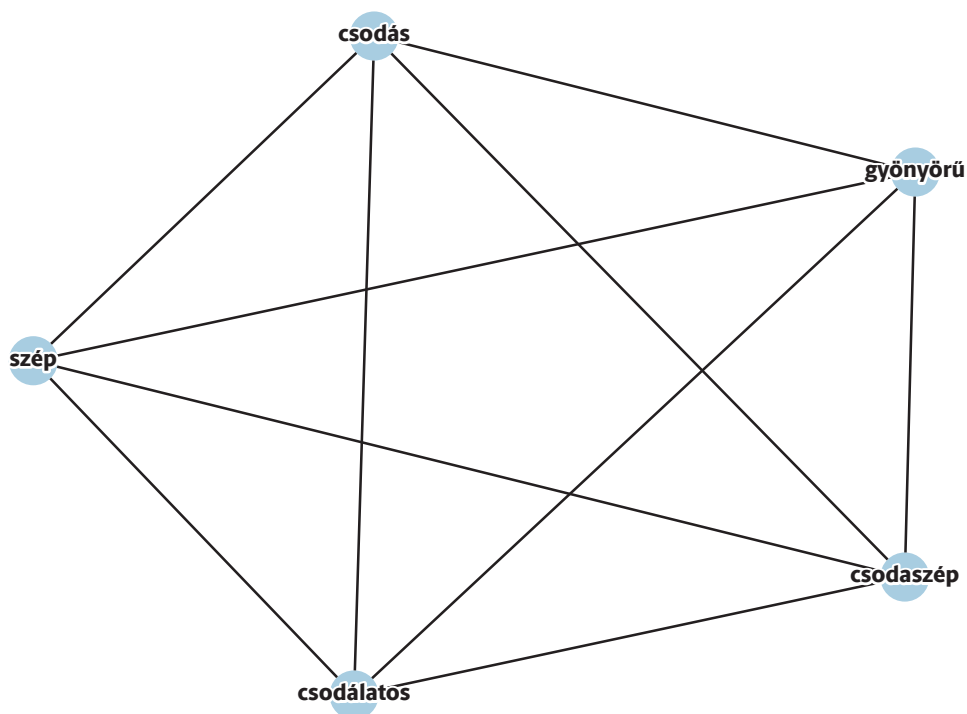
1. táblázat. A féktelen szemantikai osztályba tartozó melléknevek jelentésstruktúrája

Így tehát egy ilyen táblázatban ábrázolható jelentésstruktúra alapján képet kaphatunk arról, hogy az egy szemantikai domainbe tartozó melléknevek milyen szemantikai osztályba tartozó főneveket milyen erősséggel szubkategorizálnak.

3.3 Szemantikai relációk megragadása a melléknévi gráf lokális tulajdonságai alapján

3.3.1 Szinonima

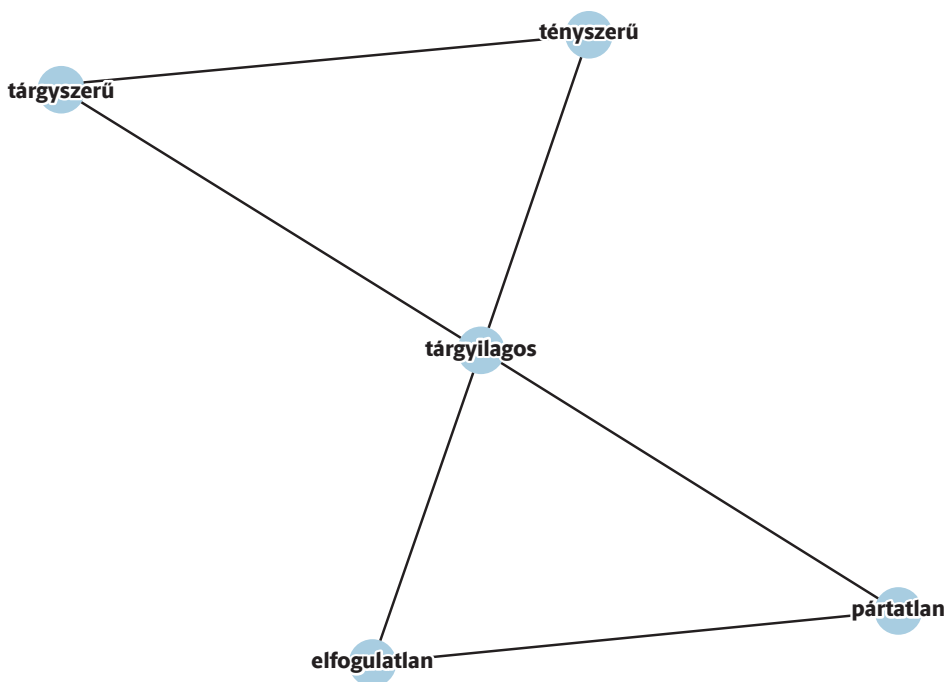
Azt már láttuk, hogy az összefüggő gráfkomponensek szemantikai domaineknek feleltethetőek meg. Ugyanakkor azt tapasztaltuk, hogy a szinonima disztribúciós fogalma is könnyedén értelmezhető a disztribúciós hasonlóság alapján létrehozott gráfokon. Mivel a gráfban két csúcs akkor szomszédos, ha a csúcs által jelölt melléknevek disztribúciós hasonlóságot mutatnak, feltételezzük, hogy a *teljes részgráfok* (ún. klikkek) lehetővé teszik a szinonimaosztályok – vagyis a melléknévi jelentések – megragadását, hiszen egy klikkben szereplő bármely két melléknév disztribúciósan hasonló lesz egymáshoz. Az alábbi ábrán a szép melléknevet tartalmazó melléknévi klikk látható.



4. ábra. A szép melléknevet tartalmazó melléknévi klikk (teljes részgráf)

3.3.2 Poliszémia

Ha a klikkek lehetővé teszik a melléknévi jelentés megragadását, akkor a melléknévi poliszémiát reprezentálhatjuk úgy, hogy egy melléknév egyszerre több klikkhez is tartozik. A következő ábra a *tárgyilagos* melléknév automatikusan kinyert két aljelentését mutatja: egyfelől lehetünk tárgyilagosaak úgy, hogy ragaszkodunk a valósághoz, vagyis úgy, hogy *tényszerűek* vagy *tárgyszerűek* vagyunk, másfelől pedig úgy is, hogy egyik félnek sem kedvezünk, ebben az esetben *elfogulatlanok* vagy *pártatlanok* vagyunk.



5. ábra. A tárgyilagos melléknév automatikusan kinyert két aljelentése

A poliszém jelentések automatikus kinyerésén túl az egyes jelentéseket előhívó, triggerelő főnévcsoportok is automatikusan kinyerhetőek a korpuszból. Így például a *tárgyilagos* első jelentése a *leírás, ismertetés* főnevek előtt fordul elő, míg a második jelentése a *megítélés, vélemény, eljárás* főnevek előtt.

Míg az alulspecifikált jelentésű főnevek (pl. *kutya, virág*) esetében a különböző aljelentések mellett megjelenő kontextusok könnyen koordinálhatók, a többjelentésű szavakra jellemző, hogy a különböző jelentést aktiváló kontextusok nem szerepelhetnek egyszerre egy szóelőfordulás mellett (Zwicky & Sadock, 1975). Melléknevekre alkalmazva ez azt jelenti, hogy ha két főnév ugyanazon poliszém melléknév különböző aljelentéseit reprezentáló klikkekhez tartozik, akkor várakozásunk szerint nem lesznek koordinálhatók, például: # *piszkos háború és aszfalt*, # *finom mozdulat és halétel*, # *hideg tea és ész*.

A koordinációs teszt széles körű alkalmazása lehetővé tenné az automatikusan előállított melléknévi jelentés-megkülönböztetések validálását, ám ennek emberi erőforrásigénye magas. Ezért jelenleg azt vizsgáljuk, hogy *előtanított nyelvmodellekből* kinyerhető-e hasonló információ. Ehhez olyan ADJ N1 és N2 koordinációs sémákat generálunk, melyekben a poliszém melléknév két mellérendelt főnévvel szerepel. Így összevethetjük, hogy mekkora valószínűséget társít a nyelvmodell a második főnévhez aszerint, hogy az első főnév azonos vagy eltérő melléknévi jelentést hív elő. Alacsony valószínűség esetén a megkülönböztetést érvényesnek tekinthetjük (mint a fenti példákban), ellenkező esetben pedig érdemes megvizsgálni, milyen okból tűnik elfogadhatónak a disztribúciósan elkülönülő szóhasználatok koordinálása, például: *alpári stílus és kifejezés, nagy fokú érdeklődés és nyitottság*.

A dinamikuslexikon-elméletek (Pustejovsky, 1991) megkülönböztetik a véletlenszerű és a szabályos poliszémiát; az utóbbira jellemző, hogy lexikális szemantikai tulajdonságokkal körülírható szócsoporthoz többé-kevésbé produktívan, szabályszerűen illeszkedő, megjósolható jelentésváltozással jár. A szabályos poliszémia gyakran észrevehetetlen a beszélők számára, és – legalábbis egyes eseteiben – meglepő módon elfogadható a poliszém jelentéseket előhívó kontextusok koordinálása (Schumacher, 2013). A kapott eredmények validálásán túl a nyelvmodellek használata segíthet az ilyen szabályos poliszém melléknevek, melléknévcsoportok azonosításában is.

3.4 A melléknévosztályok és a nyelvtechnológia

A nagy nyelvi modellek kapcsán alapvető kérdés, hogy „értik-e” a nyelvet vagy csak a szavak bonyolult korpuszbeli valószínűségi eloszlása alapján imitálják a nyelvértést. Ennek a kérdésnek az eldöntésében segíthet a modellek lexikális szemantikai képességeinek vizsgálata. Az ilyen vizsgálatok olyan adatbázisokon – ún. benchmarkkorpuszokon – alapulnak, amelyekben a célszavak jelentése valamilyen módon jelölve van. Az egyik legelterjedtebb ilyen jellegű benchmarkkorpusz a SuperGlue benchmarkadatbázis (Wang et al., 2020) részét képező Word-in-Context (WiC) kiértékelő korpusz (Pilehvar & Camacho-Collados, 2019). Érdekes módon az egyéb feladatokkal ellentétben a WiC meglehetősen nehéznek bizonyult még a GPT-3-nak is (Brown et al., 2020): few-shot tanulás esetén a GPT-3 válaszai a véletlenszerű válaszoknak megfelelő pontosságot mutattak. Ennek feltételezhető oka, hogy a WiC benchmarkkorpusz a Princeton WordNet jelentés-megkülönböztetésein alapul, így jelentésannotációja nem következetes; még az emberek sem értenek egyet a jelentések konkrét elkülönítésében minden esetben. A melléknevek gráfolapú reprezentációjának és a szaliens főnévi kontextusok kinyerésének egyik fontos célja, hogy egy olyan lexikális szemantikai tudást mérő benchmarkkorpuszt hozzunk létre, ahol (1) a jelentések elkülönítése nem a WordNet alapján, hanem korpusz alapján történik; (2) megjelennek a többjelentésűség olyan alosztályai is, amelyek hatással lehetnek a poliszémia emberi észlelésére (pl. homonímia, reguláris vagy irreguláris poliszémia). Így elvárásaink szerint kiküszöbölhetjük, hogy olyan nyelvi tudást várjunk el a nagy nyelvi modellektől, amelyekkel mi magunk sem rendelkezünk.

Irodalom

- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*. Retrieved from <https://arxiv.org/abs/2005.14165>
- Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *arXiv preprint arXiv:1805.04032*. Retrieved from <https://arxiv.org/abs/1805.04032>
- Héja, E., Ligeti-Nagy, N., Simon, L., & Lipp, V. (2023). An unsupervised approach to characterize the adjectival microstructure in a Hungarian monolingual explanatory dictionary. In Medved', M., Měchura, M., Tiberius, C., Kossem, I., Kallas, J., Jakubiček, M., & Krek, S. (Eds.), *Proceedings of the eLex 2023 conference: Electronic lexicography in the 21st century (eLex 2023): Invisible lexicography* (pp. 150–167). Brno, Czech Republic: Lexical Computing.
- Kiefer F. (2008). A melléknevek szemantikája. In Kiefer F. (szerk.), *Strukturális magyar nyelvtan 4. A szótár szerkezete*. Budapest: Akadémiai Kiadó.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *CoRR, abs/1301.3781*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26*(10).
- Pilehvar, M. T., & Camacho-Collados, J. (2019). WiC: The word-in-context dataset for evaluating context-sensitive meaning representations. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers)* (pp. 1267–1273).
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics, 17*(4) (pp. 409–441).
- Schumacher, P. (2013). When combinatorial processing results in reconceptualization: Toward a new approach of compositionality. *Frontiers in Psychology, 4* (Article 677).
- Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2020). SuperGLUE: A stickier benchmark for general purpose language understanding systems. *Advances in Neural Information Processing Systems, 32* (NeurIPS 2019).
- Zwicky, A. M., & Sadock, J. M. (1975). Ambiguity tests and how to fail them. In *Syntax and Semantics volume 4* (pp. 1–36). Brill.

1984 – A MAGYAR NYELVÉSZETBEN

KENESEI ISTVÁN | HUN-REN Nyelvtudományi Kutatóközpont

kenesei.istvan@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.8

Újabb és ezúttal szokatlan említésben részesült a közelmúltban Prószéky Gábor, az MTA Doktora, a PPKÉ ITK egyetemi tanára, a Nyelvtudományi Kutatóközpont főigazgatója, a MorphoLogic alapító igazgatója, a Széchenyi-díj és több más magas kitüntetés birtokosa stb., stb., akinek a 70. születésnapját köszönti a jelen írás. Ez az említés azonban nem a fentiekben elősorolt tudományos vagy innovatív tevékenységéhez kötődik, hanem egy fiatalkori akciójához, amit nem is egyedül követett el.

Az Állambiztonsági Szolgálatok Történeti Levéltára (ÁBTL) egy eléggé népszerű, kb. 25 ezer követővel bíró Facebook-fiókot tart fenn, amelyen naponta több bejegyzést helyez el. 2024. január 11-én megjelent egy rövid közlemény és egy ügyirat, amelynek a tartalma jelen sorok írója számára nem volt ismeretlen.¹ A poszt az alábbi titkos jelentést tartalmazta, amelyben a BM ügynökének eredeti szövegét dőlt betűvel, az ÁBTL anonim munkatársának megjegyzéseit szögletes zárójelek között közlöm:

Budapest, 1984. február 1.

Barátom január első napjaiban a "Csendes" vendéglőben hosszabb beszélgetést folytatott a következő személyekkel:

Nádasdy Árpád /ELTE BTK Angol Intézet/ [helyesen Nádasdy Ádám]

Kálmán C. György /MTA Irodalomtudományi Intézet/

Kálmán László /Filmtudományi Intézet/

Proszéki Gábor /Zenetudományi Intézet/. [helyesen Prószéky Gábor]

A beszélgetés során politikai kérdések is szóba kerültek, valamint a magyarországi ellenzék helyzete általában. A fentebbi személyek egy lazább csoportot képeznek, bizonyos szálakkal a magyar ellenzék kemény magjához. A kapcsolatot Nádasdy tartja, akinek jó személyi viszonya van Demszkyvel, Szelényivel, Bíró Ivetttel és Bencze Györggyel. Résztvett az ellenzék nagy szilveszteri összejövetelén is, amelyet egy Gárdos /a név nem biztos/ nevű pszichológus házaspár magánlakásán tartottak.

A beszélgetés során szóba került a "Mozgó Világ" ügye is. Megemlítették, az ellenzék azt tervezi, hogy az első olyan számot, amelyet már az új szerkesztő nevével jegyeztek, az ellenzék tagjai borítékban beküldik a KB-hoz Aczél György elvtárs címére.

Az ÁBTL-munkatárs a következőket fűzte hozzá az idézett jelentéshez:

¹ <https://www.facebook.com/plugins/post.php?href=https%3A%2F%2Fwww.facebook.com%2Ftitkosszolgaltileveltar%2Fposts%2Ffbid02ySDE9QDGkMWnecQQ3i8uwz6sMWQ2bzXHdeo6HQYFEcX8EpKJdn5vrurjP-HMxZxDAI>

„Docens” fedőnevű titkos munkatársat a BM III/III-2-a alosztály foglalkoztatta 1978 és 1984 között. Az ügynök az ELTE BTK oktatóiról, hallgatóiról, volt tanítványairól írt elemzéseket, részletes, magánéletüket is érintő jellemzéseket. Számos jelentés szól a Hétfői Szabadegyetem (az ún. repülő egyetem) előadásairól, a szamizdatról és a Mozgó Világ-ügyről is.

Az ügynök Antal László nyelvész, egyetemi tanár az ELTE BTK Általános Nyelvészeti Tanszékének docense volt. 1987-ben jogellenesen külföldre távozott. Az 1962. augusztus 24-szeptember 12-ig Bostonban megtartott IX. Nemzetközi Nyelvészeti Kongresszusról jelentett. Ekkor tartótisztje: Éder Mihály r. százados volt.

Tekintsünk el attól, hogy a névtelen ÁBTL-es, miután kijavította a beszélgetésben résztvevők nevét, a többi nevet meghagyta hibásan (Bíró Yvette, Bence György), valamint hogy „Docens” a repülő egyetemen nem vett részt és nem jelentett se arról, se a szamizdatokról, ahogy többet a „Mozgó Világ” ügyéről sem, továbbá Antalt 1962-től folyamatosan foglalkoztatták (először „társadalmi kapcsolatként”, majd titkos munkatársként, ami az ügynöknél jóval magasabb sarzsi volt, végül 1985-ben hagyta el családotul Magyarországot, bár az igaz, hogy a dossziéját csak 1987-ben zárták le.)

Hogy a történet miért nem volt ismeretlen számomra, azt jó két éve írtam meg egy hosszabb cikkben, amelyet Antal László, a legendás hírű tanár akkor még alig ismert BM-informátori tevékenységének szenteltem, több mint 600 oldalnyi iratanyagának áttanulmányozása, valamint az MTA Levéltárban folytatott kutatás után.² Innen tudom például azt is, hogy az ügynökök az információk forrását „barátom” megjelöléssel azonosították, ám ezek mindig ők maguk voltak, hiszen sokszor saját kezűleg írták le a jelentéseiket.

De mi adott okot arra, hogy Antal László (élt 1930–1993), a „Docens” fedőnevű titkos BM-munkatárs jelentsen, hogy egyáltalán jelenthessen a négy nyelvészről? (Bár munkahely szerint akkor csak egyikük, és szakmájuk szerint is csak ketten, Nádasy és Kálmán László számítottak nyelvésznek, Kálmán C. György pedig élete végéig irodalomtudósként dolgozott, hadd nevezzük őket ebben az írásban nyelvésznek, amire egyébként jó okunk van, mint mindjárt kiderül.³) Antal akkoriban az amerikai típusú deskriptív nyelvészet mondhatnánk legjobb, de valójában egyetlen hazai művelőjeként szerzett nevet magának – és képezett egy magányos szigetet a jól beágyazott hagyományos leíró és történeti, valamint a szárnyait akkor még csak bontogató generatív nyelvészet között. Lenyűgöző előadói stílusa, gyors észjárása, egyedi ötletei és sziporkázó megjegyzései hallgatók nagy tömegeit vonzották: amikor ő tartott órát, a terem mindig tele volt, a diákok élénk figyelemmel és gyakori nevetésekkel nyugtázták nyelvészeti poénokkal (!) megtűzdelt előadásait. Nem csoda, hogy a négy fiatal nyelvész hozzá fordult, és nem is mentek üres kézzel.

² <https://www.es.hu/cikk/2022-02-25/kenesei-istvan/a-docens-egy-rejtely-mozaikkockai.html>

³ Mint azt a hivatkozott cikkemben is megírtam, mind a négy érintettnek, köztük a két Kálmán testvérnek még meg tudtam mutatni a jelentést és egyetértésüket bírtam, hogy azt közöljem az ÉS-ben, de a Kálmán testvérek a cikk megjelenését már nem érték meg.

Az Orwell disztópikus regényéről elhíresült 1984-es év idehaza és persze másképpen sok fiatal nyelvész számára is jelentőssé vált. (A könyv frissen kiadott „jubileumi” díszkötéses példányát egyébként abban az évben sikerült megvennem, majd gondosan átcsomagolva hazacsempésztem a szokásos vámvizsgálati izgalmak közepette.) Az 1970-es évek eleje óta a Szépe György által támogatott és köréje csoportosuló generáció akkorra már nemcsak újabb tagokkal szaporodott, hanem a hazai után a nemzetközi publikációs térben is megjelent. Az MTA Nyelvtudományi Intézetének a budai Várnegyedben a Szentháromság utcában volt a székháza: a korábbi budai városháza műemlék épülete az MTA kezelésében állt és amikor a Művelődésügyi Minisztérium Szalai utcai székházából ki kellett költözzön, a nemrég kinevezett új igazgató, az uralista Hajdú Péter szerencsés kézzel választotta ki az akkor még békés úri környezetet.

Az Úri utcára (és a Hadik-szoborra) néző tágas földszinti 13. szobában többeknek volt íróasztala 1984-ben: Komlós Andrásnak, Réger Zitának, Bánréti Zoltánnak, de a Charta 77 aláírása miatt elbocsájtott Pap Marinak, illetve a nemcsak nyelvészeti, de tudományos vezetői karrierre is esélyes, ám a 80-as évek elején az MSZMP munkatársává vált Radics Katalinnak már nem. Szabolcsi Anna (és Szépe György is) más szobákban lakozott, sőt ez idő tájt Kálmán László is bekerült az Intézetbe, ha nem is ebbe a híres szobába. Sokan „vendégként” voltunk ott: az intézeti napokon, kedden és pénteken fel-fellátogattunk, különösen, ha valamilyen esemény volt: É. Kiss Katalin, Mártonfi Ferenc, Nádasdy Ádám, Varga László az ELTE-ről, Kornai Andás a SZTAKI-ból, Prószéky Gábor a szomszédban lévő MTA Zenetudományi Intézetből, Hunyadi László Debrecenből, én pedig a szegedi egyetemről. De éveket töltött közöttünk éppen akkoriban a Fulbright-ösztöndíjas Farrell Ackerman, a University of California, San Diego, mostani professzora. A 70-es évek eleje óta időnként fel-felbukkant a magyarul kiválóan beszélő amsterdami Casper de Groot, vagy a Londoni Egyetemről a „mindenevő” hungarológus Peter Sherwood. Sokszor program híján is felmentünk, hogy nyelvészetről beszéljünk, aztán együtt mentünk ebédelni, többnyire a „Lovardába”, a mostanában Pénzügyminisztériummá ki- (vagy vissza-) cifrázott hajdani BME-kollégium oldalszárnyában működő hodályszerű önkiszolgáló étterembe, rémes fémtálcáival, műanyag edényeivel és pléh evőeszközeivel.

A 13-as szobát 40 év távlatából a modern, strukturalista vagy inkább generatív nyelvészet egyfajta szülőszobájának látom. Noha voltak rendszeres előadások, a hely vonzereje inkább abban állt, hogy ha ott voltunk, mindig előjött valamilyen probléma, amit többnyire vehemensen megvitattunk egymás között, így formálva egymás gondolkodását, valamint vitakészségét. A nyilvános előadásokra az Intézet más osztályairól is bejöttek, nemcsak a saját generációnkból, hanem idősebbek is, mint persze Szépe György (amíg 1980-ban nem ment Pécsre megalapítani az unikális kísérleti bölcsészkart ott, összefogva Horányi Özsébbel, Fülei-Szántó Endrével és a helyiekkel), Kiefer Ferenc (a hazajövele után), Bakos Ferenc, Molnár Ilona, Elekfi László.

1979-ben a kandidátusi értekezésemnek az ún. munkahelyi vitáját is ott tartottuk, mert ugyan a munkahelyem Szegeden volt, de Debrecenben, Dezső Lászlónál voltam aspiráns

és Budapest látszott a megfelelő mértani középarányosnak.⁴ A Szabolcsi Anna által vezetett 7 oldalas (!) jegyzőkönyv szerint 33-an voltak jelen, közöttük (a fent említettek kivül) például Róna-Tas András, Honti László, Kanyó Zoltán, Kelemen János, Szende Tamás, Siptár Péter, Nádasdy Ádám és a külföldön tartózkodó Kiefer Ferenc voltak az opponensek. A Szépe György és a debreceni Korponay Béla társelnökségével órákon át tartó védés során tizen szóltak hozzá kritikusan, néhányan többször is. Ezt az eseményt csak azért idézem itt föl hosszabban, mert egyrészt azt mutatja, hogy milyen központja volt a 13-as szoba a nyelvtudományi történéseknek, másrészt mert ennek az évtizednek a fordulóján kezdett el a nemzedékünk tudományos fokozatokat szerezni: É. Kiss Katalin például épp az 1979-es évben lett kandidátus, Hunyadi és Radics 1981-ben, és sorra a többiek.

De térjünk vissza az 1984-es évhez! Ez volt az első olyan időszak, amikor az az ifjú nyelvésztársaság, amelyet a fentiekben megpróbáltam bemutatni, mintegy testületileg megjelent külföldön, ráadásul nem is akárhol. Ma már nehéz pontosan visszaidézni az események sorrendjét, ezért nem garantálom az emlékeim pontosságát. A 80-as évek elején már jelentősen megnőtt a külföldi vendégek száma a nyelvészetben (is). Rendszeres hazajáró lett például a californiai Santa Barbarából Hetzron Róbert, aki a szakmában a kusita és etiópiai nyelvek tudósaként volt ismert, de éles füllel hallotta meg a magyar hangsúly és intonáció finom megkülönböztetéseit is. A szakdolgozatát a Groningeni Egyetemen írt Marác László a magyar névutók és esetragok kérdéseiről készített doktori disszertációt. Mindketten gyakori vendégei voltak a 13. szobának. Valószínűleg Kiefer Ferenc és a groningeni germanista professzor Werner Abraham közötti kapcsolatnak volt köszönhető, de lehetett benne része Marácznak is, hogy az 1984 tavaszán szervezett *Topic, Focus and Configurationality* c. konferenciára népes magyar társaságot hívtak meg a helyiek, jórészt (nyugat-)német és cseh nyelvészekkel együtt, de megjelent Izraelből Julia Horvath és Bécsből Hubert Haider is.

Néhány további jól ismert külföldi nyelvész hadd soroljak fel: Eva Hajičová, Petr Sgall (Prága), Joachim Jacobs, Arnim von Stechow, Susanne Uhmann (NSZK), Sjaak de Mey, Wim Scherpenisse (Groningen) – amiből következik, hogy a konferencia tulajdonképpen „magyar show” volt, hiszen az alábbiak mind magyar nyelvi kérdésekkel foglalkoztak: Ackerman, Horvath, Komlósy, Marác (és de Mey), Hunyadi, Kenesei, Kiefer, É. Kiss, Szabolcsi, valamint a „négyes csapat”: Prószéky, Nádasdy és a két Kálmán testvér. Érdekes, hogy a magyar tárgyú kutatásokat is végző már említett amsterdami funkcionális nyelvész Casper de Groot nem volt jelen, továbbá hogy a szervezők az Általános Nyelvészeti és Német Tanszékek mellett megemlégték az akkor még létező Finnugor Tanszéket is, ahonnan azonban úgy emlékszem, senki nem vett részt (és amelynek csak később lett a professzora Honti László).

A téma aktualitása nyilvánvaló volt, hiszen a megelőző időszakban kapott egyre nagyobb figyelmet az addig elhanyagolt szórendi problémák beillesztése a generatív

⁴ Az aspirantúra a szovjet mintájú rendszerben hároméves továbbképzési ösztöndíj volt a kandidátusi értekezés megírása céljából.

nyelvtan keretei közé; magyar viszonylatban szinte egy időben jelentkezett két alternatíva is: É. Kiss Katalin (1981) nonkonfigurációs javaslata, illetve Julia Horvath (1981) konfigurációs megoldása, miközben élt még a prágai iskola „funkcionális mondatperspektíva” elnevezés alatt elterjedt hagyománya is, s ennek volt köszönhető az ő jelenlétük is.

É. Kiss Katalinnak az 1970-es évek második felében megkezdett szórendi kutatásai egyrészt a nemzetközi generativista körök számára, másrészt a formális szabályoktól mentes, mintegy impresszionisztikus leírásokban gazdag magyar nyelvészetben is friss légáramlatként jelentkeztek. É. Kiss visszanyúlt egy gyakorlatilag elfeledett és a saját korában követőkre nem talált 19. századi újításhoz, amely Brassai Sámuel nevéhez kötődött, és szembemenve a mondat szerkezet bináris elemzésével a Topik–Predikátum, és az utóbbin belül a Fókusz–VP szerkezet bevezetésével adott magyarázatot a szabad szórendre, hiszen mind a Topikban, mind a Fókuszban bármely összetevő (az előbbiben több is) megjelenhet, továbbá a VP-ben az összetevők sorrendje szabad kivéve, hogy a ragozott ige mindig szorososan követi a Fókuszt.

Ez a radikálisan új elképzelés egy csapásra azonnali hívekre és kritikusokra is talált: a Brassai–É. Kiss modell rendkívül természetes módon helyezi el a magyar mondat összetevőit, de a modell gyermekbetegségeit a bírálók azonnal elkezdték elősorolni. Az egyik legfeltűnőbb gondot az igekötők elhelyezése okozta: É. Kiss első verziójában a Fókuszban talált helyet nekik, hiszen az igekötő (látszólag) a fókuszos összetevővel állt komplementer disztribúcióban. Csakhogy a prozódiai viszonyok ennek ellentmondtak, és erre építették alternatív javaslataikat a vitázók: a „groningenieiek” közül Hunyadi László, aki elsősorban a hangsúlyviszonyokat és a kvantorokat figyelte, illetve a „négyes csapat”, akiket ott Prószéky képviselt, deskriptív szórendi és prozódiai megfigyeléseiket tették közzé, valamint e sorok írója, Kenesei István, aki a fókusz logikai értelmezését tárgyalta. Prószéky és társai talán legfontosabb eredménye a főnévi igeneves állítmányokat tartalmazó vizsgálata során a kétfajta hangsúlyozású mondatok elkülönítése volt azzal a következménnyel, hogy az igekötős igék alapesetben semleges, egyenletes hangsúlyozású, „lapos” prozódijú mondatokban állnak, míg a fókuszos mondatokban a fókusz „irtó” nyomatéka hangsúlytalanná redukálja a mögötte álló összetevőket. Mindez kiegészült egy tüzetes igeosztályozással a szerint, hogy milyen jelentés-, illetve hangsúlymintába illeszkednek bele ezek az infinitívuszt vonzó igék.

A groningeni előadások még az évben megjelentek egy afféle „könyomatos”, azaz sokszorosított kiadványban, a *Groninger Arbeiten zur germanistischen Linguistik* 24–25. számában (Abraham & de Meij, 1984), majd két évre rá egy válogatás belőlük átdolgozott formában a *Topic, focus and configurationality* c. nyomtatott kötetben (Abraham & de Meij, 1986), amelybe a magyar(országi) résztvevők közül Hunyadi, Kenesei, Kiefer, É. Kiss, Komlósy, Szabolcsi és a „négyes csapat” adtak cikket.

De ezzel még nem volt vége az 1984-es év nyelvészeti termésének, mert a groningeni nemzetközi találkozóval párhuzamosan egy jelentős hazai gyűjtemény elő-

készületei is zajlottak és a 13. szoba nagy vitáiban ezek is kezdtek végleges formát önteni. Az *Általános Nyelvészeti Tanulmányok* 17. kötetéről van szó (Kiefer & Telegdi, 1989), amelynek a tárgyköre az alcíme szerint „tanulmányok a magyar mondatköréből” volt. Jóllehet a megjelenés a rendszerváltást megelőző és a könyvkiadást is sújtó gazdasági helyzet miatt az évtized végéig elhúzódott, a kéziratokat már 1984-ben leadtuk, és legfeljebb csak az (akkori papír alapú és sajátos nyomdai jelzésekkel végrehajtott) korrektúrákba helyezhettünk el apróbb javításokat.

Ennek volt köszönhető nem csak, hogy mind a groningeri első változatok, mind a gyűjteményes kötetben publikált végleges cikkek idegen nyelven előbb jelentek meg, mint magyarul, hanem hogy a groningeri és az ÁNyT-beli tanulmányok témái több-kevesebb rokonságot mutattak. A groningeri csapat tagjai közül az ÁNyT-ben É. Kiss Katalin, Kenesei István, Komlósy András és a Kálmán–Kálmán–Nádasdy–Prószéky szerzőtársak voltak jelen, kiegészülve mindazokkal, akik a Nyelvtudományi Intézet szervezésében szintén a „Strukturális magyar nyelvtan” projekt munkálataiban részt vettek: Bánréti Zoltán (mellérendelések), Fabó Kinga (igeszemantika), Kornai András (főnévi csoport), Varga László (prozódia), Wacha Balázs (igeszemlélet), valamint Prószéky Gábor egy saját írással a határozókról. Ezek valóban csupán előtanulmányok voltak, részben mert még nem alakult ki, milyen fejezetek kerüljenek a majdani kötetbe, részben mert valóban szükség volt még az alapkérdések tisztázására, ezért többen is megengedhették maguknak, hogy alternatív elképzeléseket hozzanak nyilvánosságra, például Kornai vagy Kenesei.

Az ekkor már a külföldről hazatért Kiefer vezetése alatt működő munkacsoport nem kevés kiadói hányattatás után végül 1992-re „szülte meg” a sikeres és máig forgalomban lévő *Strukturális magyar nyelvtan* első, *Mondattan* című kötetét (Kiefer, 1992), amelybe a tartalmi szűkítés és a szerzők számának ezzel járó csökkentése miatt a következők maradtak, illetve kerültek be: a mondat szerkezete (É. Kiss Katalin), főnévi csoport (Szabolcsi Anna és Laczkó Tibor), vonzatszerkezetek (Komlósy András), alárendelt mondatok (Kenesei István), mellérendelés (Bánréti Zoltán), aspektus (Kiefer Ferenc), kiegészítve egy grammatikaelméleti bevezetővel (É. Kiss és Szabolcsi). A két lektor egyike pedig Nádasdy Ádám lett; a másik, Rácz Endre, a hagyományos magyar nyelvészet felől érkezett. Ámbár ez még kevésbé tartozik tárgyévünkhöz, két évvel később a külföldi olvasóközönség számára átdolgozott angol nyelvű változat is napvilágot látott (Kiefer & É. Kiss, 1994). Ugyan mindez már egy másik év és másik történet, de a gesztáció időszakába a groningeri konferenciákat és az ÁNyT-t be kell számítani.

Személyes emlékezetemben ez az 1984-es évszám egy másik fontos mérföldkövet is jelez. A 13-as szobában az is egyfajta összetartó kapocs volt, hogy a 70-es évek elején Szépe György egy nyelvtanoktatási problémaadagot prezentált, amiből egy magyar nyelvi tankönyvcsoomag összeállítása kerekedett ki, az ezt létrehozó társaság pedig magát *Fiatal nyelvészek munkaközösségének* nevezte el. Alapító tagjai Bánréti Zoltán, Füredi Mihály, Kenesei István, Komlósy András, Mártonfi Ferenc, Pap Mária,

Radics Katalin, Terts István, Tóth Pál voltak, majd újabb „évfolyamok” csatlakoztak hozzájuk: Heller Mária, Rohonci Katalin és Szabolcsi Anna személyében. Egy amolyan bevezető-kísérő füzetet is megszerkesztettünk 90 lap terjedelemben *Ember és nyelv* címen, és évekig terveztük, hogy egy önálló kötetet formálunk belőle, de a munka valahogy mindig elakadt. A végén én vállaltam el, hogy összeállítom, a még hiányzó fejezetekkel kiegészítem, és részben a régi „belső” tagokra, részben újonnan felkért, a társasághoz is kötődő, de a munkaközösségtől eltérő szakterületeken járatos szerzőkre támaszkodva megszerkesztem a könyvet. Az előbbieik közé Pap Mária, Radics Katalin (akik ketten békésen megfértek egymással a könyv lapjain)⁵, Réger Zita, Rohonci Katalin és Szabolcsi Anna, az utóbbiak közé Kelemen János és Pléh Csaba tartozott. A Gondolat Kiadó révén *A nyelv és a nyelvek* első kiadása is éppen 1984-ben jelent meg (Kenesei, 1984). A kiadásról kiadásra átdolgozott és alkalmasint bővülő könyvbe azután újabb fejezetek kerültek, aligha meglepően ugyanazon kör tagjaitól: Nádasdy az írásról, Prószéky (Kis Balázssal társszerzőként) a számítógépes nyelvészetről.

Ugyanebben az évben a Művelődési Minisztérium meghirdetett egy tudománytámogatási lehetőséget egyik intézményének, a Tudományszervezési és Informatikai Intézetnek a szervezésében. A szegedi Angol Tanszékről benyújtott szerény összegű pályázatomban pályázatom elnyerésével módomban állt egy kétkötetes angol nyelvű kiadványt megjelentetni, amelyben egyrészt az addigra már rendszeresen publikáló fiatal nemzedék tagjai leíró jellegű fejezeteket tettek közzé a magyar nyelvtan különféle fejezeteiről, másrészt egy nemzetközi szerzőgárda a magyar nyelv elméletileg érdekes problémáit mutatta be.

Az e kötetekkel szárnyra kelt *Approaches to Hungarian* sorozat 1985-ben megjelent első darabjában (Kenesei, 1985) az alábbi szerzők és témák sorakoztak: Kálmán László (a semleges, valamint a fókuszos mondat szórendje, továbbá inflexiók morfológia), Hunyadi László (kvantorok hatóköre és sorrendje), Komlós András (vonzatszerkezet), Kornai András (főnévi csoport), Laczkó Tibor (deverbális főnévi szerkezetek), Prószéky Gábor (melléknévi igenevek), Kenesei István (alárendelt mondatok), Kerkovits Andrea (ellipszis), Bánréti Zoltán (áthidalás [*gapping*]), Varga László (intonáció), Nádasdy Ádám (fonológia, morfofonológia).

A két évre rá megjelent 2. kötet (Kenesei, 1987) szerzői és témái: É. Kiss Katalin (a lapos VP szerkezetéről), Marác László (a projekciós elv a magyarban), Kenesei István (névmások a Kötéselv fényében), Julia Horvath (aszimmetria-jelenségek a szabad szórendben), Szabolcsi Anna (funkcionális kategóriák az NP-ben), Donka Farkas (tárgyi névmások elhagyása), Farrell Ackerman (névmási igemódosítók), Kálmán László

⁵ Emlékszem, volt bennem némi félelem, hogy mivel Pap Mari publikációs tilalom alatt állt, vajon nem lesz-e ez ártalmára a könyv megjelenésének. Másfelől meg bíztam benne, hogy Radics Katalin neve ugyanakkor kioltja a betiltási szándékokat. Végül azt hiszem, a cenzorok (vagy akikre ezek az ügyek tartoztak) észre sem vették. Legalábbis hozzám semmilyen kifogás vagy visszajelzés nem jutott el.

(fejvezérelt frázisszerkezet nyelvtan a magyarban), Casper de Groot (határozói igeneves passzív szerkezetek). A két kötet kereskedelmi forgalomba nem került, cserekiadványként terjedt, közvetve a tanszéki könyvtárat gazdagítva. Az összesen 16 kötetet megért (és egy nemzetközi periodikában folytatódó) *Approaches to Hungarian* sorozat további sorsa itt most nem releváns, de része volt annak a folyamatnak, amely során minél szélesebb körben ismerte meg a világ nyelvtudománya a hazai tudósok eredményeit.⁶

Az 1984-es évvel mintha zsilip szakadt volna fel a magyar nyelvészet nemzetközi kapcsolataiban. A Wolfgang Dressler és Kiefer Ferenc közötti együttműködés jóvoltából 1985-ben Bécsben egy közös szintaktikai szemináriumot szerveztek, majd elindult a máig tartó kétévenkénti, (többnyire) bécsi és budapesti helyszínekkel szervezett morfológiai konferenciasorozat. 1986-ban újra Groningen hívta meg a magyar nyelvészeket egy ott és Amsterdamban rendezett előadásorozatra, majd 1987 nyarán az Amerikai Nyelvészeti Társaságnak (LSA) a californiai konferenciájával egybekötött nyári nyelvészeti egyetemén (*Linguistic Institute*) vett részt sok hazai nyelvész. És innen nem volt megállás...

A harmincas éveikben járó nyelvészeknek ez a sok áttörést hozó éve az e kötetben ünnepelt Prószéky Gábornak is fordulatot jelenthetett pályáján. Programtervező matematikusként először a Zenetudományi Intézetben lett állása, onnan járt át a szomszéd Nyelvtudományi Intézetbe, a már többször is emlegetett 13-as szobába, és első szorosán vett nyelvészeti publikációja éppen a „négyes csapatnak” a segédigés cikke volt: éppen az, ami miatt felkeresték Antal Lászlót, az egyetlen szakembert, akitől saját deszkriptív elemzésükhöz segítséget várhattak – de nem kaptak. Ugyan csupán szóban ismertették az elképzelésüket, Antalt valószínűleg nem érdekelte, hidegen hagyta a probléma. Egyébként ha egy írott változatot küldtek volna el neki, sokak tapasztalata szerint (beleértve a magamét is) el se olvasta volna. Ők négyen végül üres kézzel távoztak a találkozóról – szemben Antallal, aki felhasználta a kínálatkozó lehetőséget, hogy kikérdezze az iránta bizalmat tápláló fiatal társaságot, majd tartótisztjének jelentsen egy olyan ügyben, amire konkrét megbízást se kapott, csak magánszorgalomból túlteljesített.⁷

Prószéky a segédigés cikket követő tanulmányaival, a határozókról, majd a melléknévi igenevekről, továbbá a Kálmán Lászlóval társszerzőségben konstruált „Fickó-Mackó nyelvtanról” (Kálmán & Prószéky, 1985) írott munkáival „valódi nyelvészként” is letette a névjegyét, s ezzel a háttérrel és biztonsággal indulhatott tovább a saját útján a magyar nyelvtechnológia alapjainak lerakása, majd megújítása felé.

⁶ Az újabb magyar mondattani kutatások tágabb kontextusáról részletes szakirodalmi utalásokkal egy másik, kevésbé személyes hangú cikkben írtam: Kenesei (2020).

⁷ Ezt azért merem állítani, mert Antal dossziéjában nincsen nyoma annak, hogy a „Mozgó Világ ügyben” információkat kellett volna gyűjtenie Prószékyéktől, viszont több más esetben is önként talált jelenteni való ügyeket.

Irodalom

- Abraham, W., & de Meij, S. (Eds.). (1986). *Topic, focus and configurationality*. Amsterdam: John Benjamins.
- Abraham, W., & de Meij, S. (Eds.). (1984). *GAGL: Groninger Arbeiten zur germanistischen Linguistik, No. 24–25*. Retrieved from <https://ugp.rug.nl/GAGL/issue/view/4113>; <https://ugp.rug.nl/GAGL/issue/view/4120>
- Kálmán L. & Prószék G. (1985). FMR Grammar. In *Műhelymunkák a nyelvészet és társtudományai köréből 1* (pp. 31–41).
- É. Kiss, K. (1981). Structural relations in Hungarian, a "free" word order language. *Linguistic Inquiry*, 12 (pp. 381–416).
- Horvath, J. (1981). *Aspects of Hungarian syntax and the theory of grammar* [Doctoral dissertation, University of California, Los Angeles].
- Kenesei I. (szerk.). (1984). *A nyelv és a nyelvek*. Budapest: Gondolat Kiadó.
- Kenesei, I. (Ed.). (1985). *Approaches to Hungarian 1: Data and Descriptions*. Szeged: JATE.
- Kenesei, I. (Ed.). (1987). *Approaches to Hungarian 2: Theories and Analyses*. Szeged: JATE.
- Kenesei I. (2020). A magyar mondatban 50 éve. *Magyar Nyelv*, 116(1) (pp. 1–18). http://www.c3.hu/~magyar-nyelv/20-1/Kenesei_MNy_20-1.pdf
- Kiefer F. (szerk.). (1992). *Strukturális magyar nyelvtan 1: Mondattan*. Budapest: Akadémiai Kiadó.
- Kiefer, F., & É. Kiss, K. (Eds.). (1994). *The syntactic structure of Hungarian*. New York: Academic Press.
- Kiefer F. & Telegdi Z. (szerk.). (1989). *Általános Nyelvészeti Tanulmányok XVII*. Budapest: Akadémiai Kiadó.

NYELVTECHNOLÓGIAI ESZKÖZÖKKEL (NEM) KEZELHETŐ HELYESÍRÁSI PROBLÉMÁK

LUDÁNYI ZSÓFIA | HUN-REN Nyelvtudományi Kutatóközpont

ludanyi.zsofia@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.9

1. BEVEZETÉS

Prószéky Gábort a Budapesti Műszaki és Gazdaságtudományi Egyetem másodéves mérnök informatikus (BSc) szakos hallgatójaként ismertem meg 2010 környékén. Adigra már megszereztem magyar szakos bölcsész és tanári diplomám, végzős hallgatója voltam az ELTE Nyelvtudományi Doktori Iskolájának, és azon gondolkoztam, hogyan tudnám a kétféle érdeklődési területemet, a helyesírást és az informatikát „összeházasítani”. Így ismerkedtem meg a nyelvtechnológia, a számítógépes nyelvészet területével, ahonnan egyenes út vezetett Prószéky Gáborhoz. Később, 2012-ben, utolsó éves műegyetemistaként hivatalosan is hallgatója lehettem, mivel két féléven keresztül a Pázmány Péter Katolikus Egyetem Információs Technológiai Karának vendéghallgatójaként jártam hozzá nyelvtechnológiai alapozó kurzusokra. Nagy megtiszteltetés és öröm volt számomra, hogy külsős témavezetőként segítette a BME-s önálló laborom, később BSc-s szakdolgozatom elkészítését.

Ahogy e rövid visszaemlékezés is mutatja, Prószéky Gábort születésnapja alkalmából köszöntő írásomhoz nem kellett sokáig keresnem a témát: hiszen az, hogy miként segíthetik a nyelvtechnológiai eszközök az írott szövegek helyesírásának az akadémiai normához való igazítását, az ünnepelt munkásságának egyik meghatározó területe (pl. Prószéky, 1998; Prószéky & Kis, 1999: 81–104). Köszöntő írásomban ehhez a témakörhöz kapcsolódva, a HUN-REN Nyelvtudományi Kutatóközpont nyelvi tanácsadással foglalkozó munkatársaként azt mutatom be, hogyan oszthatóak a magyar beszélőközösségben élő helyesírási problémák a gépi kezelhetőség szempontjából. A tanulmány címében szereplő kerek zárójel az értelmező szótári hagyományokhoz igazodva vagylagosságot jelent, tehát a szöveg a zárójeles rész hozzáolvasásával és anélkül is érvényes. Ebből adódóan a jelen írásban egyaránt foglalkozom azokkal a helyesírási problémákkal, amelyek kezelésében hatékony segítséget, megoldást jelentenek a nyelvtechnológiai eszközök, valamint azokkal is, amelyek különféle okokból kívül esnek a gépi kezelhetőség területén.

2. A NYELVI ÉS HELYESÍRÁSI PROBLÉMÁK KEZELÉSE

A tanulmány középpontjában – a nyelvmenedzselés-elméletre (Jernudd & Neustupný, 1987) támaszkodva – a nyelvi problémák állnak. Nyelvi problémáknak a beszélőből, a

beszédpartnerből vagy az adott interakciót kívülről szemlélő harmadik személyből valamilyen reflexiót kiváltó nyelvi jelenségeket tekintem (Ludányi et al., 2022: 76). Itt jegyzem meg, hogy a továbbiakban következetesen a nyelvi problémák *kezeléséről*, nem pedig *megoldásáról* fogok beszélni. Habár a *probléma* főnév igei kollokációi közül a *Mazzola* (Sass, 2009) adatai alapján a *megold* a leggyakoribb, Lanstyák (2018: 159–202) alapján a megoldást csak a problémakezelési módok egyik válfajának tekintem, hiszen a problémát például el is lehet kerülni vagy letagadni. Mindez fokozottan igaz azokra a problémákra, amelyekkel a nyelvhasználók az írott szövegek létrehozása és/vagy nyelvi ellenőrzése közben szembesülnek. Például ha egy korrektor nem tudja eldönteni, hosszú vagy rövid *i*-vel írja-e a helyesírási kodifikációval nem rendelkező *elh[ii]vatás*¹ szót, elkerülheti a nyelvi problémát azáltal, hogy kicseréli azt egy szinonimával, átfogalmazza a szöveget. Maga a helyesírási probléma – hogy ti. a szónak nincs helyesírási szótárban rögzített írásmódja – azonban továbbra is megoldatlan marad (Ludányi, 2023a: 40–41).

A helyesírási problémák kezelésének az alapvető célja tehát a mindenkori akadémiai helyesírási normához való igazodás. Ezt a célt szolgálja a nyelvtechnológiai eszközökkel támogatott helyesírás-ellenőrzés, illetve – lehetőség szerint a nyelvhasználói nézőpontokat, adott esetben a sajátos (szakmai) írásgyakorlatot is figyelembe véve – az emberi erővel végzett helyesírási tanácsadás is (l. 2.2 fejezet).

2.1 Gépesített helyesírás-ellenőrzés és helyesírási tanácsadás

A nyelvtechnológiai eszközökkel támogatott helyesírás-ellenőrzés megjelenése előtt a helyesírási problémák kezelése a helyesírási szabályzat és szótárak segítségével történt. A helyesírási problémák újabb, korszerű és egyszerű kezelési lehetőségei a szövegszerkesztőkbe épített helyesírás-ellenőrző programok, amelyek a 20. század utolsó évtizedében jelentek meg, majd terjedtek el (Prószéky, 1993). A kezdetleges helyesírás-ellenőrző eszközök használata azonban sok esetben újabb nyelvi problémákat generált a felhasználók számára, mivel csupán az ortográfiai szavak (Fóris, 2020: 20) szintjén tudták felismerni a normától eltérő alakokat, vagyis két betűköz vagy központozási jel között elhelyezkedő karaktorsorokkal dolgoztak. A mai környezetérzékeny helyesírás-ellenőrzők azonban statisztikai alapon többé-kevésbé képesek kiszűrni az efféle tévedéseket (Prószéky, 2019: 270–271).

A szövegszerkesztőkbe, később a böngészőkbe integrált programok mellett internetes weboldalak is segíthetik a nyelvhasználókat egyszerűbb helyesírási kérdéseik megválaszolásában. Ilyen például a *Helyesiras.mta.hu* tanácsadó portál. A HUN-REN Nyelvtudományi Kutatóközpont jogelődje, az MTA Nyelvtudományi Intézet által fejlesztett online helyesírási tanácsadó weboldal (Miháltz et al., 2013) hétféle helyesírási területen kínál gyors és automatikus választ: külön- és egybeírás, szószintű helyesírás-ellenőrző (*Helyes-e így?* eszköz), elválasztás, tulajdonnevek írása, számnevek helyesírása, keltezés, betűrendbe sorolás.

¹ A kurzívval szedett karaktorsorozat reguláris kifejezésként értelmezendő.

2.2 A helyesírási tanácsadó portál együttműködése a „humán” helyesírási tanácsadóval

A HUN-REN Nyelvtudományi Kutatóközpontban a jogelőd intézmény megalapítása óta működik nyelvi közönségszolgálat, amely e-mailben és telefonon válaszol az érkező helyesírási, nyelvhasználati kérdésekre (Ludányi et al., 2022). A több mint tíz éve működő *Helyesiras.mta.hu* portál tapasztalatai azt mutatják, hogy a helyesírási tanácsadás régi és új formái egymással folyamatos kölcsönhatásban vannak, a közönségszolgálat és a helyesírási portál egymást kiegészítve működik. Egyrészt azért, mert a felhasználói visszajelzéseket, javaslatokat idővel beépítjük a portál eszközeibe. Másrészt pedig azért, mert a helyesírási portál a felmerülő helyesírási problémáknak csak egy részét oldja meg. A tapasztalatok alapján a kérdezők egy része, mielőtt megkeresi a közönségszolgálatot, a helyesírási portál eszközeivel kísérli meg helyesírási problémája kezelését. Ha az automatikusan generált választ nem találják megfelelőnek, a közönségszolgálathoz fordulnak. A portállal kapcsolatos visszajelzések a következőképpen csoportosíthatók (Ludányi, 2023b):

- a) A portál által nyújtott javaslat megerősítésének kérése a közönségszolgálattól.
- b) A portál javaslatának elutasítása, egyet nem értés (ezen belül: egyet nem értés helyes válasz esetén, egyet nem értés hibás válasz esetén, a portál javaslatának részleges elfogadása, vagyis kiegészítési javaslat nyújtása az automatikus válaszhoz).
- c) Ellentmondás észlelése a különféle segédeszközök javaslataiban (a portál javaslata eltér a hagyományos, papíralapú szótárétól).
- d) Segítségkérés a közönségszolgálattól a portál által felajánlott többféle helyesírási javaslat közti választásban.
- e) A portál nem tudott segíteni, mert nem adott választ.
- f) A felhasználó a portálon nem a megfelelő eszközhöz fordult, ezért nem oldódott meg a nyelvi problémája.
- g) A felhasználó a portálon nem a megfelelő formátumban adta meg a bemenetet a specifikus, megszorított bemenetet váró eszközöknél, ezért nem oldódott meg a nyelvi problémája.
- h) Egyéb, a *Helyesiras.mta.hu* által előidézett egyéb nyelvi problémák esetén.

3. A HELYESÍRÁSI PROBLÉMÁK OSZTÁLYOZÁSA

A fentiek alapján az írott szövegek alkotása és/vagy nyelvi ellenőrzése során a nyelvhasználók a helyesírási problémáknak többféle típusával szembesülhetnek (Ludányi, 2023a). A nyelvtechnológiai kezelhetőség szempontjából (is) alapvető, hogy külön típusként kell számontartani a helyesírási kodifikáció által „lefedett” és „le nem fedett” helyesírási problémákat.

3.1 A helyesírási kodifikáció által lefedett helyesírási problémák

3.1.1 Egyszerű esetek

Kezelhetőségük szempontjából azokat a helyesírási problémákat tekintem egyszerűnek, amelyek megoldhatók az akadémiai helyesírási szabályzat, a helyesírási kézikönyvek és a szótárak segítségével. Ebben az értelemben tehát „egyszerű” eset az is, amikor maga a helyesírási szabály összetett (pl. szótagszámlálás szabálya), de egyértelmű útmutatást ad a nyelvhasználónak az írásmódot illetően. Azaz a nyelvhasználó a szabály elolvasása és értelmezése után le fogja tudni írni a problematikusnak ítélt alakulatot a mindenkori helyesírási normának megfelelően.

Azokban az esetekben, amikor a helyesírási szabályok leírhatók algoritmusokkal, és/vagy lexikális adatbázisokkal, kivételstákkal kezelhetők, a *Helyesiras.mta.hu* portál eszközei és a helyesírás-ellenőrző szoftverek is hatékony segítséget jelenthetnek a felhasználónak. A helyesírási portál szószintű helyesírás-ellenőrzője, a *Helyes-e így?*, valamint a tulajdonnevek (főként földrajzi nevek) helyesírási kérdéseiben segítő *Névkereső* eszköz használata lényegében a helyesírási szótárban való utánanézést váltja ki, a szótári keresés gyorsabb változatának feleltethető meg. A weboldal további, összetettebb működésű eszközei, mint például a környezetfüggetlen, attribútum-érték struktúrás formális nyelvtanon alapuló *Külön vagy egybe?* (Ludányi et al., 2013), valamint a *Számok* és a *Dátumok* eszközök pedig a helyesírási szabályok azonosítását és alkalmazását végzik el a nyelvhasználó helyett.

Léteznek azonban olyan területei is a magyar helyesírásnak, amelyeket a helyesírási portál eszközei nem képesek kezelni. Különösen a különírás-egybeírás területén fordulnak elő ilyen szabályok. A helyesírási portál „Hogyan működik?” menüpontja² részletesen tájékoztat ezekről a helyesírási területekről. Ilyenek például a mellérendelő szerkezetek, illetve összetételek helyesírásával kapcsolatban felmerülő külön- vagy egybeírás, illetve kötőjelezési problémák (pl. *testi-lelki*, *rüggkapál*, *napról napra*). A *Külön vagy egybe?* eszköz az esetekben sem tud segíteni, amikor egy többszörös szókapcsolatban két tag egyetlen egységként kapcsolódik egy harmadik elemhez (pl. *kutya-macska barátság*, *réz-arany ötvözet*). Ehhez ugyanis a programnak ismernie kellene az alkotó tagok közti szintaktikai és jelentéstani viszonyokat, ezzel a tudással azonban jelenleg nem rendelkezik. A helyesírási portál fejlesztés alatt lévő, 2.0-s verziójában (Prószéky & Várad, 2023) a mesterséges intelligenciát használó, a kontextus pontosabb felismerését lehetővé tévő interaktív technológia azonban várhatóan ki fogja tágítani a gépi módszerekkel kezelhető helyesírási problémák körét.

3.1.2 Problematikus esetek

Azt is látni kell azonban, hogy maga a helyesírási kodifikáció sem tökéletes: előfordulnak benne ellentmondások, következetlenségek. Ezeknek az eseteknek a gépi kezelése pontosan azért ütközik nehézségekbe, mert nem létezik rájuk egyértelmű megoldás.

² <https://helyesiras.mta.hu/helyesiras/default/howitworks> (2024. január 13.)

Ilyen például a *humán* szóval létrejött alakulatok írásmódja. A probléma oka az, hogy a szótárban található példák írásmódjából nem következtethető ki egyértelműen az az alapelv, hogy mi alapján íródik egybe, illetve külön a *humán* előtag az azt követő szóval.³ A szótári példákat áttekintve annyi látszik biztosnak, hogy a *humán* 'társadalomtudomány(okat tanító)' jelentésben különírandó az azt követő szótól: *humán tagozat, humán tudomány, humán gimnázium*. A *humánnak* létezik azonban az 'emberi, emberrel kapcsolatos' jelentése is, ezekben az esetekben már nem tűnik következetesnek a kodifikáció (részletesen I. Ludányi, 2023c). Kiindulva abból, hogy a gépesített helyesírási tanácsadás célja is a normaközvetítés, kérdésként merül fel, mit közvetítenek az eszközök abban az esetben, ha például a szótárban nem szereplő *humán ()faktor, humán()egészségügy* alakulatok lejegyzése okozza a helyesírási problémát. A *humán faktor* esete a jelentésviszonyokat tekintve inkább a helyesírási szótárban különírva rögzített *humán erőforrás, humán szféra* szerkezetekhez hasonló, a *humán + egészségügy* esetében a *humánpolitika* analógiája miatt elképzelhető az egybeírás, de a szótári kodifikáció ellentmondásossága miatt tulajdonképpen mindkét írásmód elfogadható: *humánegészségügy, humán egészségügy*. A *humán* szóval létrejött alakulatok írásmódja tehát azért nem írható le algoritmusokkal, mert maga a „szabály” nem egyértelmű. (Számos további ellentmondásos esetről I. Ludányi, 2023a: 37–40).

Az ellentmondás nemcsak magában a kodifikációban lehet, hanem a kodifikáció és egy kisebb-nagyobb közösség írásgyakorlata között is. Ilyen például a *kis()csoport* külön- vagy egybeírása. A *Magyar helyesírási szótár* (Tóth, 2017) csak óvodai kontextusban írta egybe a *kiscsoportot*, 'kis létszámú csoport' jelentésben a különírást javasolja. Az egyes szakmák írásgyakorlata (pl. szociológia, társadalomtudományok) azonban ellentmond ennek, mivel következetesen az egybeírást alkalmazza. Ez a gyakorlat indokolható is, mivel ezekben az esetekben nem pusztán kis létszámú csoportról van szó, hanem egy bizonyos, konkrétan meghatározott szám alatti csoportról. Kérdéses azonban, hogy miként járjon el egy helyesírási tanácsadó program ezekben az esetekben: szorítkozzon pusztán a kodifikált helyesírási norma közvetítésére, és 'kis létszámú csoport' jelentésben írassa külön (*kis csoport*), vagy pedig figyelembe véve a szakmai írásgyakorlatot és a jelentésváltozást, bírálja felül a kodifikált helyesírási normát, és írassa egybe (*kiscsoport*)?

3.2 A helyesírási kodifikáció által le nem fedett helyesírási problémák

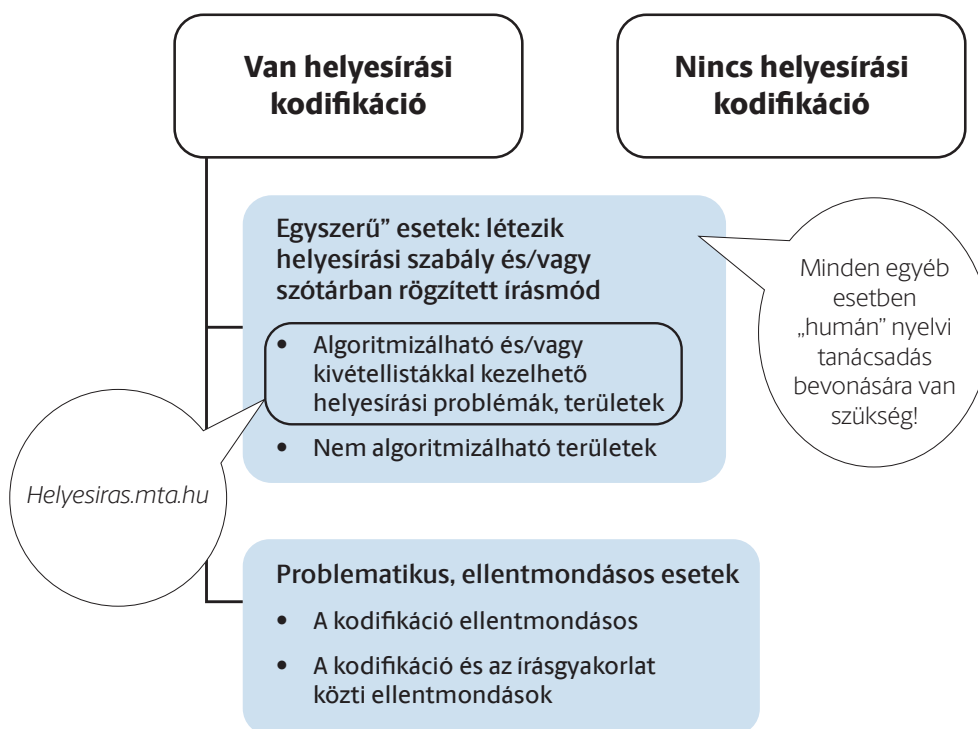
A helyesírási problémák gyakori forrása a helyesírási kodifikáció hiánya is, amely több szinten jelentkezhet. A kodifikáció hiánya vonatkozhat egyrészt egy-egy adott szóalakra (l. az említett *elh[ii]vatás*-t, valamint számos újonnan átvett kölcsönszót, pl. *stream* [sztrím?], *gender* [dzsender?]). Másrészt egész helyesírási területek, például azonos fogalomköri csoporthoz, kategóriához tartozó szavak helyesírása is „lefedetlen” lehet a kodifikáció szempontjából. Például nincs szabályozva az olyan újabb jelenségek körök

³ Nem minden esetben kérdéses azonban a *humán* egybe- vagy különírása az utána következő szóval. A 2.2 fejezet címében például egyértelműen különírandó a *humán*, mivel egy minőségjelzős szerkezet újabb minőségjelzője

helyesírása, mint a tanösvények, a számítógépes keretrendszerek stb. Ezekben az esetekben a helyesírási probléma gépi eszközökkel történő megoldása, vagyis a norma közvetítése azért nem lehetséges, mert – nincs mit közvetíteni. Emellett azonban a nyelvtechnológiai eszközök abban a tekintetben közvetve mégis a segítségünkre lehetnek, hogy a nagyméretű korpuszok, illetve a korpuszadatok kinyerésére szolgáló korpuszlekérdező eszközök (Sass, 2022) segítségével feltérképezhető az írásgyakorlat, így a közönségszolgálathoz (vagy akár egy majdan létrehozandó intelligens helyesírási segédeszközhöz) forduló kérdezőnek valós nyelvi adatokon alapuló helyesírási tanácsot lehet adni. Természetesen ez nem jelenti azt, hogy a kodifikáció által le nem fedett esetekben egyértelműen a kialakult írásgyakorlat szolgálja követése jelenti a megoldást, de a helyesírási problémák egy részénél, különösen az idegen nyelvekből kölcsönzött neologizmusok idegenes és magyaros írásmódja közti választás esetében hatékony segítség, sőt döntő jelentőségű lehet az úzus vizsgálata.

4. ÖSSZEZÉS

A magyar beszélőközösségben élő helyesírási problémák gépi kezelhetőség szempontjából (is) történő csoportosítását az 1. ábra foglalja össze.



1. ábra. A helyesírási problémák osztályozása

Jól látható, hogy a helyesírási problémáknak csak egy igen kis szelete kezelhető gépi eszközök segítségével. Az ebbe a szűk halmazba tartozó helyesírási problémák esetén valóban igen gyors és hatékony megoldást jelentenek a nyelvtechnológiai eszközök, ezt egyértelműen mutatják is a Kutatóközpont közönségszolgálatához érkező, a *Helyesiras.mta.hu* portállal kapcsolatos pozitív visszajelzések.

A helyesírási tanácsadás számítógépes nyelvészeti eszközökkel történő támogatása tehát mindenképpen fontos és szükséges, de önmagában nem elégséges megoldás a létező helyesírási problémák kezelésére. A jelen tanulmány is amellett érvel, hogy elengedhetetlen a gépesített helyesírási tanácsadás mellett annak hagyományos, „humán” formájának működtetése is. Nem csupán azért, mert a gépesített helyesírási tanácsadás a helyesírási problémáknak csak a kodifikált írásmóddal/szabályokkal rendelkező, algoritmusokkal leírható és/vagy kivétellistákkal kezelhető eseteiben képes segítséget nyújtani. A helyesírási problémák ezen részhalmazának ugyanis növelhető a mérete: a neurális hálók, a mesterséges intelligencia nyelvtechnológiai alkalmazásai minden bizonnyal jelentősen ki fogják bővíteni az automatikusan kezelhető helyesírási problémák körét. Továbbá egy mesterséges intelligencián alapuló csetbotalkalmazás várhatóan a felhasználókat is segíteni fogja akár a *Helyesiras.mta.hu* megfelelő eszközének használatában, akár abban, hogy a jelentéstől, a kontextustól függően ajánljon írásmódot a hozzá forduló felhasználónak. Annak azonban, hogy a kérdező az ellentmondásos, illetve szabályozatlan helyesírási esetekben is rugalmas, az ő (és gyakorlatközössége) nézőpontját, helyzetét figyelembe vevő választ kapjon, a tisztán gépi eszközökkel történő megvalósíthatósága erősen kétséges. A helyesírási tanácsadó portál jelenlegi és majdan megvalósítandó 2.0-s verziójának is alapvető feladata ugyanis a helyesírási norma közvetítése, így legyen bármilyen korszerű nyelvtechnológiai alapokra helyezve, akkor sem tud teret nyitni az akadémiai helyesírási szabályozást rugalmasabban, az egyes gyakorlatközösségek igényeihez igazítani szándékozó törekvéseknek.

Egy korszerű helyesírási tanácsadó segédeszköz fejlesztésekor azonban mindenképpen szükséges tisztázni olyan elvi kérdéseket, mint hogy:

- Felülbíráhatja-e egy helyesírási tanácsadó program a mindenkor kodifikált helyesírási normát?
- Helyesírási kodifikáció hiányában hozhat-e egy helyesírási-ellenőrző eszköz kodifikációs döntést? (A kérdés tulajdonképpen költői, hiszen a program válasza nagy valószínűséggel kodifikációs döntésként, „a helyes írásmódként” fog értelmeződni a beszélők körében, még ha nem is ez a fejlesztők szándéka.)

Mindemellett pedig nem lehet figyelmen kívül hagyni azt a tényt sem, hogy a nyelvhasználók is – még ha el is fogadják a gépi eszközök helyesírási javaslatait – alapvetően jobban bíznak a „humán” nyelvi tanácsadásban. Ahogy ezt az egyik levélíró is megfogalmazta a közönségszolgálat munkatársai által készített kérdőívben (Ludányi et al., 2022: 81–83): „Jó volt igazi emberektől, szakértőktől kapni direkt, egyenes és érthető választ”.

Ezekkel a tanulságokkal zárom Prószékly Gábornak szánt írásomat, akinek ezúton, 70. születésnapján köszöntve mondok köszönetet a nyelvi tanácsadás, a kutatóközpontbeli közönségszolgálat ügyeinek mindenkori támogatásáért.

Irodalom

- Fóris Á. (2020). *Lexikológiai és lexikográfiai ismeretek magyar (magyar mint idegen nyelv) tanároknak*. Budapest: L'Harmattan Kiadó, Károli Gáspár Református Egyetem.
- Jernudd, B. H., & Neustupný, J. V. (1987). Language planning: for whom? In Laforge, L. (Ed.), *Proceedings of the International Colloquium on Language Planning* (pp. 69–84). Québec: Les Press de L'Université Laval.
- Lanstyák I. (2018). *Nyelvalakítás és nyelvi problémák*. Somorja: Fórum Kisebbségkutató Intézet, Gramma Nyelvi Iroda.
- Ludányi Zs. (2023a). A helyesírási problémák kezelési stratégiái a nyelvi tanácsadásban. *Alkalmazott Nyelvészeti Közlemények*, 16(2) (pp. 29–46).
- Ludányi Zs. (2023b). *A nyelvi tanácsadási nemzetközi és hazai gyakorlatai, különös tekintettel a Nyelvtudományi Kutatóközpontra*. Habilitációs értekezés. Budapest: ELTE Nyelvtudományi Doktori Iskola.
- Ludányi Zs. (2023c). A nyelvi tanácsadó szolgálat postaládájából 14. *Amega*, 30(3) (pp. 36–37).
- Ludányi Zs., Miháلتz M. & Hussami P. (2013). Különírás-egybeírás – automatikusan. In Váradi T. (szerk.), *AlkNyelv-Dok7. Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből* (pp. 116–130). Budapest: MTA Nyelvtudományi Intézet.
- Ludányi Zs., Domonkos Á., Kocsis Á., & Jakab D. (2022). A nyelvi menedzselés szemlélete és a nyelvi tanácsadás. In Deme A. & Kuna Á. (szerk.), *Tanulmányok a nyelvészet alkalmazásainak területéről* (pp. 73–107). Budapest: ELTE Eötvös Kiadó.
- Miháلتz M., Hussami P., Ludányi Zs., Mittelholcz I., Nagy Á., Oravecz Cs., Pintér T., & Takács D. (2013). Helyesírás.hu – Nyelvtechnológiai megoldások automatikus helyesírási tanácsadó rendszerben. In Tanács A. & Vincze V. (szerk.), *MSZNY 2013. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 135–147). Szeged: Szegedi Tudományegyetem.
- Prószékly G., & Kis B. (1999). *Számítógéppel emberi nyelven. Intelligens szövegkezelés számítógéppel*. Bicske: SZAK Kiadó.
- Prószékly G. (1993). Nyelvművelés számítógéppel? (A számítógépes helyesírás-ellenőrzés új útjai). *Magyar Nyelvőr*, 117(4), 509–511.
- Prószékly G. (1998). Nyelvhelyesség és számítógépes szövegírás. *Magyar Tudomány*, 105 (Új folyam 43)(4), 491–493.
- Prószékly G. (2019). Nyelv, nyelvhasználat és a magyar helyesírás szabályainak 12. kiadása. In Bozsik G. & Ludányi Zs. (szerk.), *Szabályzat, oktatás, gyakorlat. Helyesírásról sokszínűen. A 2015-ös és a 2017-es Nagy J. Béla helyesírási verseny előadásai, feladatai és egyéb tanulmányok* (pp. 265–571). Eger: Linceum Kiadó.
- Prószékly G., & Váradi T. (2023). A magyar nyelv digitális fenntarthatóságának támogatása. *Magyar Nyelv*, 119(4), 482–489. <https://doi.org/10.18349/MagyarNyelv.2023.4.482>
- Sass B. (2009). „Mazsola” – eszköz a magyar igék bővítményszerkezetének vizsgálatára. In Váradi T. (szerk.), *Válogatás az I. Alkalmazott Nyelvészeti Doktoranduszkonferencia előadásaiból* (pp. 117–129). Budapest: MTA Nyelvtudományi Intézet.
- Sass, B. (2022). Principles of corpus querying: A discussion note. *Acta Linguistica Academica*, 69(4), 599–614. <https://doi.org/10.1556/2062.2022.00581>
- Tóth E. (szerk.). (2017). *Magyar helyesírási szótár*. Budapest: Akadémiai Kiadó.

ETTÉL MÁR SÜLT HERMELINT HRANOLKIVAL? A Termini-szótárról dióhéjban¹

M. PINTÉR TIBOR, P. MÁRKUS KATALIN

Károli Gáspár Református Egyetem, Bölcsészeti- és Társadalomtudományi Kar
m.pinter.tibor@kre.hu | p.markus.kata@kre.hu | DOI: 10.18135/PG70.2024.10

Prószéky Gáborral először egy majdnem tízórás vonatút során találkoztam. Szepsziszentgyörgyre utaztunk (talán még emlékszik is rá), ahol előadást tartott. Később, egyik első munkahelyem épp a MorphoLogic Kft. lett (ti. MPT-nek), amely inspiráló volt – a mai napig szeretettel gondolok arra az időszakra és Gábor mindenkori közvetlenségére, vidámságára, világlátására. Lexikológiai, lexikográfiai előadásaiban mindig kitért arra, hogy a kis nyelvek szótárait online kellene működtetni (ehhez képest a MorphoLogic Kft. végezte el angol–magyar kétnyelvű szótárain XML-esítését és online térbe helyezését – PMK). Hát, az évek múlnak, mi pedig épp egy online szótár munkálataiban is részt veszünk. Egy szótárban, amelyet épp a Nyelvtudományi Kutatóközpont hoztolt. Álljon itt egy rövid értekezés arról, hogy is gondoljuk a magyar–magyar online szótárat – bizonyosságként a magyar nyelv többközpontúságára.

Bevezetés

A Termini magyar–magyar szótár és adatbázis (termini.nytud.hu/htonline/) a külső régiók magyar nyelvhasználatát dokumentálja, így egyedülálló szótártípusnak számít a hazai szótárpalettán. A magyar nyelv leírása szempontjából különösen fontos, hiszen a szótárba azok a magyarországiaktól eltérő, idegen eredetű szavak kerülnek, amelyeket a Kárpát-medencében, hazánk határain túl élő magyarok használnak. A szótár hét, Magyarországot körülvevő ország magyarlakta régióinak szavait tartalmazza: Erdély, Felvidék, Vajdaság, Kárpátalja, Horvátország, Muravidék, Örvidék. A gyűjtemény 2007 óta online is elérhető. A szótár szerkesztőinek elsődleges célja, hogy lehetővé tegyék más magyar nyelvváltozatok sajátos szavainak és szójelentéseinek (pl. *hermelin*, *hranolki*, *kofola*, *nanuk*) megismerését, másrészt, hogy az említett régiók beszélői is megismerhessék e szavak közmagyar megfelelőit (ha vannak ilyenek). A Termini-szótár „nyitott” szótárként működik, a közösség is beleszólhat a szerkesztésébe: ha az olvasók új szócikket, esetleg meglévő szócikkhez új jelentést szeretnének javasolni, írhatnak a szerkesztőknek. Mindegyik címszó használatának szemléltetésére lehetőség szerint élőnyelvi példamondatot, szövegrészletet tartalmaz (a forrás pontos megjelölésével), lehetővé téve, hogy a szó vagy szerkezet használati sajátosságait grammatikai és pragmatikai kontextusban vizsgáljuk. Tanulmá-

¹ Jelen tanulmány alapjául az alábbi írás szolgált: Benő et al., 2021.

nyunkban a szótár újszerűségeit, valamint érdekességeit szeretnénk részletesebben áttekinteni.

A Termini-szótár és adatbázis a maga nemében sajátos vállalkozás a magyar lexicográfiában. Sajátos a lefedett nyelvváltozatok tekintetében, sajátos a mikrostruktúra kidolgozottságában, illetve sajátos a készítés folyamata tekintetében is, mivel olyan, nyelvész szakemberek által szerkesztett, adatbázis-alapú szótárról van szó, amely elsődlegesen az online megjelenésre van paraméterezve.

Informatikai háttér

A szótár kizárólag online felületen létezik, egy időben nyolc országból szerkesztik. A változtatások azonnal látszanak a találati listában, amely értelmezhető akár egy papírszótár oldalainak is. A többfelhasználós platform alapja egy relációs adatbázis, melyhez a bemeneti (szerkesztőfelület) és kimeneti (lekérdezőfelület) oldal tartozik. Ez a technológiai megoldás, amelyet a mai lexicográfiában már széles körben alkalmaznak, a Termini-szótár készítésének kezdeti fázisában még újdonságnak számított.

Mind a relációs adatbázis, mind az online felület lehetőséget biztosít statisztikák készítésére és információk kinyerésére, illetve az online megjelenés folyamatos lehetőséget biztosít képek és hangfájlok közzétételéhez és azok folyamatos revíziójához. A szótár mögötti adatbázis nyelvészeti, elsősorban lexicográfiai szempontból rendkívül gazdag szerkezetű – igyekszik minél árnyaltabb képet adni a címszó mögötti lexikális és konceptuális tartalmakról. Az egyes lekérdezések folyamán lesz adatból információ, amely lexicográfiai, lexikológiai vagy adatbányászati szempontból is érdekes.

A technológia meghatározza a szótár szerkesztését: a papíralapú szótárak lineáris szerkesztési folyamatát felváltotta a technológiailag fejlettebb, adatbázis-alapú munkafolyamat, ahol a szerkesztés folyamán létrehozott változások azonnal láthatók (a szerkesztés folyamatáról a szerzők e-mailben értesíthetők), azonban a szerkesztés folyamán használt felület (nézet) jelentősen eltér a véglegesen megjelenített szócikkek vizualizációjától.

Az osztott adatbázis lehetővé teszi, hogy a szócikkeket helytől függetlenül lehessen szerkeszteni. A szerkesztőfelület nem jelent nagy kihívást a használók és a szerkesztők számára sem grafikai, sem technikai, mivel könnyen és egyszerűen kezelhető. A szerkesztők jogosultak bármelyik szócikket szerkeszteni vagy a szócikkekhez (csak a szerkesztők számára hozzáférhető) megjegyzést fűzni.

Interaktív szótár

A szótár nem pusztán lexikai gyűjtemény, hanem a nyelvhasználat, nyelvváltozatok, lexikai és morfológiai adatbázisa – adatbányászatra, valamint a tárolt tartalom különféle formában, különböző rendezési elvek alapján történő megjelenítésre is alkalmas (akár szótárformában is).

A háttérben működő adatbázisból különféle lehetséges utakat bejárva a felhasználó a lekérdezésnek megfelelő struktúrájú szócikkeket kap. A lexikai adatot és információt

keresve három felület közül választhat, így juthat el a saját maga által keresett információt tartalmazó „saját szótár” megjelenítéséig. A honlap nyilvános felülete – felhasználóbarát módon – csupán az alapfunkciókat felkináló *Keresés* mezőt tartalmazza. A szótár elsősorban nyelvészek által használt része a *Szójegyzékek* menü, mely *lexikográfiai adatbányászként* működik. A kereső/lekérdező a szócikkstruktúra bármelyik elemében kereshet.

A *szójegyzékek generálása* menüben többféle szójegyzék is előállítható, melyek létrehozását négy alaptípusba sorolt kategóriaegyüttes teszi lehetővé: így a szerkesztés időszakára vonatkozó szócikkek (*Időszak*), a szótárral foglalkozó nyelvészek lexikográfiai kutatásaihoz szükséges szócikkek (*Kutatási listák*), a szótári mikrostruktúra bizonyos részeit érintő és a lexikográfiai kutatásokat segítő szócikkek (*Fejlesztési listák*), valamint a fejlesztők munkásságát vagy az aktuális állapotát tükröző listák (*Dokumentációs listák*) előállítása egyaránt lehetséges. A listák és szócikkek megjelenítése a szócikk mikrostruktúrája és metaadatai mentén paraméterezhető.

A nézetek sajátos típusa a *Megjegyzések* rovat/elem, amelybe a szerkesztés folyamatán a szócikkekhez megjegyzések fűzhetők, a „félkész” állapotban lévő szócikkek pedig *rejtett* szócikké alakíthatók, amelyek így csak a szerkesztők számára jelennek meg, a regisztráció nélküli használók pedig nem láthatják őket.

A lexikográfiai, lexikológiai munkák pontosítására, valamint a határon túli magyar nyelvváltozatok elemeinek behatóbb vizsgálatára különféle paraméterekkel rendelkező listák is szolgálnak. A kvantitatív kutatások adatigényét kiszolgáló listák a szójegyzékekhez hasonlóan igény szerint alakíthatóak, segítve vagy adatokkal ellátva a szerkesztőket.

A különféle nézetek adta lehetőségnek köszönhetően az olvasó vagy szerkesztő csak azokat az információkat látja, amelyekhez hozzá szeretne férni.

A Termini-szótár az egyetlen olyan komplex szótár, mely speciálisan a magyar nyelv határon túli változatainak sajátos lexikális elemeit, jelentéseit tartalmazza.

A szótár multimédiás elemei

Napjainkban a digitális szótárak egyre növekvő népszerűsége vitathatatlan. Legnagyobb előnyük a könnyű és gyors információhozzáférés, amely mindennapi életünkben bárhol, bármikor lehetséges. Az elektronikus tárolás ezen túl a multimédiás eszközök használatának széles skáláját is lehetővé teszi, így a szótárhasználók nem kizárólag a papírszótárban megszokott, „hagyományos” szócikkeket olvashatnak, hanem lehetőség nyílik vizuális és akusztikus megoldások alkalmazására is. A képjelentésleírást támogató elemként történő megjelenítése a szótártörténetben már hagyományosnak tekinthető, gondoljunk csak a képes szótárakra. Ha csak a magyar szótárak történetben tekintünk vissza, akkor több száz évvel korábban, már 1653-ban megtalálhatjuk Comenius képes „szótárát”, amely a vizuális lehetőségek alkalmazásának előfutára, valamint megemlíthetjük még az Értelmező kéziszótár első kiadását is, amely egyes címszavakat a definíción túl képek segítségével magyarázott. A képek a

szótárban enciklopédikus tartalomnak minősülnek: nem a nyelvi jelet, hanem a világot láttatják az olvasóval, segítségükkel a jelentések pontosíthatók, egyfajta megerősítést adnak az olvasónak, valamint a szöveges definícióknál sokkal több, a címszóval kapcsolatos tipikus jellemzőt tudnak közvetíteni.

A Termini-szótár képei, amelyek a szócikkekben az egyes jelentésekhez külön-külön is beilleszthetők, legtöbbször főneveket, konkrét tárgyakat jelölnek. Az illusztrációk alatt a forrás és a képaláírás is olvasható, amelyek leíró jellegük miatt egyfajta magyarázatként is szolgálnak. Ezt láthatjuk a *blokk* szócikk '(lakótelepi) panelház' jelentésénél, ahol a következő képaláírás áll: „Jellegzetes szocializmus-korabeli blokk Szereden”.

Hupka (1989) felosztása alapján a vizuális elemek többféle típusba sorolhatók. A Termini-szótár három fő típust alkalmaz ezek közül. Az első csoportba azon képek tartoznak, amelyekben kizárólag a tárgy jelenik meg: igyekezni kell a legtipikusabb elem(ek)et kiválasztani. Jó példa erre a típusra a Fv, Ka *baklászán* 'padlizsán' címszó illusztrációja, hiszen itt a zöltség formájának, alakjának és színének a megjelenése elegendő az értelmezéshez; de az élet egyéb területein is találunk hétköznapi tárgyakat: *kalendár* 'naptár', *bandaszka* 'benzin, olaj szállítására használt marmonkanna; víz szállítására használt műanyag kanna', *batri* 'szárazelem', ahol hasonlóan nem volt szükség tágabb környezetre a jelentés bemutatásához. A szótárban találunk még márkaneveket is, amelyek szintén ide sorolhatók, hiszen a vizuális elemek márkához kapcsolódó termékeket ábrázolnak, így a képek a termékeket egyszerűen be tudják mutatni (pl. *ticsinki* 'ropi', *chrumki* 'mogyorós pufi', ahol a termékek csomagolása vagy kinézete önmagáért beszél).

A következő csoportban már a tárgyat környezetében kell bemutatni, ez a típus rengeteg plusz információval gazdagíthatja az olvasót. Megemlíthetjük például a *bábovka* 'kuglóf' szócikket, ahol a sütemény izlésesen tálalva jelenik meg az olvasó előtt; a *ko-pacski* 'stoplis futballcipő' képén már a futball-labda és a futballpálya is látható, tehát a tágabb környezet segít az értelmezésben; a *dodávka* 'furgon' esetében a járművet egy raktárbejárat előtt, a kocsibeállón láthatjuk, illusztrálva ezzel haszongépjármű-jellegét.

A harmadik csoportba tartozó képek a tárgyat használat közben mutatják be, hiszen az értelmezést itt a működés, a funkció bemutatása illusztrálja a legjobban. Szemléltetésül megnézhetjük a *kolobeska* 'roller' képén megjelenő rollerező fiatal vagy a *bublifuk* 'szappanbuborék-fújó gyermekjáték' alatt szereplő gyereket, aki éppen buborékot fúj, de a *bazén* 'medence' képe is pontosan a használatot és a tárgy funkcióját illusztrálja.

Esetenként melléknevekhez, igékhez is lehet képet csatolni, ilyenek például az *autoritativ* 'tekintélyelvű, tekintélyt parancsoló', a *degusztál* 'ételt, italt (meg)kóstol' szócikkek, ahol a melléknevet mint emberi tulajdonságot egy kifejező mozdulattal, valamint szemléletes környezettel lehet ábrázolni; az ige pedig a cselekvés végzése közben bemutatható.

A vizuális tartalom mellett a szótárban hanganyagok is megjelennek: mivel a példamondatok egy része előszóban is elhangzott. A hanganyagból viszonylag kevés van, mivel ezen terület új fejlesztésnek számít.

Összegzés

A Termini magyar–magyar szótár és adatbázis egyedülálló vállalkozás a magyar nyelv lexikográfiai feldolgozásában. Ezt látszik alátámasztani a szócikkszerkezet összetettsége, információgazdagsága, vagy akár a szövegek kontextusban idézett példamondatok eredetének jelzése. Olyan interaktív online szótár, ahol a szócikk adatstruktúráját a lekérdező a saját „információéhségéhez” tudja alakítani. Jobban megnézve még az is megtudható, hogyan lehet hranolkit (sült krumpli, hasáburgonya) és hermelint (szlovákiai gyártású camembert) enni.

Irodalom

- Benő A., Lanstyák I., P. Márkus K. & M. Pintér T. (2021). A Termini magyar–magyar szótár és adatbázis szerkezeti és szerkesztési sajátosságai. *Magyar Nyelvőr*, 145(4) (pp. 378–389).
- Hupka, W. (1989). *Wort und Bild: Die Illustrationen in Wörterbüchern und Enzyklopädien*. Tübingen: Max Niemeyer Verlag (Lexicographica, Series Maior 22).

A SZÓTÁR ÖSSZEKÖT

MURÁTH JUDIT¹, SZÖLLŐSY ÉVA²

¹PTE KTK Egyetemi Könyvtár és Tudásközpont,

Terminológiai Dokumentációs Központ | ²PhD, lexikográfus

murath.ferencne@ktk.pte.hu | szeva1@freemail.hu | DOI: 10.18135/PG70.2024.11

1. Bevezetés

Egyedül nem megy. Ezt mindenki tudja, aki egy-egy szótár írásával, megjelentetésével és forgalmazásával valahogy kapcsolatba került. A szakszótár meg végképp sokembe-eres műfaj. Aki szakszótárírássra adja a fejét, legyen az lexikográfus vagy egy-egy szakterület specialistája, azért teszi, mert munkájával az érdekeltek szakmai ismereteinek és nyelvi kifejezőképességének gyarapítását kívánja szolgálni. Akkor lesz jó és használható a szótárunk, ha használói széles körének igényeivel tisztában vagyunk. A szótár tehát összeköt. Összeköti továbbá a szakembereket és lexikográfusokat, de összeköti azokat is, akik az évszázadok alatt megszületett művek adatait szeretnék egybegyűjteni, és végül bennünket is az Ünnepelettel. A mi ismeretségünk és barátságunk is ezer szállal kapcsolódik a szótárhoz. Ha nincs ez a barátság, egész biztosan nem jön létre az a közös munka, amelynek eredményeként eddig közel száz számítástechnikai/informatikai szótárt és lexikont sikerült felkutatni, adatait rögzíteni és közreadni adattárunkban. Ebbe a folyamatba kívánunk most e jeles alkalomból első ízben rövid betekintést nyújtani.

A PTE Egyetemi Könyvtár és Tudásközpont Terminológiai Dokumentációs Központjában épülő elektronikus adatbázis, amely a magyarral kapcsolt szakszótárak és -lexikonok bibliográfiáját kívánja kutatók és érdeklődők számára rendelkezésre bocsátani, széles körű összefogással készül. A projekt célja az egynyelvű (magyar), a kétnyelvű és a többnyelvű szakszótárak és -lexikonok lehetőleg teljes körű bibliográfiájának elkészítése a 16. századtól napjainkig. Hosszú kutató- és adatrögzítő munkálatok után 2019-ben megnyitottuk az adatbázist, és azóta minden érdeklődő számára hozzáférhetőek a „Publikált” kategóriában lévő adatlapok. Az eddigi kutatások és gyűjtőmunka eredményeként pillanatnyilag 25 szakterületről 1006 adatlap érhető el (<https://szakszotar.lib.pte.hu>).

2. A számítástechnikai/informatikai szótárak és lexikonok adatbázisunkban

Az adatbázisba felvett, magyar nyelvű, valamint magyarral kapcsolt számítástechnikai/informatikai szótárak és lexikonok kutatója Szöllősy Éva, mellette még Czékmán Orsolya, Seidl-Pécs Olivía, Szamosmenti Marianne, Muráth Judit és Farkas Flóra. Az 1963 és 2020 között megjelent munkák rendkívül érdekes, tarka képet mutatnak, és a bennük dokumentált tartalmakkal alátámasztják Prószéky Gábor azon megállapítását, hogy „...a magyar nyelvközösségnek az interkulturális kommunikáció megváltozása tekintetében nagy esélye a számítógép és az elektronikus kommunikáció megjelenése” (Prószéky, 2017: 332). E nagyon új, az élet minden területét behálózó szakterület szókin-

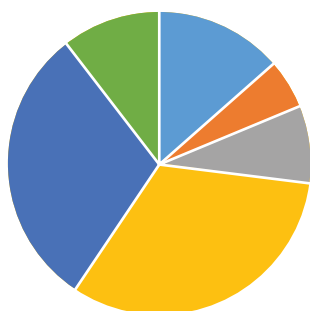
csének fejlődését őrző dokumentumok bemutatása nem egyszerű feladat. A kutatás és dokumentálás közben a kérdések sokaságával találtuk magunkat szembe, amelyek közül itt és most hármat szeretnénk kiemelni.

Mindjárt az adatgyűjtés elején felmerült a szakterületi besorolás kérdése. Hosszú ideig a műszaki tudományok és a matematika is magáénak vallotta ezt az új szakterületet (Czékmán, 2020: 237; Seidl-Pécs, 2020: 280). Ugyanez elmondható a statisztikáról is (Hegedűs, 1977). Lassan azonban rendeződött a helyzet. Időközben, pontosabban 1995-ben már az ETO-ba is bekerült önálló szakcsoportként, először „004 Számítógép-tudomány, Számítástechnika” besorolással és elnevezéssel (Barátné Hajdu, 2007: 31).

A következő kérdés, hogy milyen szerzőséggel, hogyan készültek e szótárak és lexikonok. A korai munkák között több szótár nemzetközi együttműködésben készült. Kiváló példa erre a vállalatirányítási számítógép-alkalmazás fogalmainak kétkötetes terminológiai szótára, amely bolgár, német (NDK), cseh, szovjet és magyar (Szófia, Drezda, Prága, Moszkva, Minszk, Budapest) részvétellel készült. A magyar partner a budapesti Rendszertechnikai Vállalat volt. Az egyik kötet 851 fogalom egységes, egynyelvű értelmező szótára, a másik pedig a kifejezések ötnyelvű (magyar, német, angol, orosz, esperantó) szótára. A két kötet kiadását számos hazai és külföldi irodalomkutatás és konzultáció előzte meg. A magyarországi kiadás egyik szerkesztője Síklaky István közgazdász és informatikus volt (Síklaky, 1972a, 1972b). Hasonlóképpen nemzetközi összefogással készültek az 1970-es években a könyvtárügy, a kartográfia, a statisztika és a számítástechnika területén készített szótárak (5 db).

Gyakran alkalmazott eljárás egy-egy idegen nyelvű munka, azaz elsősorban egy-egy angol vagy német nyelvterületen megjelent szótár, képes szótár vagy lexikon fordítása. 1963 és 2005 között 13 ilyen munka került a magyar olvasók kezébe, az egynyelvű értelmező szótártól a képes szótáron át a 11 nyelvű szakszótárig. A következő csoportba a magyar szerzők többnyelvű szótárai tartoznak, ezek között vannak a határon túli magyar iskoláknak készült szótárak is (8 db). A két legnépesebb csoport magyar szerzők angol–magyar szótárai (31 db), valamint ugyancsak magyar szerzők magyar nyelvű munkái (29 db). Említést érdemelnek még a kifejezetten iskolai célra megírt szótárak (10 db). Összesen 82 szótár (négy munkának két csoportban is helye van).

Milyen szerzőséggel, hogyan készült?



A harmadik kérdés, amelynek megválaszolása még a továbbiak során is fejtörést okoz, a munkák műfaj szerinti besorolása. Egyrészt a címek alapján nem mindig lehet besorolni egy-egy adott szaklexikográfiai kategóriába bizonyos műveket, ezért jó gyakorlatnak bizonyult, hogy minden művet kézbe veszünk. Másrészt az értelmező szótárak és a lexikonok közötti határok gyakran elmosódnak, csakúgy, mint az értelmező szótárak és a kétnyelvű szótárak közötti határok. Egy-egy adott mű szerzői ugyanis praktikumra törekednek, és azokkal az információkkal töltik meg munkájukat, amelyek tapasztalataik szerint a legjobb helykihasználással segítik az adott szakterületen tájékozódni kívánó olvasót.

3. Kitekintés

A számítástechnikai/informatikai szótárak és lexikonok rögzítése, az adataik pontosítása adatbázisunkban lassan a végéhez közeledik. A nyomtatott szótárak és lexikonok mellett fontos további feladatunk a még fel nem fedezett, elektronikusan tárolt szókincsgyűjtemények felkutatása, és azok adataival a meglévő gyűjteményünk kiegészítése.

Irodalom

- Barátné Hajdu Á. (2007). A magyar ETO-kiadás munkálatai. Hazai és nemzetközi tapasztalatok. *Könyv, Könyvtár, Könyvtáros*, 16(2) (pp. 20–39). <http://epa.oszk.hu/01300/01367/00122/pdf/02muhelykerdesek.pdf> (2024. 02. 13).
- Czékman O. (2020). A magyar matematikai szótárak a 20. században. In Muráth J. (szerk.), *Magyar szaklexikográfia* (pp. 225–251). Budapest: Tinta Könyvkiadó.
- Prószéky G. (2017). A számítógép, az elektronikus kommunikáció és az internet hatása. In Tolcsvay Nagy G. (szerk.), *A magyar nyelv jelene és jövője* (pp. 321–335). Budapest: Gondolat Kiadó.
- Seidl-Pécs O. (2020). A műszaki szaklexikográfia Magyarországon. In Muráth J. (szerk.), *Magyar szaklexikográfia* (pp. 273–301). Budapest: Tinta Könyvkiadó.
- Síklaky I. & Ruisz R. (szerk.). (1972a). *A vállalatirányítási számítógép-alkalmazás fogalmainak értelmező szótára*. Budapest: Számítástechnikai Oktató Központ. Main collaborator: Faragó Sándor.
- Síklaky I. & Ruisz R. (szerk.). (1972b). *A vállalatirányítási számítógép-alkalmazás fogalmainak többnyelvű szótára: magyar–német–angol–oros–eszperantó*. Budapest: Számítástechnikai Oktató Központ. Main collaborator: Faragó Sándor.

SOSEMVOLT CIGÁNY SZÓTÁR: a latin–cigány szójegyzéktől a lehetséges romani spacy-ig¹

ROSENBERG MÁTYÁS | PhD-hallgató

matyas.rosenberg@gmail.com | DOI: 10.18135/PG70.2024.12

Bevezetés

A romani nyelvvel szemben minden korban komoly elvárásokat támasztanak. Az elvárásokat az emberek alapvetően olyan nyelvek alapján fogalmazzák meg, amelyek több beszélővel bírnak, és amelyeknek használói nyelvi, társadalmi és gazdasági szempontból eltérő helyzetben vannak a romanit beszélőkhöz képest, ezért a saját nyelvükre is másképp tekintenek, annak más funkciókat tulajdonítanak. Ezek az elvárások gyakran olyan nyelvek mintájára épülnek, amelyek széles körű beszélői bázissal rendelkeznek, továbbá jelentős kulturális és gazdasági erőforrást képviselnek, mint amilyen a magyar, a román vagy a szlovák nyelv. Az ilyen nyelvek elitjei gyakran olyan ambiciózus célokat tűznek ki nyelvük fejlesztésére, mint például a minőségi szótárak készítése, inspirálódva más nagyobb nyelvek hasonló törekvéseiből, ezzel is emelve nyelvük presztizsét. Ez az elit aztán általában a kisebb nyelvek felé fordul és megpróbálja azt is „felzárkóztatni” – jelentsen a kisebb nyelv ez esetben bármit is, ami miatt kevésbé vetül rá figyelem.

A romani nyelvű elit nem így tesz: más, kisebbségben élő népek iskolázottabb tagjaihoz hasonlóan elsősorban nem a saját igényeiket fogalmazzák meg, hanem a többségi társadalom a romani nyelvvel és általában a cigányokkal szemben támasztott igényeinek kívánnak megfelelni, még ha adott esetben interiorizálják is azokat. Ezek az elvárások pedig rendre nem teljesíthetők, jellemzően olyan tényezők miatt, amelyeket – hegemon szemszögből – régóta cipelt hiányosságnak is tekinthetnénk. Írásom amellett kíván érvelni, hogy a romani nyelv helyzete, bár számos kihívást rejt, nem tekinthető reménytelennek. Az előttünk álló akadályok ellenére léteznek lehetőségek és stratégiák, amelyek révén a romani nyelv és annak beszélői megerősíthetik pozícióikat a társadalomban, elősegítve ezzel nyelvük megőrzését és fejlődését.

¹ Intertextuális utalás Bartos (1958) által szerkesztett *Sosemvolt cigányország – Szegkovács cigány történetek* c. gyűjteményére, melyben megjelenik az önrendelkezésért küzdő, állam nélküli nemzet fogalma, valamint Szuhay (2012) *Sosemlesz Cigányország* c. művére, amely bemutatja, hogy a magát cigánynak/romának mondó vagy a mások által cigánynak mondott csoportok hogyan szerkesztik meg saját cigány/roma vagy a cigánytól elhatárolódó képüket, vagyis azt, amelyet magukra érvényesnek tartanak, illetve hogyan próbálják érvénytelené tenni azt a képet, amelyet akár mások fogalmaztak meg róluk, akár korábban maguk fogalmaztak meg önmagukról.

A lemaradás

A magyarországi írott romani források rövid történeti áttekintése során egyfajta állandó lemaradást észlelhetünk. A legkorábbi ismert dokumentum, a Vistai Farkas (1797–1798) által összeállított latin–cigány–magyar szójegyzék keletkezési ideje sem biztosan tudható, de a kutatók egyetértenek abban, hogy a szöveget ténylegesen lejegyző személy – akinek kiléte máig bizonytalan – nem rendelkezett megfelelő romani nyelvtudással. A következő két évszázad folyamán – Sztojka (1886) 90%-ban tükörfordításokat tartalmazó munkáját leszámítva – nem jelent meg anyanyelvi szerző által írt nyelvi témájú munka, és az egyetlen bizonyosan elkészült szótár sem maradt fenn. A többségi társadalom politikai és vallási irányzataihoz és mintáihoz való igazodás, valamint a restancia ledolgozása tükröződik abban, amikor magas presztízsű (nek vélt) szövegek fordításai látnak napvilágot, így a cigány nyelvű Kommunista kiáltvány (Choli, 1975), amely 16 évvel előzte meg Máté evangéliumának romani fordítását (Choli, 1991), és amelyet hamar a „Szentírás cigány nyelvre történő lefordításaként” értelmeztek, annak ellenére, hogy a teljes bibliafordítás csak 2008-ban készült el (Vesho-Farkas, 2008), és amellyel szintén komoly szakmai problémák merülnek fel.

A 2000-es évek óta jelentős növekedés mutatkozik a romani szótárak kiadásában, így az elmúlt kb. 230 év során összesen 49 önálló művet tartunk számon. Az első időszak szótárai inkább amatőr gyűjtemények, olyan kuriózumok, melyeket szigorú tudományos kritériumok alapján nehéz értékelni: kevés, sokszor egyetlen adatközlőre támaszkodnak, ortográfiájuk inkonzisztens, és szerkesztőik gyakran nem rendelkeznek megfelelő romani nyelvtudással. Előnyük, hogy több nyelvjárást is lefednek és deskriptív jellegűek, a magyarizmusok módszeres kipurgálásán túl nem alakították a szóanyagot. Az 1967–1984 közötti rövid időszak jelenti Magyarországon a romani nyelv tudományos csúcspontját: képzett szakemberek kezdenek a romani nyelvvel foglalkozni, akik szisztematikusabb és tudományos alapokon nyugvó terepmunkát végeznek, ezáltal megbízhatóbb adatok látnak napvilágot, még ha a megvalósítás minősége olykor kívánivalót is hagyott maga után.

1984-től a leíró szemléletet a preskriptív váltja fel, ahol a szótárak neologizmusok bevezetésével próbálják a romani nyelvet a többi európai nyelvhez közelíteni, ezzel bővítve használatának lehetőségeit. Ebben az időszakban már az oláh cigány dialektusok dominálnak, és bár a korábbi kutatásokra támaszkodnak, azok feldolgozása közel sem megfelelő. Az utóbbi két évtizedben a szanszkrit bőrbé bújtt hindi kölcsönzések beáramlása a szótárak minőségének további csökkenését eredményezte. A korábbi időszak terepmunkájával ellentétben a szótárírók nem végeznek gyűjtő tevékenységet, nem rendelkeznek elegendő szakmai tudással és az alapvető romani ismereteik is megkérdőjelezhetők.

A nehézségek és az igények

Az elmúlt évszázadok során nem jött létre olyan romani szótár, amely legalább az alapvető igényeket kielégítené, ami számos jól ismert tényezőre vezethető vissza: a

romani nyelvváltozatok heterogenitása, a cigányság transznacionális jellege, az ortográfiai dilemmák, a nyelvi asszimiláció, az identitásbeli különbségek közösségenként és beszélőnként egyaránt, a romani beszélőkkel szembeni diszkrimináció és alacsony iskolázottságuk stb. mind hozzájárulnak ehhez. Az adatgyűjtést megnehezíti továbbá a nyelv titkos funkciója miatti bizalmatlanság (ld. *te na dasla avri le gaženge* 'ne adjuk ki [a nyelvet] a gázsóknak'), mely az anyanyelvűt egyébként épp úgy sújtja, mint a közösségen kívüli kutatót. A minőségi szótárakhoz elengedhetetlen az alapos terepmunka és a megfelelő korpusz, a korábbi hiányosságok – többek között az adatközlők kiválasztása, a lejegyzői kompetencia, valamint lejegyzési és nyomdai hibák – számos problémát vetnek fel. A jó minőségű korpusz és a nyelvújítás végsősorban elengedhetetlen, de ezek a feladatok, bár kihívást jelentenek, nem reménytelenek, mindössze sosem kezdődtek el, az elenyésző számú próbálkozással pedig komoly szakmai kifogások merültek fel.

A szükségletek a romani nyelvű oktatási, kulturális és más anyagok tekintetében igen sokrétűek, mégis hiányoznak az alapvető eszközök, mint például általános iskolai és gimnáziumi tankönyvek, szótárak, nyelvtanok és nyelvtankönyvek, sőt alapvetően nincsenek romani nyelvű könyvek. Az 1987 és 2010 közötti időszakot kivéve romani nyelvű újságot nem adtak ki, hírportálok pedig ezt követően sem állnak rendelkezésre, és a romani nyelvű filmek vagy filmfeliratok is hiányoznak. Ez különösen visszas, hiszen az oktatásban egyre több bilingvis diák tanul egyre tovább, akik minél tovább maradnak az oktatási rendszerben, annál valószínűbb, hogy monolingvis magyar beszélővé válnak. Emellett számos olyan újságíró ismert, aki képes lenne romani nyelven publikálni, ám nincsenek meg a szükséges platformok vagy források, hogy munkájukat piacképessé tegyék.

A lehetőségek

A romani nyelv egy pozitívuma, hogy számos beszélője van, ha a kutatások számára sokszor láthatatlanok is. Tapasztalataink szerint a romani nyelv titkos használata is addig tart, míg fel nem ismerik, hogy több száz éve léteznek szabadon hozzáférhető, írott romani források, illetve, hogy jó promptolás esetén akár még a ChatGPT 4 is rávehető a cigány nyelvű kommunikációra és fordításra – még ha igen kezdetleges és változatos módon is teszi azt. Az elmúlt 11 év terepmunkája során számos helyszínen megfigyeltem, hogy az ott-tartózkodásom végére mind a több száz településen több hasznót láttak a nyelvről való nyílt diskurzusban, mint a nyelv elrejtésében, így ez biztosan nem jelenthet komoly és tartós problémát, hiszen minél több pozitív visszacsatolás érkezik, annál kevesebb a negatív prekonceptió a többségi társadalommal szemben és fordítva.

A gyűjtési és adatfeldolgozási technikák jelentős fejlődésen mentek keresztül. Már nem szükséges súlyos fonográfot cipelni, hogy aztán a nagy melegben megolvadjon a viaszhenger, és minden tönkremenjen, mint a Csenki testvérek esetében. A táska méretű audiokazettás diktafonok és a vállra vehető VHS kamerák ideje is letűnt, a mai eszközök kiváló minőségű adatgyűjtést tesznek lehetővé. Ráadásul az utóbbi időben,

részen a COVID–19-járvány következtében egyes cigány közösségek maguk is megosztanak bárki számára elérhető videókat online platformokon, amelyek sokszor többórnyi értékes nyelvi anyagot tartalmaznak.

A digitális korban, bár sok eszköz áll rendelkezésre, nem minden megoldás elérhető automatikusan. A nyelvtechnológia, mint például a szövegfeldolgozás és a gépi fordítás, kulcsfontosságú a kommunikációhoz és az információhoz való hozzáférésben. A tokenizálás az egyik alapvető lépés az NLP-ben. Bőséges forrásokkal rendelkező nyelveknél ez a folyamat viszonylag egyszerű, míg a kevesebb írott forrással bíró nyelveknél ez komoly kihívást jelent. Azonban az első lépések megvalósítása elérhető:

greeting = "Baxtalo rakhadjimasko djes le eftavardeše beršenge Prószečky Gáboreske!"

word_tokenize(greeting)

['Baxtalo', 'rakhadjimasko', 'djes', 'le', 'eftavardeše', 'beršenge', 'Prószečky', 'Gáboreske', '!']

Az ezt követő lépések, így például a PoS tagging a romani nyelv esetében még nem megoldott, mivel a géptanulás-alapú modellek és algoritmusok hatékony működése nagy mennyiségű adatra épül, ami a romanihoz hasonló kevésbé dokumentált nyelveknél akadályt jelent (Bird, 2006), ráadásul a romani nyelvű közösségek főként szóbeli hagyományokon alapulnak, ami korlátozza az írott anyagok mennyiségét és a nyelvi regiszterek változatosságát (Grenoble & Whaley, 2006). Mindez megnehezíti a nyelvmodellhez, beszédfelismeréshez, gépi fordításhoz vagy hangulatanalízishez szükséges, következetes írásbeli formák meghatározását, a nyelvi és dialektális diverzitás pedig további kihívásokat jelent (Crystal, 2000). Megjegyzendő, hogy a romani nem rendelkezik olyan összetett morfológiai szerkezetekkel, mint például a magyar vagy a török, mégis összetettebb, mint az angol, ami többletmunkát jelent modellalkotás során (Comrie, 1989). Mindezek ellenére a jövőben nagy valószínűséggel egy lehetséges AI fordítás eredményét is olvashatjuk majd:

['Boldog', 'hetvenedik', 'születésnapot', 'Prószečky', 'Gábornak!']

Következtetések

Az NLP a kevésbé dokumentált, kisebb nyelvek számára jelentős kihívásokat állít a nyelvtechnológia terén, azonban ezek leküzdése hozzájárulhat a nyelvi sokszínűség digitális korszakbeli megőrzéséhez és fejlesztéséhez. Ez azt követeli meg, hogy a romani nyelv esetében is kifejlesszünk adaptált NLP-eszközöket, hasonlóan a magyar nyelvre szabott HunPoS, HuSpaCy nyelvelemző eszközökhöz. Fontos lenne továbbá a már létező nyelvmodellek (pl. PULI) továbbtanítása/adaptálása a romani nyelvre. Ezen új eszközök hozzájárulhatnának a modern és részletesebb romani nyelvű adatbázisok és korpuszok előállításához. Emellett az AI-fordítás során felmerülő lexikai hiányok azono-

sítása elősegítheti a nyelvújítást, és a közösségek bevonása a kutatásba aktivizálhatja őket, hangsúlyozva a tanulás fontosságát. Ezáltal a romani nyelv vélt vagy valós „leamaradása” ha nem is válna egy csapásra „ledolgozhatóvá”, de nagy lépést tehetnénk abba az irányba.

Irodalom

- Bartos T. (szerk.). (1958). *Sosemvolt cigányország – Szegkovács cigány történetek*. Budapest: Európa Könyvkiadó.
- Bird, S. (2006). NLTK: The Natural Language Toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions* (pp. 69–72).
- Choli Daróczi, J. (1975). *Le kommunishtenge partosko cipipe*. Rom Som.
- Choli Daróczi, J. (1991). *Nyevo teshtamento. Amare Rajesko le Jesusesko Nyevo Jekkethanipe. Le shtar evandyeliiumura*. Budapest: Romano Kulturalno haj sittyimasko Jekhipe.
- Comrie, B. (1989). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.
- Crystal, D. (2000). *Language Death*. Cambridge: Cambridge University Press.
- Grenoble, L. A., & Whaley, L. J. (2006). *Saving Languages: An Introduction to Language Revitalization*. Cambridge: Cambridge University Press.
- Sztojka (Nagyidai) F. (1886). *Ő császári és magyar királyi fensége József főherceg magyar és cigány nyelvgyök-szótára. Románé álvá. Iskolai és utazási használatra*. Kalocsa: Malatin Nyomda.
- Szuhay P. (2012). *Sosemlesz cigányország*. Budapest: Osiris.
- Vesho-Farkas, Z. (2008). *Biblia – Dulmutano thaj nyevoteshtamenticko suntoiskiripe*. Budapest: Szent Jeromos Katolikus Bibliatársulat.
- Vistai Farkas M. (1797–1798). *Cigány–magyar szótár*. (Kézirat).

KERESÉS KORPUSZBAN 3: BRÜSSZELRŐL JELENTJÜK

Prószéky Gábornak, 70. születésnapjára

SASS BÁLINT | HUN-REN Nyelvtudományi Kutatóközpont
sass.balint@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.13

Kedves Gábor! Többek között a Te rendszeres felkéréseid nyomán jó pár éve tartok előadásokat és publikálok (Sass, 2017, 2022) arról, hogy szerintem hogyan lehet nyelvészeti kutatás céljára korpuszból megfelelően kinyerni a szükséges adatokat, és ezekre építve hogyan lehet egy szép és módszertanilag helyes korpuszvezérelt kutatást elvégezni. Kerek születésnapod alkalmából arra gondoltam, hogy itt az ideje, hogy megtudjuk, hogy nekem magamnak ez hogyan megy. :) *Isten éltesse sokáig!*

Inspirált ennek a tanulmánynak a megírására az is, hogy nemrég vált hozzáférhetővé a Magyar Országgyűlésben 2014 és 2023 között elhangzott beszédek és hozzászólásokat tartalmazó ParlaMint 4.0 korpusz (Ligeti-Nagy et al., 2023). Ennek a korpusznak a gazdag annotációs rétegei nagyon sokféle kutatásra és összevetésre adnak lehetőséget. Számomra a legizgalmasabbnak a kormányoldali és az ellenzéki kommunikáció jellegzetességeinek vizsgálata és összehasonlítása tűnt.

Egy aspektust, vagyis igazából egyetlen szót választottam ki, és ez az utóbbi években sokak által nagyon sokszor emlegetett, sok ráarakódott jelentésréteggel bíró városnév: a *Brüsszel*. Hogyan jelenik meg a kormányoldali és az ellenzéki beszédben ez a szó, hogyan és mire használja a két oldal, mik a hasonlóságok és mik a különbségek a két oldal vonatkozásában?

A kutatási hipotéziseim felállítása után az adatkezelés technikai részéről szólok, itt bemutatom az általam elkészített és a jelen kutatás elvégzéséhez is használt új, általános célú adatfeldolgozó eszközt; majd a vizsgálatok, az eredmények és a tanulságok bemutatása következik.

1. Hipotézisek

A vizsgálandó hipotéziseimet a sajtóban megjelentek, a közbeszéd és saját benyomásaim alapján fogalmazom meg.

H₁: A kormányoldal számára inkább aktív *Brüsszel*: egy olyan ágens, aminek (akinek?) szándékai vannak, csinál valamit; az ellenzék számára inkább passzív szereplő: egy város, ahol valami történik.

H₂: A kormányoldal kommunikációs paneleket, előregyártott frázisokat használ, amiket aztán sokan és sokszor ismételnék. Az ellenzék nem vagy sokkal kevésbé él ezzel az eszközzel.

H₃: A kormányoldal számára *Brüsszel* inkább negatív, az ellenzék számára inkább pozitív. A kormányoldal kommunikációjában *Brüsszel* általában ellenünk tesz valamit; az ellenzék kommunikációjában pedig inkább értünk.

2. Adatfeldolgozás: csql

A ParlaMint 4.0 korpusz szabadon elérhető a NoSketchEngine (NoSkE) korpuszkezelő rendszerben (Rychlý, 2007) a https://www.clarin.si/ske/#concordance?corpname=parlamint40_hu címen. A felület a lekérdezés támogatásán túl (pl. CQL, szűrés) számos hasznos kiegészítő funkciót biztosít (pl. gyakoriságlista-készítés, kollokációkeresés), mégis gyakran az a leginkább célravezető, ha az adatokat letöltés után táblázatos formában saját magunk kezelhetjük és dolgozhatjuk fel egyéni igényeink szerint.

Ilyenkor viszont sok esetben abba a nehézségbe ütközünk, hogy a NoSkE letöltési lehetőségei meglehetősen korlátozottak. A felkínált négyféle fájlformátum lényegében ugyanazt az adathalmazt tartalmazza, és ugyanazokat a problémákat rejti. Most a **.txt** formátumon mutatom ezt be. A konkordancia találatok sorozata, egy találat négy mezőből áll: egy fejlécmező metaadatokkal, valamint pontosan három tartalmi mező: bal kontextus, találat, jobb kontextus. A tartalmi mezők *szóközzel* elválasztva tartalmazzák az egyes tokeneket, egy tokennek a reprezentációja pedig a hozzá tartozó lekérdezett jellemzők értékeit *perjellel* (/) elválasztva. Ebből egyenesen következik a probléma: gondot okoz, ha az adattartalomban szóköz vagy perjel szerepel, és ez természetesen elő is fordul. A dizájn hibája az, hogy nem sikerült extrémális elemet találni a reprezentáció formai részének megjelenítésére. Nyilvánvaló megoldás lehetne az XML használata, de ezt a NoSkE nem valósítja meg, az **.xml**-ként hirdetett letöltési formátumában is a most leírt szóköz-perjel reprezentáció jelenik meg az XML mélyén.

A létrehozott **csql** nevű eszköz egyik funkciója éppen az, hogy kezeli a fenti problémás formátumot, azaz egyértelműen beazonosíthatóvá teszi az egyes tokeneket és az egyes jellemzőket (1. ábra). A szóközők kezeléséhez szükséges az eszköz rendelkezésére bocsájtani egy olyan, a NoSkE-ből letöltött kiegészítő fájlt, amely ugyanazt a konkordanciát tartalmazza, amit vizsgálni akarunk, de kizárólag a szóalakok vannak benne megjelenítve; a perjelek kezeléséhez pedig paraméterben meg kell adnunk, hogy melyek azok a mezők, amelyek perjeleket tartalmazhatnak.

- a) `nyolc/nyolc/NUM/Case=Nom NumType=Card Number=Sing`
`Brüsszel/percben/Brüsszel/perc/NOUN/Case=Ine`
`Number=Sing előadni/előad/VERB/VerbForm=Inf Voice=Act`
- b) `['nyolc', 'nyolc', 'NUM', 'Case=Nom++NumType=Card++Number=Sing']`
`['Brüsszel/percben', 'Brüsszel/perc', 'NOUN', 'Case=Ine++Number=Sing']`
`['előadni', 'előad', 'VERB', 'VerbForm=Inf++Voice=Act']`

1. ábra. Példa a szóköz-perjel reprezentáció hátrányaira. Az ábra a) részén a *nyolc Brüsszel/percben előadni* szöveg reprezentációja látható 4 jellemző (szóalak, szótő, szófaj, morfológia) megjelenítésével. Látjuk, hogy szóalakban előfordul perjel, morfológiában pedig szóköz. Az ábra b) részén látjuk a csql által kitisztázott viszonyokat.

A **csql** másik funkciója az, hogy lehetővé teszi, hogy a meglévő három tartalmi mezőt tetszőleges továbbiakkal egészítsük ki. Azaz szabad kezdet ad abban, hogy az adathalmazunk milyen (származtatott) jellemzőit szeretnénk vizsgálni, kiegészítőleg meghatározhatjuk, hogy mi az, ami számunkra releváns az adott kutatás szempontjából. Ezt egy Pythonban implementált függvény segítségével valósíthatjuk meg. Feljegyezhetjük külön mezőbe adott token adott jellemzőjét: például a jobb kontextus első szavának szófaját, vagy a találat utolsó szóalakját. Szabadon származtathatunk további jellemzőket a meglévőkből. Külön mezőt szentelhetünk például annak, hogy a találati szó nagybetűs-e, hogy a találat hány tokenből áll, mi a jobbra eső legközelebbi ige, hány főnév jelenik meg a kontextusban vagy a ParlaMint-korpuszokban meglévő függőségi annotációt felhasználva akár azt, hogy hány testvérbővítménye van egy adott bővítménynek. A könnyű visszakereshetőséghez hasznos a sorszám, ezért azt alapértelmezetten mindig hozzávesszük a fejlémező metaadataihoz.

Az eszköz szabadon hozzáférhető a <https://github.com/sassbalint/csql> címen. Az alábbi vizsgálatok során a NoSkE felület funkcióit és a **csql** eszköz lehetőségeit egymást kiegészítve használtam.

3. Adathalmaz

A lekérdezés, amellyel a vizsgálati adatokhoz hozzájutottam, nagyon kézenfekvő volt: a

brüsszel

kereső kifejezést használtam egyszerű keresésben a ParlaMint 4.0 korpusz https://www.clarin.si/ske/#concordance?corpname=parlamint40_hu címen található keresőfelületén. Azaz minden olyan szóalakot lekértem, amely bármilyen módon tartalmazza a **brüsszel** karaktersorozatát, ez mind a vizsgálat tárgyát képezi most. Ez a lekérdezés minden szóba jöhető adatot teljes körűen lefed (Sass, 2022: principle #2). A metaadatok közé a beszélőt (**speech.speaker_id**) és a kormányoldal-ellenzék státuszt (**speech.date speech.party_status**) vettem be; a jellemzők közül a szóalakot, a szótövet, a szófajt, a morfológiai jegyeket, és az adott tokenhez tartozó függőségi fej szótövet jelenítettem meg.

Biztam benne, hogy jelentős mennyiségű adatot kapok. A 6853 találat a korpusz 36,3 millió szavas méretét tekintve – azaz a 189 találat/millió szó (röviden: 189/M) gyakoriság – talán nem tűnik soknak, de a nagybetűs szótöví szavak listájának elejét átnézve azt látjuk, hogy a helyek között Magyarország, az Európai Unió, Európa és Budapest után az ötödik helyen szerepel Brüsszel, azaz a városok között a második leggyakoribb. Néhány jellegzetes példa az 1. táblázatban látható.

azonosító	szóalak	előfordulási szám
1	<i>brüsszeli</i>	2101
2	<i>Brüsszel</i>	1654
3	<i>Brüsszelben</i>	1492
4	<i>Brüsszelből</i>	309
...		
5	<i>brüsszeleznek</i>	7
6	<i>Brüsszel-Strasbourg-Soros-féle</i>	6
7	<i>Brüsszel-ellenes</i>	5
8	<i>Brüsszel-Strasbourg-Soros-triumvirátus</i>	4
9	<i>brüsszelitáknál</i>	4
...		
10	<i>brüsszelitábbak</i>	3
...		
11	<i>Brüsszel/percben</i>	1
...		
12	<i>brüsszeliörület-vonat</i>	1

1. táblázat. Példák a *brüsszel* egyszerű lekérdezésre kapott adathalmazból. A 6853 elemű találati listában 79 különféle szóalak fordul elő. Látjuk, hogy a szónak nagyon kiterjedt bokra alakult ki: a ragozott alakok (3, 4) mellett megjelennek a különféle összetételek (6, 8, 12) és a képzett alakok, mint például a *brüsszelez* (5) vagy a *brüsszelita* (9, 10), sőt egészen speciális képződmények is, amilyen a *Brüsszel/perc* (11) mértékegység. A három leggyakoribb alak együtt az összes előfordulás több mint háromnegyedét adja.

4. Vizsgálatok és eredmények

Vizsgálataim célja a kormányoldal és az ellenzék *Brüsszel*-használatának összevetése. E két kategória mellett a korpuszban létezik egy „nem besorolt” kategória is. Ezt a továbbiakban nem veszem figyelembe, bár ránézésre ebben a csoportban is a kormányoldali kötődésű emberek (Dömötör Csaba, Orbán Balázs, Schanda Tamás stb.) vannak túlsúlyban.

név	#
Rétvári Bence	532
Dömötör Csaba	306
Nacsa Lőrinc	230
Orbán Viktor	215
Vejkey Imre	192

2. táblázat. A *Brüsszelt* legtöbbször használó megszólalók. Mindannyian kormányoldaliak.

A 6853 találat eloszlása: 3727 (54,4%) kormányoldal, 1799 (26,2%) ellenzék (és 1327 [19,4%] nem besorolt). A kormányoldali szövegmenyiség 13,3 millió szó, az ellenzéki 16,9 millió szó, eszerint a *brüsszeles* kifejezések gyakorisága rendre 279/M, illetve 107/M. A kormányoldalon tehát majdnem háromszoros ezen kifejezések gyakorisága. A megszólalók gyakorisági listájából is jól látszik a kormányoldali túlsúly, az első 5 helyezett mindegyike a Fidesz–KDNP-hez kötődik, ahogy ez a 2. táblázatból látszik.

A konkordancián lefuttatott kollokációvizsgálatban megjelennek a jól ismert szavak: *bürokrata, szankció, elhibázott, szankciós, politika, Soros, képvisel, bevándorláspárti, háború, elit*. A főnevek legtöbbször a *brüsszeli* jelzővel szerepelnek, *Soros* általában a *Brüsszellel* koordinációban jelenik meg, a *háború* a *brüsszeli szankciók* kísérője, a *képvisel* igével pedig legtöbbször a *Brüsszelt* vagy *Brüsszel érdekeit képviseli* szókapcsolatban találkozunk.

A *Brüsszelt* középen tartalmazó 5-gramok közül 53 előfordulással magasan a leggyakoribb az *elhibázott brüsszeli szankciós politika*. Az 5-gramok között *elhibázottat* és *szankciót* is bárhol tartalmazó találat 142 darab van, ez önmagában az összes előfordulás 2%-a (!), azaz minden ötvenedik *Brüsszel* ebben a frázisban jelenik meg. A kifejezés eloszlása szélsőséges: 116 db (8,7/M) kormányoldal vs. 8 db (0,5/M) ellenzék.

Az aktív/passzív szerep megragadására a legprimitívebb megközelítés egyszerűen annak vizsgálata, hogy a *Brüsszel* milyen raggal milyen gyakori. A ragtalan alanyi előfordulás ágéntív használatra utalhat, a tárgyi vagy helyhatározói előfordulás pedig páciensire. Az eredményeket a 3. táblázat mutatja be.

alak	kormányoldal	ellenzék	kormányoldal%	ellenzék%	%-arány
<i>Brüsszel</i>	910	426	24,4%	23,7%	~1
<i>Brüsszelben</i>	825	378	22,1%	21,0%	~1
<i>Brüsszelbe</i>	103	85	2,8%	4,7%	1,7
<i>Brüsszelt</i>	82	56	2,2%	3,1%	1,4
összesen	3727	1799			

3. táblázat. Különböző ragozott alakok gyakorisága. Ezen az absztrakt szinten inkább hasonlóságokat találunk. A *-ba/-be* és a *-t* mutat eltérést.

Mit mutatnak meg a szélsőségesen az egyik oldalhoz kötődő kifejezések? Ebben a vizsgálatban a *Brüsszel* és a hozzá tartozó függőségi fej alkotta kétszavas egységeket vizsgáltam **csql** és további elemzés segítségével. Először összegyűjtöttem ezeket a kétszavas egységeket, majd lekérdeztem a gyakoriságukat a kormányoldali, valamint az ellenzéki alkorpuszban, végül a kettő aránya szerint sorba rendeztem őket. Így a lista elején a kormányoldalra, a végén pedig az ellenzékre jellemző elemeket kaptam meg (4. táblázat).

sorszám	kifejezés	kormány	ellenzék	K/E	E/K
1	<i>brüsszeli elit</i>	40	–	∞	0,0
2	<i>Brüsszelből kap</i>	31	–	∞	0,0
3	<i>brüsszeli intézmény</i>	21	–	∞	0,0
4	<i>brüsszeli politikus</i>	30	2	15,0	0,1
5	<i>Brüsszelben dolgozik</i>	22	2	11,0	0,1
6	<i>brüsszeli politika</i>	128	12	10,7	0,1
7	<i>brüsszeli vezető</i>	19	2	9,5	0,1
8	<i>brüsszeli javaslat</i>	18	2	9,0	0,1
9	<i>Brüsszel tesz</i>	19	3	6,3	0,2
10	<i>brüsszeli döntéshozó</i>	22	4	5,5	0,2
11	<i>brüsszeli Bizottság</i>	31	6	5,2	0,2
12	<i>brüsszeli szankció</i>	139	27	5,1	0,2
13	<i>brüsszeli bürokrácia</i>	33	7	4,7	0,2
14	<i>brüsszeli bürokrata</i>	117	26	4,5	0,2
15	<i>brüsszeli döntés</i>	26	6	4,3	0,2
16	<i>brüsszeli terv</i>	13	3	4,3	0,2
17	<i>Brüsszelben van</i>	34	10	3,4	0,3
18	<i>Brüsszel akar</i>	10	3	3,3	0,3
19	<i>Brüsszelben képvisel</i>	39	12	3,3	0,3
20	<i>Brüsszelben tesz</i>	17	6	2,8	0,4
21	<i>Brüsszelből érkezik</i>	13	5	2,6	0,4
22	<i>Brüsszelbe küld</i>	13	5	2,6	0,4
23	<i>brüsszeli forrás</i>	18	8	2,3	0,4
24	<i>Brüsszelben mond</i>	16	10	1,6	0,6
25	<i>Brüsszelben dönt</i>	11	7	1,6	0,6
26	<i>Brüsszel ad</i>	12	9	1,3	0,8
27	<i>Brüsszel van</i>	10	8	1,3	0,8
28	<i>Brüsszelbe megy</i>	12	10	1,2	0,8
29	<i>Brüsszelnek ad</i>	9	8	1,1	0,9
30	<i>Brüsszel mond</i>	13	12	1,1	0,9
31	<i>Brüsszelben megszavaz</i>	10	11	0,9	1,1
32	<i>Brüsszelben aláír</i>	8	10	0,8	1,3
33	<i>Brüsszeltől vár</i>	7	13	0,5	1,9
34	<i>Brüsszelbe kimegy</i>	5	20	0,3	4,0
35	<i>Brüsszelt megállít</i>	4	17	0,2	4,3
36	<i>brüsszeli pénz</i>	4	24	0,2	6,0
37	<i>Brüsszelre mutogat</i>	–	31	0,0	∞

4. táblázat. A célszó és a hozzá tartozó függőségi fej alkotta (jelző+főnév vagy bővítmény+ige) kifejezések elemzése. A táblázat a legalább 20-szor előforduló kifejezéseket tartalmazza, a kormányoldalra jellemzőek fent, az ellenzékre jellemzőek lent helyezkednek el.

Az alábbiakban a zárójeles számokkal a 4. táblázat példaira utalok. A kormányoldalra jellemző kifejezések legtöbbször erős negatív kontextusban jelennek meg. A *brüsszeli elit* (1) *korrupt és bevándorláspárti*, „*lábbon lőtte az európai gazdaságot*” és „*figyelmén kívül hagyta az állampolgárok döntését*”. A *brüsszeli intézmények* (3) „*példa nélküli lépésre készülnek*”, „*ránk akarnak erőltetni*” valamit és „*kétarcú politikát folytatnak*”. A *brüsszeli politikusok* (4) *baloldaliak*, „*nyíltan bevándorláspártiak*” és „*vádolnak minket*”. Még a *Brüsszelben dolgozik* (5) is szinte mindig úgy jelenik meg a korpuszban, hogy a magyar ellenzék Magyarország érdeke ellen dolgozik Brüsszelben.

Az ellenzékre jellemző kifejezések viszont inkább semleges, érzelemmentes kontextusban fordulnak elő (*Brüsszelben aláír* [32], *Brüsszelbe kimegy* [34]), vagy a Fidesz–KDNP-re utalnak, a kormányoldali frázisokat idézik (*Állítsuk meg Brüsszelt!* [35]). A *Brüsszelre mutogat* viszont inkább pozitívan állítja be *Brüsszelt*, mint aki nem szolgált rá arra, hogy a bajokért őt hibáztassák. Nem véletlen, hogy ebben a kontextusban a *mutogat* csak harmadik személyben fordul elő, és a kormányoldal sohasem használja.

A kormányoldal jelentős számban használ olyan kifejezéseket, amelyek *Brüsszelnek* erőteljes szándékot tulajdonítanak, egyúttal kifejezik az ezzel a szándékkal való szembenállást, és ezt jellemzően érzelmileg fűtött szavakkal teszik. *Brüsszel* például „*a háború árát az európai emberekkel fizetteti meg*”, vagy „*nem tesz semmit*”, vagy „*szankciós lépéseket tesz*”, vagy „*hiába akar valamit elérni*”, vagy „*új adókat akar ránk kényszeríteni*”, vagy „*kényszerbetelepítést akar*”, de mi „*nem fogjuk beengedni őket, akárhogy is akarja Brüsszel*”.

5. Összefoglalás és tanulságok

Ebben a részben áttekintjük, hogy mit mondhatunk az 1. részben megfogalmazott hipotézisekről a 4. részben elvégzett vizsgálatok alapján.

A H_1 hipotézisünk az volt, hogy *Brüsszel* a kormányoldal számára inkább aktív, szándékokkal bíró cselekvő ágens; az ellenzék számára pedig inkább passzív szereplő. A 3. táblázatban bemutatott egyszerű vizsgálat nem mutatta ezt ki, annyi látszott, hogy a passzívabb *-ba/-be* ragos, illetve tárgyragos forma kicsivel jellemzőbb az ellenzék megnyilvánulásaira. Amint láttuk, a kormányoldal sok olyan kifejezést használ, ami *Brüsszelnek* szándékot tulajdonít, ráadásul olyan szándékot, ami a kormányoldal szándékaival szemben áll, ami ellen „harcolni kell”, és ez okot ad az érzelmmel telített megfogalmazások használatára. Azt mondhatjuk, hogy a H_1 kormányoldalra vonatkozó része igaznak bizonyult, és úgy tűnik, hogy a kormányoldal–ellenzék különbség nem az aktív–passzív tengelyen, hanem inkább az érzelmi telítettség skálán jelentkezik.

A H_2 hipotézisünk az volt, hogy a kormányoldal hajlamos kommunikációs paneleket, előregyártott frázisokat használni, és ezeket számos alkalommal ismétli; az ellenzék nem él ezzel az eszközzel, viszont gyakran utal a kormányoldal paneleire kritikus hangon. Láttuk, hogy az *elhibázott*, a *brüsszeli* és a *szankció* szavak különféle kombinációiból képzett panel kiemelten gyakori a korpuszban: önmagában az előfordulások 2%-át adja, és 20-szor gyakoribb a kormányoldali alkorpuszban, mint az ellenzéki-

ben. Az ellenzéki megszólalók minden esetben a Fidesz–KDNP-re utalva gunyorosan és/vagy számonkérően ejtik ki a kifejezést (például: „*hát, egyetlenegy magyar ember szavazta meg ezeket az elhibázott brüsszeli szankciókat*”). Az egyetlen kivétel Toroczkai László, aki komolyan használja, egy alkalommal. A kifejezést 39 (!) különböző kormányoldali megszólaló használja, így az a sejtés is igaz, hogy nemcsak sokszor, hanem sokan is ismétlik ezt a panelt. Azt az ellenzéki oldalon népszerű vélekedést tehát, hogy a kormányoldal előre megfontolt szándékkal és összehangoltan folytatja ezt a fajta kommunikációt, a nyelvi vizsgálatunk alátámasztja. Az *Állítsuk meg Brüsszelt!* panel gyakrabban jelenik meg az ellenzéki kommunikációban a Magyar Országgyűlésben, ez összefügghet azzal, hogy a kormányoldal viszont plakátokon hirdette ezt az üzenetet az egész országban.

A **H₃** hipotézisünk szerint *Brüsszel* a kormányoldal számára inkább negatív, az ellenzék számára pedig inkább pozitív szereplő; a kormányoldal kommunikációjában *Brüsszel* általában ellenünk tesz valamit, az ellenzék kommunikációjában pedig inkább a mi érdekünkből. Fentebb láttuk, hogy a kormányoldalra jellemző kifejezések nagyon gyakran erős negatív kontextusban jelennek meg, és szinte mindig Magyarország-ellenes, a magyarok érdekeivel ellentétes élt kapnak. Az ellenzékre jellemző kifejezések ugyanakkor többnyire semlegesek, néhány kifejezés – mint például a *Brüsszelre mutogat* – pedig ha közvetlenül is, de határozottan pozitív szereplőként hivatkozik *Brüsszelre*. A másik oldal *Brüsszel*hez való viszonyának percepciója megjelenik olyan egyedi alakulatokban is, mint az *Orbán–Brüsszel-háború*. A **H₃** esetében is igaz, hogy a kormányoldal tekintetében teljesen helytállóan bizonyult a hipotézis, az ellenzék tekintetében viszont csak részben. Azonban az így is egyértelmű, hogy *Brüsszel* megítélésében markánsan eltér egymástól a két oldal.

Irodalom

- Ligeti-Nagy, N., Dodé, R., Jelencsik-Mátyus, K., Varga, Zs., Héja, E., & Váradí, T. (2023). A new ParlaMint corpus for Hungarian – 30m tokens of annotated parliamentary data. In Berend G., Gosztolya G. & Vincze V. (szerk.) *XIX. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 433–446). Szeged: SZTE TTIK, Informatikai Intézet.
- Rychlý, P. (2007). Manatee/Bonito – A Modular Corpus Manager. In *1st Workshop on Recent Advances in Slavonic Natural Language Processing* (pp. 65–70). Brno: Masaryk University. <https://nlp.fi.muni.cz/trac/noske>.
- Sass B. (2017). Keresés korpuszban: a kibővített Magyar történeti szövegtár új keresőfelülete. In Forgács T., Németh M. & Sinkovics B. (szerk.) *A nyelvtörténeti kutatások újabb eredményei IX.* (pp. 267–277). Szeged: SZTE Magyar Nyelvészeti Tanszék.
- Sass, B. (2022). Principles of corpus querying: A discussion note. *Acta Linguistica Academica*, 69(4) (pp. 599–614).

HANTI MOTÍVUMOKRÓL NÉGY TÉTELBEN

SIPOS MÁRIA | HUN-REN Nyelvtudományi Kutatóközpont

sipos.maria@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.14

1. A hagyományos hanti női ruhákon a posztócsikokból összeállított motívumok aprólékosan kidolgozottak, amint az 1. képen látható. Az obi-ugor tárgyakon fellelhető stilizált, geometrikus mintákat, nevükkel (ló, nyúl, sirály, nyírfaág, coboly, róka, rénszarvasagancs, medve stb.), jellegzetes felhasználási helyükkel (ruhaszegély, lábbeli, varrózacskó, gyöngyfüzér, bölcso stb.) számos kutató felgyűjtötte (Szjazi, 2000; Seskin & Sabalina, 2001). Napjainkban gyakran megfigyelhető, hogy a hantik saját maguk számára készített ruháin a jól ismert geometrikus motívumok lényegesen nagyobb méretben, esetenként leegyszerűsítve jelennek meg. Ezt a tendenciát az etnográfia tárgyalja is a női ruházat készítésének változásai között (Moldanova, 2017). A 2. fotó ezt illusztrálja.



1. kép (Khanty Yasang 3631: 11)



2. kép (Khanty Yasang 3570: 1)

2. A fent bemutatott jelenség megfigyelhető más díszített tárgyakon is. Az alábbi felvételen egyrészt egy hagyományörző táborban, szakemberek felügyelete mellett a 2000-es években elkészített nyírkéreg doboz vehető szemügre (3. kép, balra), másrészt egy szuvenírboltban árusított, turistáknak szánt változat (3. kép, jobbra).



3. kép

3. A motívumok életének egy további lehetséges színtere, amikor elveszítik a kapcsolatot azokkal a tárgyakkal, amelyeket eredetileg dekoráltak. Ahogy Oroszország számos más területén, úgy a helyi adminisztráció a Hanti-Manysi Autonóm Körzet központjában, Hanti-Manszijszkban is az őshonos kisebbségek motívumkincsét alkalmazza különféle városi színtereken, amikor hasznosnak találja a helyi identitás építését, hangsúlyozását (Várnai, 2016). Az említett, egészen más méretekben létező motívumok nagy felületeken köszönnek vissza ebben a városban is: plakátok szegélyén, tartálykocsi festésén, korlátokon, magas panelházak falán stb.

4. A motívumok életének újabb színterére ér a turista, ha ellátogat Hanti-Manszijszk szuvenirboltjába, ahol a tárgyakon a terület turisztikai közhelyei (pl. rénszarvas, mamut) mellett az obi-ugor népek ismert ornamentikája is visszaköszön. A szakemberek a jó szuvenir motiváltságát igen összetettnek tartják. Zi Yan Duan és szerzőtársainak segítségével összefoglalva: a jó emléktárgy egyrészt szól az érzékszervekhez, vagyis valamilyen értelemben vonzó benyomást kell keltenie. Másrészt komolyan kötődik a környék kézműves technikáihoz, de azt rugalmasan alkalmazva, vagyis meg is újíthatja. A termék akkor vált ki megbecsülést, ha az elkészítése komoly időráfordítással jár (mivesség), és a készítő kreativitása is megnyilvánulhat. Harmadrészt a helyhez erősen kötődik anyagában, a tárgy funkciójában, a díszítés módjában és a felhasznált motívumokban is. Összességében a tárgy akkor lesz hiteles, ha a vásárló azt a csatornát is hitelesnek vélheti a háttérben, amely révén a termék a helyi készítőtől eljut a boltokba (Duan et al., 2023).

A fenti kritériumokból szinte semmi nem valósul meg a Hanti-Manszijszkban vásárolt vodkás poharak esetében – igaz, azt sem mondhatjuk, hogy a motívumok „megűszös” módon elnagyoltak; éppen ellenkezőleg: meglepően kicsinyek (4. kép).



4. kép

Sem a tárgy funkciója (speciálisan vodkaiásra szolgáló pohárka), sem az anyaga (üveg), sem a dekorálás módja (minuciózus vésés az üveg felületén), sem az aranszegegy a pohár szélén az ugyancsak arany „Jugra” felirattal nem kapcsolódik az obi-ugor kultúrához – de még a filigrán, stilizált vonalrajzok (pl. sámándob) sem a hantik használati eszközein szoktak megjelenni.

Mindezzel együtt e poharak tökéletes manifesztumai annak, ahogyan a többségi kultúra alkalmazza és – az egzotikumfaktor kiaknázásával – beilleszti saját ízlésvilágába a kisebbségi kultúra bizonyos elemeit. Nem utolsósorban pedig kiválóan alkalmasak arra, hogy vodkával megtöltsük őket, és emeljük az ünnepelt tiszteletére! (5–6. kép)



5. kép



6. kép

Ahogy a hantik jókívánsága szól:

тӕлаη јӕӕ, тӕлаη кӱр! egészséges kezet, egészséges lábat!

Irodalom

- Duan, Z. Y., Tan, S.-K., Choon, S.-W., & Zhang, M. Y. (2023). Crafting a place-based souvenir for sustaining cultural heritage. *Heliyon*, 9(5). <https://www.sciencedirect.com/science/article/pii/S2405844023029687>
- Moldanova, 2017 = Молданова, Т. А. (2017). Трансформация декорирования традиционного костюма хантов реки Тромъеган в XXI веке. In Е. Ф. Фурсова (Ред.), *Народный костюм в Сибири*. Новосибирск: Изд-во ИАЭТ СО РАН.
- Seskin & Sabalina, 2001 = Шешкин, П. Е., & Шабалина, И. Д. (2001). *Мансийский орнаменты*. Санкт-Петербург: Филиал издательства Просвещение.
- Szjazi, 2000 = Сязи, А. М. (2000). *Орнамент и вещь в культуре хантов нижнего приобья*.
- Várnai Zs. (2016). Az őslakos kisebbségi identitás vizuális megjelenése egy arktikus szibériai városban – A nyelvi tájkép vizsgálata Dugyinkában. *Folia Uralica Debreceniensia*, 23, (pp. 275–292).

IRÁNYOK, ÉPÜLETEK, ERŐK: metaforák a magyar minőségügyi és jogi szaknyelvben

SÓLYOM RÉKA | Károli Gáspár Református Egyetem,
Bölcsészet- és Társadalomtudományi Kar | solyom.reka@kre.hu | DOI: [10.18135/PG70.2024.15](https://doi.org/10.18135/PG70.2024.15)

Bevezetés

A tanulmány a szaknyelvi nyelvhasználat metaforáinak témakörét vizsgálja a kognitív metaforaelmélet keretében. Két területen elemez metaforikus kifejezéseket: a minőségügy és a jog magyar szaknyelvében.

Az elemzések során három, a szerző vonatkozó kutatásai (Sólyom, 2020, 2023) alapján gyakori forrástartományhoz – az IRÁNYOKhoz, az ÉPÜLETEKhez és az ERŐKhez – kapcsolódó metaforikus kifejezések bemutatására kerül sor. Korpuszként magyar nyelvű szakszövegek szolgálnak: három minőségügyi témájú ISO-szabvány terminusai, valamint *Magyarország Alaptörvényének* szövege

Szaknyelvi terminusok létrejötte

A *terminusalkotás* természetes és szükséges folyamat. A szaknyelvekben megjelenő új terminusokkal szemben felmerülnek olyan elvárások, mint a szabatosság, az egyértelműség vagy a magyarosság, de ezek a követelmények a valóságban nem mindig teljesülhetnek (Fóris, 2010: 426), sőt, a vonatkozó terminológiai munkák hangsúlyozzák a kreativitás fontosságát is a szaknyelvi terminusalkotásban (Temmerman & Van Campenhout, 2014: 1).

Napjainkban a szaknyelvek bővülése rendkívüli mértékben felgyorsult; a globalizációnak köszönhetően általánosan jellemző a nemzetközi méretű szaknyelvi szabványosítás (Fóris & Bölcskei, 2019), és szükséges a szakszókincs árnyalt felosztása, csoportosítása (Kurtán, 2003: 157).

Metaforikus kifejezések elemzési lehetőségei kognitív szemantikai keretben

A fogalmi metaforák kognitív megközelítésének magyar és idegen nyelven is nagy szakirodalmi háttere van: Lakoff & Johnson (1980) óta számos elmélet látott napvilágot. A jelen tanulmánynak nem célja ezek bemutatása, viszont az elemzett metaforikus forrástartományokat ebben az elméleti keretben vizsgálja. Ezeknek a megközelítéseknek az értelmében a „fogalmi metafora két fogalmi tartományból áll: a konkrétabb forrás- és az elvontabb céltartományból” (Kövecses & Benczes, 2010: 81). Előbbi segítségével konceptualizáljuk az utóbbi, általában nehezebben megérthető vagy elvontabb entitást.

A jelen elemzés három gyakori, a vizsgált minőségügyi és jogi szakszövegek esetében prototipikusnak (Tolcsvai Nagy, 2013: 115) bizonyuló forrástartományú metaforát elemez. Az elemzés korpuszát három minőségügyi ISO-szabvány – az MSZ EN ISO 9001:2015, az MSZ EN ISO 13485:2016 és az MSZ EN ISO 14001:2015 –, valamint *Magyarország Alaptörvényének* szövege adja.

A minőségügyi szabványok terminusainak metaforikussága

A következőkben a fenti, minőségügyi előírásokat tartalmazó magyar nyelvű ISO-szabványok metaforikus terminusainak elemzésére kerül sor. Az elemzések hosszabb szövegrészteteket egyik szabvány esetében sem idéznek vagy emelnek ki.

„A szabvány közmegegyezéssel létrehozott és egy hivatalosan elismert szervezet által jóváhagyott, tevékenységekre vagy azok eredményeire vonatkozó dokumentum, amely (...) szabályokat, irányelveket, megoldásokat, jellemzőket fogalmaz meg a rendező hatás adott feltételek között kivitelezhető legkedvezőbb fokának elérése céljából” (1995. évi XXVIII. törvény a nemzeti szabványosításról, idézi: Pónyai, 2010: 3–7; Bölcskei, 2011: 275; Sólyom, 2020: 86). Olyan szakszövegtípus tehát, amelynek témájából és funkciójából fakadóan elsődleges feladata, hogy egyértelműen, a félreérthetőség elkerülésével fogalmazzon meg tudnivalókat. Ennek az elvnek nem mond ellent az sem, hogy a szabványok terminusainak szemantikai felépítése is sok esetben metaforikus. Ezek közül kiemelkedően nagy arányban jelennek meg a három elemzett szabványban az IRÁNYOK (63 db), az ÉPÜLETEK (25 db) és az ERŐK (15 db) forrástartományt tartalmazó metaforikus kifejezések. A fő kategóriákon belül megfigyelhetőek további alcsoportok is; ezek bemutatására kerül sor az alábbiakban (a példák után zárójelben a szabványok rövidített neve és az oldalszámok olvashatók).

Tipikus csoportot alkotnak az IRÁNYOK forrástartomány esetében a *fontos, kint-bent* irányokat tükröző metaforikus terminusok, például: *belső források, külső források; belső és külső kommunikáció* (ISO 9001, pp. 29, 30); *A tervezésnek a kimeneti adatai; a kockázatirányítás lehetséges bemenete; felső vezetőség* (ISO 13485, pp. 23, 31, 35); *külső szervezet; belső audit* (ISO 14001, pp. 19, 35).

Egy másik csoport térbeli (cél)pontokra utaló terminusokat tartalmaz, például: *a (...) tudatosság előmozdításának biztosítását* (p. 20); *Felelősségi körök, hatáskörök* (ISO 13485, pp. 20, 24); *stratégiai irányvonal* (ISO 14001, pp. 9, 23).

Az ÉPÜLETEK forrástartományú metaforák esetében tipikus a szintekre utaló alcsoport, például: *kockázatalapú gondolkodásmód; a szervezeti ismeretek alapulhatnak (...)* (ISO 9001, pp. 23, 29); *A célok (...) különböző szinteken alkalmazhatók (...)* (p. 17); *hivatkozási alapként használt irányvonalak* (ISO 14001, pp. 17, 20).

Egy másik alcsoport a ház részeit tartalmazza, például: *a (...) dokumentáció szerkezete; a nemmegfelelések okainak kiküszöbölése* (ISO 13485, pp. 18, 34); *a fenntartathatóság környezeti pillére* (ISO 14001, p. 8).

A házzal kapcsolatos tevékenységek is megjelennek egy alcsoportban, például: *a minőségirányítási rendszer követelményeinek beépülése (...)* (ISO 9001, p. 23);

beépíti a környezetirányítást a szervezet átfogó irányítási rendszerébe (ISO 14001, p. 9).

Az ERŐK forrástartományú metaforikus kifejezések elsősorban teherre, súlyra utalnak a korpuszban, például: *a szervezetnek (...) fenn kell tartania (...) a környezetet* (ISO 9001, p. 28); *a szervezetnek (...) fenn kell tartania egy vagy több dossziét (fájlt)* (ISO 13485, p. 18); *a szervezetnek folyamato(ka)t kell (...) és fenntartania (...)* (ISO 14001, p. 29).

Az Alaptörvény vonatkozó metaforái

A másik elemzett forrás Magyarország Alaptörvényének szövege. Itt az *ember* forrástartományú, megszemélyesítésnek számító metaforikus kifejezéseket leszámítva az IRÁNYOK (62 db), az ÉPÜLETEK (32 db) és az ERŐK (19 DB) forrástartományú metaforák fordulnak elő a leggyakrabban. Hasonló alcsoportok figyelhetők meg ebben a korpuszban is a vizsgált három metafora esetében, mint a szabványoknál.

Az IRÁNYOK forrástartomány esetében megjelennek a *fent-lent* irányokra vonatkozó kifejezések, például: *legfőbb ügyész* (p. 15), valamint térbeli célpontok: *feladatkörükbe tartozó* (p. 15); *a törvény által a köztársasági elnök hatáskörébe utalt döntés* (p. 16); *az ügy a személyek széles körét érinti* (p. 21); *gazdasági tevékenységének körét* (p. 25).

Az ÉPÜLETEK forrástartományú metaforák esetében itt is megjelennek a ház részei, például: *az emberi lét alapja* (p. 2); *családi kapcsolat alapja* (p. 4); *a bírósági szervezet többszintű* (p. 21). Megfigyelhetők továbbá a házzal kapcsolatos tevékenységekre utaló metaforikus kifejezések is, például: *állami berendezkedés* (p. 3); *támogatásra jogosult* (p. 10).

Az ERŐK csoportjába tartozó metaforikus kifejezések is több alcsoportot alkotnak szemantikai felépítésük alapján. Megjelennek egyrészt az erő és a gyengeség ellentétére utaló kifejezések, például: *részekre szakadt nemzetünk* (p. 2); *a törvényes rend megdöntésére (...) irányuló* (p. 27); *elháríthatatlan akadályba ütközik* (p. 28).

Egy másik alcsoportot alkotnak a teherhez, súlyhoz köthető kifejezések, például: *terhelik azok a kötelezettségek* (p. 7); *a nemzetgazdasági egyensúly helyreállítása* (p. 24).

A kontrollt, szabályozást kidolgozó terminusok pedig a harmadik tipikus kategóriát alkotják az ERŐK forrástartományú metaforákon belül, például: *a Kormány hozzájárulásához kötheti* (p. 24); *az Alkotmánybíróság működése nem korlátozható* (p. 30).

Összegzés

A tanulmány rövid áttekintést adott arról, hogy két, egymástól teljesen különböző témájú magyar nyelvű szakszövegtípus esetében hogyan jelenik meg három, a kognitív nyelvészeti elemzésekből jól ismert metaforikus forrástartomány: az IRÁNYOK, az ÉPÜLETEK és az ERŐK. Világossá vált, hogy mindhárom metaforatípus több alcsoportot alkot a szövegekben; érdekes és további kutatásokra sarkalló jelenség továbbá az is, hogy mindegyik szövegtípus esetében a leggyakrabban előforduló forrástartomány az IRÁNYOK volt a három vizsgált metafora esetében, ezt követték az ÉPÜLETEK kategóriájába tartozó példák, végül az ERŐK kategória jelent meg.

Források

- Magyarország Alaptörvénye: <https://njt.hu/jogszabaly/2011-4301-02-00.11> (2022. június 15.)
- MSZ EN ISO 9001:2015 (Minőségirányítási rendszerek. Követelmények) szabvány
- MSZ EN ISO 14001:2015 (Környezetközpontú irányítási rendszerek. Követelmények alkalmazási útmutatóval) szabvány
- MSZ EN ISO 13485:2016 (Orvostechikai eszközök. Minőségirányítási rendszerek. Szabályozási célú követelmények) szabvány

Irodalom

- Bölcскеi A. (2011). A szabványügy és a magyar nyelv: a szabványosítás online forrásai az információszerzés, a terminológiai munka és az oktatás szolgálatában. *Magyar Nyelvőr*, 135(3) (pp. 274–285).
- Fóris Á. (2010). A szaknyelvkutatás modelljei és módszerei: szociolingvisztikai megközelítés. *Magyar Nyelv*, 106(4) (pp. 424–439).
- Fóris Á. & Bölcскеi A. (szerk.). (2019). *A szabványosítás fordítási és terminológiai vonatkozásai*. Budapest: L'Harmattan Kiadó, Károli Gáspár Református Egyetem.
- Kövecses Z. & Benczes R. (2011). *Kognitív nyelvészet*. Budapest: Akadémiai Kiadó.
- Kurtán Zs. (2003). *Szakmai nyelvhasználat*. Budapest: Nemzeti Tankönyvkiadó.
- Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago and London: The University of Chicago Press.

ROMANI INTERAKCIÓS RÍTUSOK AZ ONLINE TÉRBEN*

SZALAI ANDREA | HUN-REN Nyelvtudományi Kutatóközpont

szalai.andrea@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.16

A romani online használatát vizsgáló kutatások eddig jórészt ortográfiai kérdéseket és a dialektusok online reprezentációját, valamint emblematis- emancipációs funkció- jú, gyakran a roma aktivizmushoz köthető írott szövegeket vizsgáltak (Leggio, 2020), a spontán közösségi online nyelvhasználat jellemzőire kevesebb figyelmet fordítottak (Hámori, 2018). A pandémia időszaka olyan helyzetet teremtett, amely ráirányította a figyelmet az online kommunikáció „grassroot”, közösségi igények által alakított formái- ra (Prószék, 2017).

A személyes érintkezést korlátozó járványügyi rendelkezések az összetartozás, szo- lidaritás fenntartásában alapvető fontosságú interakciós rítusok láncolatában törést okoztak (Božic, 2021). A kapcsolattartás online formái ugyanakkor felértékelődtek, és olyan új, online kapcsolati rítusok keletkezésének lehettünk tanúi, mint az egészség- ügyi dolgozókat köszöntő kollektív taps, az erkélyekről közvetített ismétlődő zenei performanszok és közös éneklés, továbbá a különféle online kihívások, például a Face- bookon terjedő hagymaevés rítusa vagy a táncos kihívás a hazai roma közösségekben. Az új helyzet a kutatói terepmunkamódszerekben is változást hozott (Fotta & Gay y Blasco, 2024), a terepmunka abban az értelemben is több színterűvé vált, hogy nem kerülhette meg az online térben zajló interakciók vizsgálatát (netnográfia).

A korlátozások a sűrű hálózatú roma közösségek kapcsolatcentrikus morális rendjé- vel élesen szembenálló viselkedést követeltek meg, lehetetlenné tették a szolidaritás és osztozás alapvető fontosságú gyakorlatait, a rokonlátogatást, a nagy létszámú tár- sadalmi összejöveteleket (esküvők, gyászritusok, virrasztók) vagy a konfliktusmegoldást célzó nyilvános eskük megtartását. Magától értetődő volt a kérdés, vajon hogyan me- nedzselik ezt a saját értékrendjük szerint amorális viselkedést normatívvá tevő helyze- tet a gábor közösségek.

Míg a roma közösségek társadalmi egyenlőtlenségekből fakadó, kiszolgáltatott helyze- tét több kutatás vizsgálta (Gay y Blasco & Fotta, 2023), arról kevés szó esett, hogy a közös- ségek hogyan reagáltak a járvány okozta krízishelyzetre, és milyen támogató kapcsolati gyakorlatokat hoztak létre. Az alábbiakban röviden egy, a járvány idején kialakult és azóta is fennmaradt online kapcsolati gyakorlatot, a *lajvo manglimo* 'online ima' interakciós rítusát vizsgálom romani–magyar–román háromnyelvű gábor roma közösségekben.

* A kutatást az NKFIH 129378. számú pályázata támogatta.

Az online ima kapcsolati ritusa: résztvevői keret és nyelvi repertoár

A pandémia idején világszerte megfigyelhető volt a vallásos diskurzusok megélenkülése. Ezt jelzi az is, hogy miután 2020. március 11-én a WHO világvárványnak nyilvánította a COVID-19-járványt, az „ima”, „COVID-ima” keresőszó előfordulása minden korábbinál magasabb volt a Google-keresések között (Sinding-Bentzen, 2021), sőt, egyes kormányok kríziskommunikációjában is megjelentek a vallás morális tekintélyére hivatkozó diskurzusok, amelyek segítségével legitimálni igyekeztek a korlátozó rendelkezéseket (Wodak, 2021).

Nem volt ez másként a járvány által erősen érintett gábor közösségekben sem, ahol a bizonytalanság és az izoláció, továbbá a járvány első időszakában a helyi magyar és román társadalom és sajtó irányából érkező bűnbakképző, áldozathibáztató diskurzus – amely a gábor romákat tette felelőssé a vírus terjedéséért – felerősítette a transzcendenshez fordulás igényét. Ez életre hívta a közösségtagok által szervezett, Facebookon meghirdetett, élő közvetítésben zajló online ima (*lajvo manglimo*) gyakorlatát, amely egyszerre a szakralitáshoz fordulás és a roma közösség tagjai közötti szolidaritás megerősítésének ritusa. Míg az ima korábban főként az imaházi színtérhez kötődő intézményi vagy az otthon terében zajló privát gyakorlat volt, a COVID időszakában online, etnicizált kapcsolati ritussá vált (Kádár, 2017), amely nagyrészt romani nyelven zajlott, és a résztvevői is főként gábor romák. Az online ima olyan közös tevékenységgé vált, amely transzlokális, sőt országhatárokon átívelő transznacionális online gyakorlatközösséget hozott létre, összekapcsolva akár különböző országokban tartózkodó résztvevőket is. Bár „offline” személyközi kapcsolatokra épül, de meg is haladja azokat, mivel olyan személyek is bekapcsolódhatnak, akik egyébként személyesen ritkán találkoznának.

Ezeket a ritusokat gábor ismerőseim közösségi médiahálózatának perifériális tagjaként, az online interakciók közönségéhez tartozó megfigyelőként, illetve alkalmanként ismerőseim meghívására, aktív résztvevőként kísértem figyelemmel. Ekkor számomra is beszédjogot biztosítottak a csoportos Messenger-hívás formájában zajló beszédeseményben. Később élő Facebook-közvetítés formájában még nagyobb nyilvánosságot szolgáltattak meg az önkéntesként *lajvo manglimot* hirdető férfiak. Néhány hét elteltével készítettek egy időbeosztást is, hogy ne legyenek átfedésben az egyes imaalkalmak, ezzel megteremtették a ritus időbeli folyamatosságát is. Személyes és közösségi indítatásból, az adventista imaházi vezetőktől függetlenül létrejött egy ismétlődő, sajátos résztvevői kerettel jellemezhető, multimodális, hibrid ritus.

A nyelvi repertoárt tekintve az online ima jelentős részben romani nyelven zajlott, kivéve a Szentírás-részletek magyar vagy román nyelvű felolvasását. Az első néhány hónapban a résztvevők jórészt csak gábor romák voltak. Az imát éppen meghirdető, a ritust vezető-moderáló roma férfi a közösség meghatalmazottjaként, annak „hangjaként” fordul Istenhez (Goffman terminusával élve animátori szerepben). A moderátor és a vele közös offline térben jelenlevő, gyakran csupán hallható, de nem látható családtagok és az online térben szimultán jelenlevő közönség közösen konstruálta meg a ritust. A közösségtagok írott kommentekkel járultak hozzá azt interakcióhoz, üdvö-

zölték egymást, imakéréseket fogalmaztak meg, gyógyulást kívántak, érzelemkifejező emotikonokat, like-jeleket, GIF-eket küldtek. Míg a szóbeli interakcióban domináns volt a romani, az írott kommentek többnyelvűek voltak, ad hoc ortográfiával írott romani, román és magyar nyelvű hozzászólásokat egyaránt olvashattunk.

Az így szerveződő online ima nem szerint differenciált rítus, amely reprodukálja a gábor közösségek „offline”, nyilvános térhasználatát és a nyilvános megszólaláshoz való hozzáférés mintáit az online interakcióban is. A vezető mindig férfi, és csupán néha hallhatók – még ritkábban láthatók – a vele közös térben tartózkodó női családtagok, akik alkalmanként csatlakoznak az imához és az énekléshez. Az online COVID-imákban a nyelvi tabukhoz való viszony is átalakult: a betegség és haláleset említésével kapcsolatos tabuk részleges felfüggesztése figyelhető meg például az imakérések megfogalmazásában. (Az imákból mindenki megtudhatta, hogy mely családok és személyek vannak kritikus helyzetben, a betegségről nyíltan beszéltek, míg korábban a téma kerülendő volt.)

A COVID-ban érintettség különböző fázisaihoz, az azok közötti állapotváltozásokhoz igazodó rituális ciklus jött létre: az otthon tartózkodók, a kórházban ápolott betegek kórteremből közvetített imái és a hazabocsátott gyógyultak hálaadó imái – Van Genep elméletében – az átmenet rítusaiként értelmezhetők, amelyeken legalább online résztvevőként végigkísérhették egymást a közösség tagjai. Családi körben zajló imát, előre meghirdetett, nyilvános *lajvo manglimok*at és zárt, meghívásos csoportokban zajló imákat ugyanúgy láthattunk, ahogyan részesei lehettünk a kórházi kórteremben együtt imádkozó, maszkot, pizsamát viselő gábor férfiak csoportos bejelentkezéseinek. (A pizsamában mutatkozás a testképpel összefüggő tabuk miatt korábban elképzelhetetlen lett volna, a rendkívüli helyzet azonban átmenetileg felülírta ezt a tabut is.)

A rítus visszatérő, állandó strukturális elemei a következők. Köszöntés, az online csatorna megfelelő működésének ellenőrzése; rövid ima és/vagy ének; Szentírás-részlet felolvasása (magyarul vagy románul), a közönségtől érkező kérések, imaszándékok felolvasása/ismertetése; könyörgés, ima a betegekért (150 nagycsaládot felsoroló imalistára is volt példa), minden gábor romáért Erdélyben és azon kívül, a helyi magyarokért és románokért, az orvosokért és ápolókért, az ország vezetőiért stb. A betegeket gyakran romani névvel, származási helyre és rokonsági státusra utalva említik, ezzel voltaképpen a roma rokonsági hálózat reprezentációja és megerősítése is zajlik az online rítus keretében. A könyörgéseket az érzelmetli előadásmód jellemezte, olyan nonverbális jegyekkel, mint a sírós hang, egyenetlen beszédritmus, felkiáltó intonáció. Ezt ismét éneklés követte, végül a rítust az elköszönés, áldások, jókívánságok sora és a következő imaalkalom idejének, platformjának, vezetőjének bejelentése zárta, ami voltaképpen a rítuslánc szervezésének aktusaként értelmezhető.

A rítus expanziója és intézményesülése: miért érdekes mindez nyelvészeti szempontból?

Bár az online ima a krízismenedzselés interakciós rítusaként jött létre 2020 tavaszán, a járvány lecsengése után is fennmaradt, sőt elindult az intézményesülés útján. A résztvevők köre bővült. A rítus multietnikussá vált, mivel a kora reggeli online imákba

a gábor résztvevőkön túl egy-egy más romani dialektust beszélő (ún. gyergyói roma és cörhár/korturár) roma résztvevő is bekapcsolódott. Alkalmanként megjelentek az adventista felekezet magyar és román tagjai és vezetői, prédikátorai is, és az online közönségben feltűntek magyar és román ismerősök, gyülekezettagok is. Nemzetközi online imaheteket is szerveztek, ahol a rítus tolmács bevonásával főként angolul és románul zajlott, amerikai vendégpredikátorok részvételével. Ekkor a romani használata a főként egyénileg előadott vallásos énekekre (*sentó djili*) és az üdvözlésekre korlátozódott. Úgy tűnik, hogy az adventista egyház a gábor romák evangelizációjának lehetséges terét ismerte fel a roma kezdeményezésben, és ma már az online ima és egyéb vallási események számára külön Facebook- és Youtube-csatorna jött létre (*Speranța pentru Gabori*, azaz „Reménység a gáboroknak” néven).

Az utóbbi három évben a gábor romani dialektusban előadott „szent énekek” repertoárja jelentősen bővült, azok romák által készített szövege gyakran kivetítve olvasható, azaz a vernakuláris változat írott formájával való kísérletezésnek lehetünk tanúi. Gábor roma kórus is működik, amelyet a missziós alkalmakon gyakran nem roma muzsikusként kísérnek. Ilyenkor előfordul, hogy szolidaritásuk jelzéséül a rítus előadásának idejére gázások „romává válnak”: például egy román adventista énekesnő és zongorista romani viseletbe öltözik és romani nyelven énekel. A kivetített romani énekeket a magyar és román gyülekezeti tagok és prédikátorok is együtt éneklék a romákkal. A kezdetben csak gábor romáknak szánt kapcsolati rítus az online nyilvánosság csatornáinak köszönhetően többnyelvű, multietnikus, intézményesülő rítussá vált.

A rítusnak a nyelvhasználati funkciók bővülése szempontjából is nagy a jelentősége: a romani számára egy új, formális nyelvhasználati szintér jött létre, ahol a romani a közösségen kívüli nyilvános kommunikációban és írott formában is megjelenhet. Olyan témákkal kapcsolatban is használják azt, amelyre korábban korlátozottan volt lehetőség: például vallásos témák megvitatása, bibliai szövegek magyarázata, vallásos énekek és versek szövegének kollaboratív megalkotása, az online közvetítéssel és új kommunikációs technológiákkal kapcsolatos „techtalk” vagy épp a járvánnyal kapcsolatos információk. A használati lehetőségek bővülése utat nyitott egy spontán nyelvi tervezési, korpusztervezési folyamatnak, amely magával hozhatja a romani szókincs, terminológia bővülését. Alakulóban van egy romani „nettalk”, és a járvány témájához kapcsolódó fogalmak romani megnevezésére is formálódnak kifejezések.

A fentiekén kívül miért lehet érdekes számunkra mindez? Ezek a multimodális műfajok hasznos forrást jelenthetnek egy leendő romani nyelvi korpusz és egy, a közösségek élő nyelvhasználatára támaszkodó romani szótár számára is. Lehetőséget kínálnak továbbá egy spontán, a közösségek nyelvi gyakorlatából és kommunikációs igényeiből kiinduló, többközpontú, a dialektális sokféleséget természetesnek tekintő, azok iránt toleráns romani nyelvtervezési folyamat tanulmányozására is. Ahogyan más kisebbségi és veszélyeztetett nyelvek, például az uráli nyelvek esetében létrejöttek nyelvtechnológiai támogatással készült korpuszok, morfológiai elemzők, a romani nyelv dokumentációját, digitális támogatását és megtartását is hatékonyan előmozdít-

hatná a Prószéky Gábor több évtizedes munkásságához köthető innovációk alkalmazása, a roma közösségek és a nyelvtechnológiával, lexikológiával és romanival foglalkozó nyelvészek együttműködésében.

Irodalom

- Bentzen, J. S. (2021). In crisis, we pray: Religiosity and the COVID-19 pandemic. *Journal of Economic Behavior & Organization*, 192, 501–543.
- Božic, S. (2021). Interactional Ritual Chains and Sustainability of Lockdowns: Quarantines, "Social Distancing" and Isolation during the COVID-19 Pandemic. *Sociologija i Prostor*, 59, 3–34.
- Fotta, M., & Gay y Blasco, M. (Eds.). (2024). *Ethnographic Methods in Gypsy, Roma and Traveller Research: Lessons from a Time of Crisis*. Bristol University Press.
- Gay y Blasco, P., & Fotta, M. (Eds.). (2023). *Romani Chronicles of COVID-2019: Testimonies of Harm and Resilience*. New York–Oxford: Berghahn.
- Hámori, Á. (2018). „Jó lett, Taven baktali!” Egy orális kultúra vizualitása: a cigány nyelv meg-JEL-énesei a YouTube szemiotikai terében. In Tódor, E. M. et al. (szerk.), *Nyelvi tájkép, nyelvi sokszínűség II* (pp. 13–31). Cluj-Napoca: Scientia.
- Kádár, Z. D. (2017). *Politeness, Impoliteness and Ritual: Maintaining the Moral Order in Interpersonal Interaction*. Cambridge: Cambridge University Press.
- Leggio, D. V. (2020). Romani on the Internet. In Matras, Y., & Tenser, A. (Eds.), *The Palgrave Handbook of Romani Language and Linguistics* (pp. 515–537). Palgrave.
- Prószéky, G. (2017). A számítógép, az elektronikus kommunikáció és az internet hatása. In Tolcsvai Nagy G. (szerk.), *A magyar nyelv jelene és jövője* (pp. 321–335). Budapest: Gondolat.
- Wodak, R. (2021). Crisis communication and crisis management during COVID-19. *Global Discourse*, 11(3), 329–353.

SEGMENTATION OF TEXTUAL ARTEFACTS IN DIGITAL HUMANITIES PROJECTS

CRISTINA VERTAN¹, WALTHER VON HAHN²

¹Herder Institute for Eastern European History | ²University of Hamburg

cristina.vertan@herder-institut.de | WvHahn@gmx.de | DOI: [10.18135/PG70.2024.17](https://doi.org/10.18135/PG70.2024.17)

Derived mainly from a European/American language technology background, digital humanities projects have used in a rather blind way for decades annotation as “golden tool”. Less attention was paid to the type and scope of such annotations. This paper discusses segmentation decisions as an obligatory prerequisite of any annotation project. Also it introduces a number of parameters which should be the ground for these decisions. The paper tries to address a number of issues, some of them known from the research on rich morphology and less resourced languages, in which Gábor Prószték plays an important role (Prószték & Merényi, 2012).

1 Introduction

The tasks of representation, processing and analysis of textual artefacts is extending nowadays from mere language technology applications (e.g. machine translation or information extraction for marketing purposes) and corpus linguistics, to all fields of humanities. Methods, Models and Tools used once either for rather engineer applications or pure linguistic research are now one of the first (and most important) steps in the digitization and digital usage of textual objects in other disciplines as history, cultural heritage, ethnology, musicology or dedicated cultural area research (e.g. classical Ethiopic, old cultures on the American territory). This scope extension was done for almost two decades without a deep reflection on new data types, to which such methods and tools were exposed.

As a result, many projects in digital humanities either artificially limit themselves to some positive shallow representation or simply ignore text features (like vagueness and uncertainty of natural language) which in the hermeneutic (analogue) analysis plays a major role. (Dilthey, 1883). This gap between the physical digital analysis and the “traditional” (analogue) research on categories like Dilthey’s “Verstehen”. They often led to misunderstandings and less enthusiasm for the new methods among humanities’ researchers.

During the last years increased awareness about the potential of deep and text-type oriented representation in the humanities was gained together with the massive expansion of computational power, but looks like the representation of an increasing number of textual features. They can lead to better models for the artefacts.

In this paper we argue that not only the number of represented objects but their selection out of the text plays a central role. Although this may seem trivial we will show, focusing on the “segmentation” problem and its implication for annotation, that the contrary may be true. We claim that such multidimensional analysis of a project scope, available data and technical limitations has to be done right at the beginning of each DH-project, independently of the technical paradigm followed (deep-learning, ruled-based or mixed) and might avoid illusions about possible results.

The paper mainly concentrates on research projects and their practical success or failure. The contents of this paper may sound a bit trivial to non-experts in Digital Humanities, but the given examples might avoid illusions about possible results of projects. The central issue is how to prepare (represent) texts for digital processing of any kind.

2 Written artefacts, annotation and the segmentation problem

In the following sections we will use the term written artefacts for all sorts of objects containing language (in contrast to art objects or topographical pictures, audio or audio-visual objects). By language we understand written natural language but also other similar notation types like musical or phonetic alphabets.

Written artefacts in natural language can be represented though different script types: alphabetic (e.g. all languages from the Latin, Germanic, Slavic, Finno-Ugric families as well as Arabic, Jewish, Greek), syllabic (e.g. the Japanese Katakana or Hiragana), abugida (a mixture between alphabetic and syllabic as Amharic and classical Ethiopic (Gééz) or logographic (with symbols representing a concept like Maya glyphs, Egyptian hieroglyphs, Sumerian or Chinese).

Digitization of written artefacts is the prerequisite for any DH-Project and comprises the digital image of the text as well as any further transformation aiming at the enrichment with knowledge which may serve the research purpose (annotation). We distinguish among:

- a) metadata-annotations (referring the entire object) and
- b) content-annotation (on different parts of the text).

The current paper discusses the content-annotation, whilst one should keep in mind that the metadata granularity, i.e. the level at which the metadata are inserted (e.g. entire collection, book, page) faces also the segmentation problem.

2.1 Annotations-Levels and Processing Pipelines

In DH-projects we distinguish between layout, linguistic and domain/application annotations. They are created automatically or manually. In contrast to layout or domain annotations, linguistic annotations are often hidden in the visualization. They are mostly used in order to detect domain specific information, so their scope is strictly related to the aim of the research. Many projects grab automatically pre-compiled processing pipelines: tokenizer → lemmatization → part-of-speech tagger, followed in rare cases by a syntactic

annotation of noun or verbal phrases. Occasionally semantic roles or links to language specific subsets are annotated.

The problem of this type of approach is, that such pipelines are applied without a deep definition of the annotated segments and the appropriateness of this employed segmentation to the particular research problem, respectively underlying language. In the next section we will argue that the decision on text-segmentation is a prerequisite for any successful DH-project dealing with textual artefacts. We will discuss different levels of segmentation and the possible pitfalls by annotations, keeping always an eye on the software economy.

3 Segmentation Levels

In this section we will discuss possible segmentation levels in written artefacts languages with alphabetic scripts and give some hints about challenges when dealing with other language types as mentioned in section 2.

However before deciding on a proper segmentation, the first decision to be taken is the alphabet which will be the basis for the annotation. It often happens in DH projects that languages based on non-Latin scripts are transliterated in Latin script. However, not all (historical) scripts have standardized transliterations (e.g. for Ottoman Turkish there exist several transliteration schools). Thus, one should decide which text version (original script or one of the transliterations) will be the basis for the annotation. Secondly, one should decide if there is a one to one linear correspondence between transliteration and original script alphabet. In case of Semitic languages e.g. transliteration may contain the vocalization, while the original script is might be missing it. The most complicated case is an annotation, where the text must have a right-to-left direction and the annotation (of separable verb forms, e.g.) have to be annotated left-to-right.

In order to illustrate different levels of annotation-complexity we present in Figure 1 three cases. Case 1: Annotation in a traditional basic language technology project (Part-of-Speech-Annotation /PoS in a German text). The focus here is on the annotation of the supposedly most relevant PoS (nouns, verbs, adverbs, adjectives). The second case shows a more complex annotation model in the annotation of classical Ethiopic (Vertan, 2018a). Here the annotation is done on several layers, there is a more fine-grained segmentation, and the original script and transliteration are both involved. The third case shows an even more complex situation in which annotation also involves segments at the text level, references to external sources (ontologies) (von Hahn & Vertan, 2019).

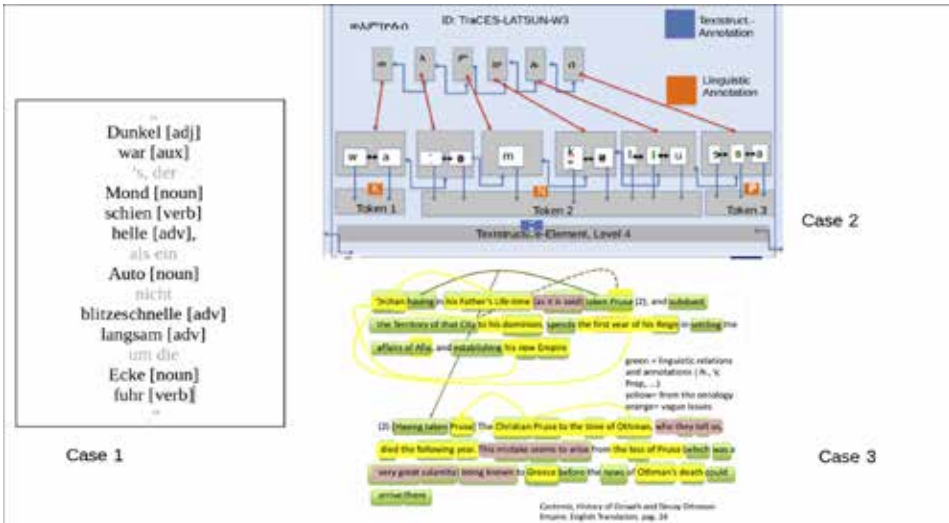


Figure 1 Different segmentation levels

From the above examples one can see that different levels of segmentation may be considered. Our claim is, that before starting to design the annotation model one should go through the segmentation scale presented in Table 1 across a multidimensional parameter scale presented in Table 2.

Segments	Targets, blind spots
Characters: letters (a, ț, í, ç, Ĥ, Ȧ, ...) separators (:, empty character), single numbers (1, 6, 8, ...)	For annotation of transcription doubts (in Edition projects), or when the letter itself is a meaning carrier; for the detection of abbreviations or segment borders (take into account, that non-European languages may have other separators as the blank) and that one letter may be encoded with 2 Unicodes (like the case of Ĥ in Amharic)
Punctuation marks (blank, ,, ?, -, ı, ::)	Often neglected and just erased from the source text, punctuation marks deliver information on the level of discourse as well as syntax (relative clause). They must be separated from the meaning segments (words, letters).
Letter groups (sch)	E.g. for Semitic languages in which one letter may be transliterated by a group of characters in the Latin alphabet. (consonant-vowel)
Morphemes (ver-, -ul,)	To mark postfixed articles (like in Bulgarian or Romanian) or prefixed particles like in German
Tokens (Dunkel, war's, der, ቀንከም)	Group of characters between two separators. In the Amharic example ቀንከም we have a combination of conjunction (ቀ=and) and pronoun (ንከም = together). In the German example "war's" is an abridged form of 2 words (war es =it was)

Words (Common or proper names, Verbs, etc.)	Usually the main segment in DH and language technology projects
Compounds (Donaudampfschiffahrtsgesellschaft, Alu mini umher stellung)	Must be fragmented into its (sets of) parts and thus produces ambiguity. The second example has 12 readings!
Multiword lexemes (The United States of America)	Often these segments should be isolated together as meaning carriers. One should decide for every singular case if only the idiom as such or also its components need to be segmented for the projects' result.
Idioms (Solving that math problem turned out to be a piece of cake for her)	Very important for semantic representation as the meaning of the idiom differs from the simple combination of the meaning of its parts.
Syntactic categories	It is important to define the category and the grammar (constituent vs. dependency)
Concepts [IDEA]	Concepts in the sense of the "triangle of reference" can have word equivalents per language and can be suitable for synonym analysis or language comparisons. However, concept hierarchies (ontologies) cannot be built with words.
Propositions (logical combination of group of words)	Whereas propositions do not comprise questions or relative clauses, negations with their scope are standard versions of (declarative) propositions. Propositions can be orders, modal utterances, declarations, e.g.
Text segments beyond sentence level	It is primarily used for summarization or discourse analysis. Textual segmentation may comprise different (possibly hierarchical segmentation levels). In critical edition projects one may face the problem of overlapping text segments (different editions with different segmentation layers). Syntactically ambiguous sentences may require overlapping annotations too.

Table 1 Segmentation Layers

One particular aspect, which may affect any of the above mentioned levels is the (partial) destruction of one or more segments. This has to be represented and the representation form is essential for any further processing. In texts based on hieroglyphs for example it is crucial not only to mark a specific area as destroyed but also to distinguish about the position and the type of the destruction. This is something to be kept in mind by any project working with historical /archeological material.

It is apparent that there is no "natural" segmentation of texts across any of the levels in Table 1. Thus the choice of the segmentation levels depends on a series of parameters which we summarize in Table 2.

Segmentation parameter	Consideration
Project outcome	What has to be visualized; which algorithms have to be run in the background?
Available technical capacities	A large collection of texts in which each letter is annotated (i.e. represents a segment) may require very fast storage and processing capacity.
Available software and/or representation language	Does the representation allow for reasoning? Representing and annotating uncertainty at all levels in the text may lead to a very slow performance of the used reasoner; moreover it may exceed the logic implemented within the available reasoner. If e.g. the number of uncertain places at the level of single characters are limited one may omit this segmentation level.
Language-type	Is a transliteration level needed? Is the text an agglutinative language (e.g. Hungarian)? Is it a rich compound language (like German). Which is the written orientation of the language and of the mark-up?
Text-type	Inscription, novel, legal text, legend.

Table 2 Parameters for the choice of segmentation layers

3 CONCLUSION

In this paper we present a possible workflow for DH or language technology projects dealing with annotation of textual artefacts. We argue that the first step in such projects is to set the segmentation levels according to a set of parameters. Although written from a perspective of projects in the European language setting, the paper tries to go beyond these borders and address issues in other language families. Possible representation formats which are flexible enough to model this multidimensional problem are presented in (Vertan, 2018b).

References

- Dilthey, W. (1883). *Einleitung in die Geisteswissenschaften (Großdruck): Versuch einer Grundlegung für das Studium der Gesellschaft und ihrer Geschichte* (p. 357). Leipzig: Teubner.
- von Hahn, W., & Vertan, C. (2019). Modelling linguistic vagueness and uncertainty in historical texts. In *Proceedings of the Workshop on Language Technology for Digital Historical Archives in conjunction with RANLP-2019* (pp. 34–38). Varna: INCOMA Ltd. http://doi.org/10.26615/978-954-452-059-5_007
- Prószyński, G., & Merényi, Cs. (2012). Language Technology Methods Inspired by an Agglutinative, Free-Phrase-Order Language. In *Multilingual Processing in Eastern and Southern EU Languages* (pp. 182–206). Cambridge: Cambridge University Press.
- Vertan, C. (2018a). Data Modelling for Historical Corpus Annotation. In Burghardt, M., & Müller-Birn, C. (Eds.), *INF-DH 2018 – Workshopband, 25. Sept. 2018, Berlin*. <https://doi.org/10.18420/infhd2018-17>
- Vertan, C. (2018b). Supporting hermeneutic interpretation of historical documents by computational methods. In Piotrowski, M. (Ed.), *Proceedings of the Workshop on Computational Methods in the Humanities 2018* (pp. 77–82).

FELEZŐ BÁLTÓL A PARANCSPULI-IG

YANG ZIJIAN GYŐZŐ | HUN-REN Nyelvtudományi Kutatóközpont,
Pázmány Péter Katolikus Egyetem
yang.zijian.gyozo@nytud.hun-ren.hu | DOI: 10.18135/PG70.2024.18

Majdnem napra pontosan 14 évvel ezelőtt, 2010-ben a Felező bálon ismertem meg Gábort, aki bevezetett a nyelvtechnológia világába. Azóta a témavezetőm. Együtt számtalan témát körbejártunk, a kínai karakterek okozta problémáitól kezdve, a gépi fordításon és transzformereken át a PULI-modellekig. Bár sokszor a világot kevésbé érdekli, de számos területen sikerült a magyar nyelv *state-of-the-art* eredményeket elérnünk. Jelenleg az egyik zászlóshajónk a PULI. Sikere töretlen, amit bizonyít a számos tudományos és sajtómegjelenés, és emellett a PULI-modellek nevei védjegyekké is válnak. A jelen tanulmány ezt a sikertörténetet tekinti át a szerző szemszögéből.

1. Bevezetés

A Gáborral való találkozást mondhatnám véletlennek vagy sorsszerűnek, de talán a gondviselés a legjobb szó rá. A pszichológia, a nyelvek és a robotika érdekelt. Végül a Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar mérnök informatikus szakára vettek fel. Elég hamar be kellett látnom, hogy a fizika nem az erősségem. Több tárgyból is megbuktam, és évet is kellett ismételnem. Egy csoporttársammal eldöntöttük, hogy egyetemet váltunk. Már az átjelentkezésemet intéztem, amikor kaptam egy meghívót a Felező bálra. Kiderült, hogy a bent maradás miatt annyi kreditet sikerült teljesítenem melléktárgyakból, hogy meghívtak a bálra, ami a hagyomány szerint egy vacsora a Mongol étteremben a tanárokkal. Mivel nem „szabályosan” értem el a félidőt, nem akartam elmenni a vacsorára. De az egyik csoporttársam megkért, hogy kísérem el, mivel ő sem ismert sok embert, de szeretett volna egy jót vacsorázni. A jó vacsora reményében beadtam a derekam. Már a desszertnél tartottunk, amikor mind a ketten meguntuk az eseményt, és elhatároztuk, hogy hazamegyünk.

Ekkor jött oda Prószék Gábor Levendovszky Jánoshoz (nála is megbuktam neurális hálózatokból) megbeszélni egy projektet. Miután megbeszélték, Levendovszky tanár úrnak haza kellett mennie, így ott maradt Gábor velünk. Elkezdte kérdezni, hogy ki milyen szakos, és hogy valakit érdekel-e a nyelvtechnológia. Félbeszakítottam, és megkérdeztem tőle, hogy mi is az a nyelvtechnológia. A bál után meg is kerestem azonnal Gábort. Bizonyítékul szolgáljon az első levelem Gáborhoz:

„Tisztelt Prószéky Gábor tanár úr!

Yang Zijian Győző vagyok a Pázmány Itk kar hallgatója, és még a felező bálon találkoztunk. Önálló laboratóriumra szeretnék keresni valami témát, és érdeklődni szeretnék, hogy esetleg fordulhatnék-e önhöz, és kérem a közeljövőben (holnap, holnap után) egy időpontot, vagy mikor találhatom meg önt az karon.

Köszönettel:
Yang Zijian Győző”

Kaptam is egy témát tőle, majd még azon a nyáron elmentem a MorphoLogichoz szakmai gyakorlatra, Gábor lett a témavezetőm és a többi már történelem...

2. A kínai írás okozta néhány probléma informatikai megoldása

A BSc-témám, Gábornak köszönhetően, a kínai írás okozta problémák megoldása volt. A szakdolgozatom, amivel TDK-n is részt vettem, három problémát igyekszik megoldani egy programcsomag (lásd 1. ábra bal oldala) segítségével. Íme a dolgozat témája a szakdolgozattéma-bejelentésből:

1. **„Egy adott kínai karakter és környezete vizsgálata:** A program célja, hogy megvizsgálja az adott karakter környezetét, hogy alkot-e egy másik karakterrel értelmes kifejezést. A kínai nyelvben léteznek két vagy több karakterből álló szavak, kifejezések, amelyek nyelvtani vagy egyéb okokból különváltak, és ez a program arra ad lehetőséget, hogy ezeket a különvált kifejezéseket a környezetet bejárva megtalálja.”
2. **„A kínai fonetikus ábécé szerinti szótagoló program elkészítése:** A kínai karakteres írásban nincsenek feltüntetve szóhatárok, így ha egy mondatot, vagy kifejezést fonetikus ábécé segítségével átírunk, szintén nem lesznek láthatóak a szóhatárok. A szótagoló program a kínai nyelvtannak megfelelően szótagolja a fonetikus ábécével írt szöveget.”
3. **„Kézzel írt kínai karakter felismerése:** A feladat adott rajzoló felületen, egérrel rajzolt kínai karakter felismerése neurális hálózat segítségével. A program egy a felhasználó által egérrel rajzolt képet alakít át feldolgozható formába, majd neurális hálózat segítségével asszociál egy adott mintahalmazból egy létező kínai karakterre.”

A harmadik feladat egész jól sikerült (lásd 1. ábra jobb oldala), a Kohonen-hálózatot¹ alkalmaztam a probléma megoldására. Azért kaptam kritikát is; szerencsésnek mondhatom magam, hogy Bartos Hubát kaptam bírálónak, aki így kezdte és fejezte be az egyik feladat bírálatát: „A 2. probléma esetében a kiinduló problémafelvetésben van

¹ <https://www.nnwj.de/kohonen-feature-map.html>

furcsaság. ... A szóhatárok megállapításának problémája ennél sokkal nagyobb és lényegesebb probléma, mind az írásjegyes, mind a hangjelölő írásmódban, ennek a megoldásához viszont a javasolt program nem visz közelebb.” A védésen még nem így volt, de utólag már teljesen egyetérttek ezzel a bírálattal.



1. ábra. A program főmenüje és égbolt

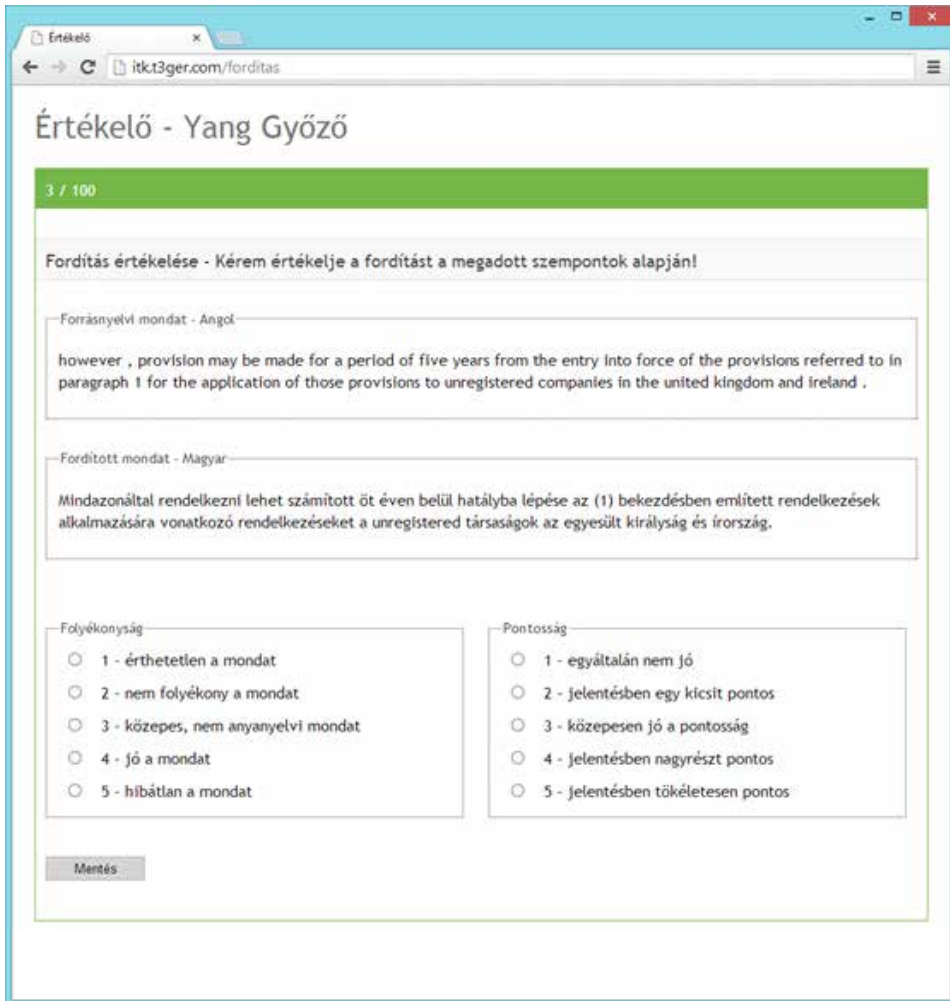
3. Hierarchikus modell alapú gépi fordítás minőségének becslése

A mester éveim alatt kezdtem el azt a témát, ami kitarzott a doktori iskola végéig, itt kaptam Gábor mellé egy második konzulenzst, Laki Lászlót, akihez azóta is szoros barátság fűz. A diplomatervemben ez állt:

„A dolgozatom során felépítettem egy kifejezés alapú illetve egy hierarchikus gépi-fordító-rendszert angol–magyar nyelvpárra. A gépi fordítókat betanítottam a Hunglish korpusszal (Halácsy et al., 2005), majd lefordítottam vele egy tesztszöveget, és kiértékeltem a minőségét referenciafordítással történő és referenciafordítás nélküli kiértékelő módszerrel. A gépi fordító felépítéséhez a Moses keretrendszert (Koehn et al., 2007), a kiértékeléshez a Moses és a QuEst (Specia et al., 2013) keretrendszert használtam.”

„A munkám során felépítettem a QuEst keretrendszert magyar nyelvre, és kiértékeltem vele az általam felépített kifejezés alapú és hierarchikus modellen alapuló gépi fordító által generált kimenetet. Az angol–spanyol nyelvpár esetén létezik 17 alaptulajdonság, ami elég a kiértékeléshez. Ez alapján én is optimalizáltam magyar nyelvre a QuEst rendszerét. Az általam optimalizált rendszer 23 alaptulajdonsággal dolgozik.”

„A QuEst kiértékeléshez szükség volt emberi értékelésekre, amihez készítettem egy értékelő weboldalt (lásd 2. ábra), amit a doktori témámhoz is felhasználtam.”



2. ábra. T3ger kiértékelő felülete

4. epQue: Gépi fordítás minőségét becslő programcsomag

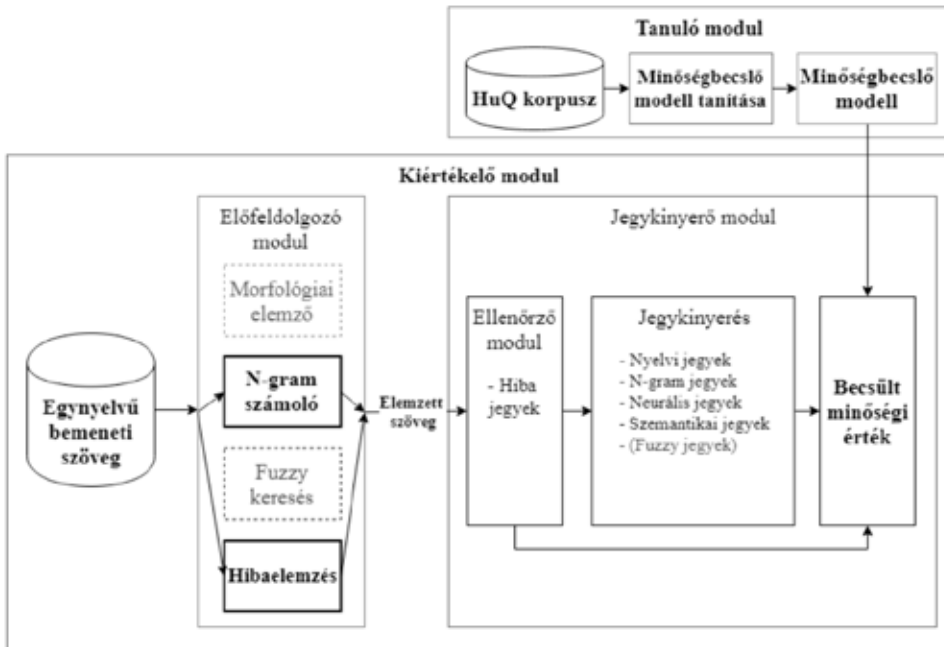
A Pázmányon a kutatócsoport a 314-es szobában volt, ezért a π fontos eleme lett a kutatásaimnak. Hamar kitaláltam, hogy lesz egy π Rate nevű valami a kutatásomban, ezért elkezdtem a kalózos témakört felépíteni. A PhD témám a gépi fordítás minőség-becslése volt. A doktori disszertációm (Yang, 2019) három nagyobb témát ölel fel, idézet a tézisfejezetemből:

1. **A Hun-QuEst rendszer és a HuQ korpusz:** „A minőségbecslés módszere gépi tanuláson alapszik. A modell, jegyek segítségével, a forrásnyelvi és a gép által lefordított mondatokból különböző nyelvfüggetlen és nyelvspecifikus minőségi mutatószámokat nyer ki. Majd a mutatószámok gépi tanuló algoritmussal be-tanítottam emberi kiértékelésekre. A modell tanításához tanítókorpuszra lenne szükség, azonban a kutatás ideje alatt nem állt rendelkezésre angol–magyar nyelv-

vü emberi kiértékeléssel rendelkező párhuzamos korpusz. Ezért az angol–magyar minőségbecslő rendszer tanításához létrehoztam egy kézzel kiértékelte tanítókörpuszt (Yang et al., 2016). Ennek segítségével létrehoztam egy angol–magyar minőségbecslő rendszert. A felépített rendszeren különböző méréseket végeztem el. Először az angol–spanyol nyelvre optimalizált alapjegykészletet mértem le angol–magyar nyelvre, majd megvizsgáltam a Specia és társai (Specia et al., 2013) által implementált 76 jegykészletet is. Ezt követően saját szemantikai jegyekkel kísérleteztem (Yang et al., 2018). A szemantikai jegyekhez egy angol–magyar szótárt, a WordNetet, a szóbeágyazási modellt és a látens szemantikai analízis módszerét használtam. Végeztem jegykiválasztást is, ami azt jelenti, hogy kevesebb releváns jeggyel sikerült további eredményjavulást elérnem. Ezáltal, kevesebb erőforrással magasabb minőséget értem el. A WordNet jegyeket angol–spanyol és angol–német nyelvpárokra is kipróbáltam. Mindkét esetben jobb eredményt értem el az alapjegykészlethez képest.”

2. **A MaTros rendszer:** „A második kutatás során a minőségbecslés módszerét használtam a különböző gépi fordítórendszerek kimeneteinek kombinálására (Laki & Yang, 2018). Az általam létrehozott kompozit rendszer egy kifejezés alapú statisztikai, egy hierarchikus statisztikai és egy neurálhálózat-alapú gépi fordítórendszer kimenetét kombinálja. A rendszer a minőségbecslés módszerével kiválasztja a három rendszer fordításából a legjobb fordítást, és az lesz a rendszer végső kimenete. A módszeremet négy különböző nyelvpárra teszteltem: angol–magyar, angol–német, angol–olasz és angol–japán. Az eredmények alapján rendszerszinten a kompozit minőségbecslés módszerével gépi fordítórendszerem minden esetben jobb minőséget eredményezett, mint az általa felhasznált rendszerek önmagukban. Angol–magyar nyelvpár esetében nyelvfüggő jegyekkel tovább tudtam növelni a rendszer minőségét.”
3. **A π Rate rendszer (lásd 3. ábra):** „A harmadik kutatásban a minőségbecslés módszerét kiterjesztettem egynyelvű szövegek minőségének becslésére. A kutatás célja az volt, hogy megvizsgáljam az interneten elérhető, emberek által produkált szövegek hibáit, valamint a minőségbecslés módszerével, létrehozok egy automatikus hibadetektáló programot. Az kutatásom arra mutatott rá, hogy az emberek által létrehozott egynyelvű szövegek, a gépi fordítók által generáltakkal ellentétben, nagyrészt nem nyelvtani hibákat tartalmaznak (Dömötör & Yang, 2018). Ezen szövegek minőségi problémái inkább az internetezők írási szokásaiból adódnak, mint például az ékezetek vagy az írásjelek elhagyása. Az általam létrehozott, egynyelvű minőségbecslő rendszer (Yang & Laki, 2017) jól alkalmazható a korpusznyelvészetben vagy a természetesnyelvi elemző rendszerek előfeldolgozó moduljaiban.”

Idézem Gábort a védésemről: „...az is jó tudós, aki magától mindent kitalál, de az, aki a témával alázatos és tud a szakmába úgy belemerülni, hogy közben felveszi azokat az információkat, amiket mások mondanak...” A védésen is és utólag is teljesen egyetértek vele.



3. ábra. A π Rate rendszer architektúrája

5. Nyelvészeti Diákolimpia

2014-ben Kínában rendezték meg a Nemzetközi Nyelvészeti Diákolimpiát². Gábor felkért, hogy legyek a Magyar Nyelvészeti Diákolimpia csapatának a kísérője. Rendkívül megtisztelő feladat volt számomra, három éven keresztül kísértem a csapatot, azóta is a szervező csapat tagja vagyok. A magyar csapat évről évre sikeresebb eredményeket ér el:

- 2013: 1 dicséret
- 2014: 1 dicséret
- 2015: 1 bronz
- 2016: 1 bronz, 2 dicséret
- 2017: 2 dicséret
- 2018: 1 ezüst, 2 bronz
- 2019: 2 bronz, 2 dicséret, 1 legjobb megoldás különdíj
- 2021: 1 ezüst, 1 bronz, 2 dicséret
- 2022: 3 bronz, 1 legjobb megoldás különdíj
- 2023: **Csapatbronz!**, 1 ezüst, 1 bronz

A 4. ábrán látható az eddigi egyik legnagyobb sikerünk. Érdekeség, hogy mára már hagyománnyá vált, hogy a csapat a Gábor házában lévő zászlót viszi minden évben (Kína óta) a nemzetközi döntőre.

² <https://ioling.org>

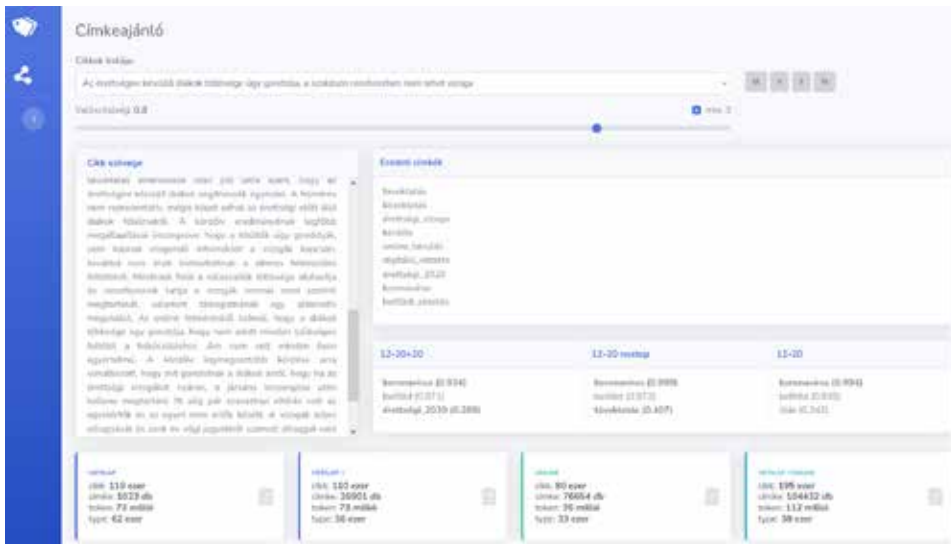


4. ábra. Az első csapatbronzérem és a világot megjárt zászló

6. MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoport

A doktori cím megszerzése után pár évig még a Pázmányon maradtam az MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoportban. A kutatási irányomat meghatározta a neurális hálózatok térhódítása. Szerencsém volt olyan kutatásokban részt venni, mint:

- Szövegek stílusának meghatározása neurális hálózattal (Dömötör, et al., 2022): a kutatásunkban a fastText (Joulin et al., 2017) eszközzel osztályoztunk különböző típusú (irodalmi, tudományos, jogi, sajtó, beszélt nyelvi, komment) szövegeket.
- Ékezet-visszaállítás (Laki & Yang, 2020): az ékezet-visszaállító rendszer tanításához egy neruálshálózat-alapú gépfordító-rendszert használtunk, amely transzformermodellt használ. A módszert kiterjesztettük 14 különböző nyelvre, a rendszerünk minden nyelvre több mint 98%-os pontossággal tudta visszaállítani az ékezet nélküli szavakat.
- Zéró kopulák automatikus felismerése (Dömötör et al., 2020): a kutatásunk során létrehoztunk egy eszközt, amely a zéró kopulás mondatok automatikus felismerésére és a zéró kopulák mondatba való beillesztésére alkalmas. In-domain tesztkorpusz esetén az eszköz közel 90%-os pontossággal tudta megfelelő helyre beilleszteni a zéró kopulát.



5. ábra. Címkeajánló rendszer

– Kulcsszógenerálás sajtószövegek számára (Yang et al., 2020): Létrehoztunk egy címkézőrendszert, amellyel sajtószövegek automatikus tematikus címkézését tudjuk megvalósítani. Ez volt az első komoly megbízásom Gábortól (lásd 5. ábra).

A nyelvtechnológia közben elmozdult a transzformerek (Vaswani et al., 2017) irányába. Így a kutatási témám is elmozdult:

- Transzformeralapú gépi fordítás (Laki & Yang, 2022): létrehoztuk a legjobb minőségű angol–magyar neurális gépfordító-modelleket. Kísérleteztünk a Marian keretrendszerrel (Junczys-Dowmunt et al., 2018), az mBART (Liu et al., 2020), az mT5 (Xue et al., 2021) és az M2M100 (Aharoni et al., 2019) modellekkel. A kutatásainkban megmérettettük az általunk tanított modelleket olyan ipari gépfordító-rendszerekkel, mint a Google Fordító, a Bing/Microsoft Fordító, a Yandex Fordító, az eTranslation vagy a DeepL.
- Szöveg-összefoglaló (Yang, 2022): létrehoztuk az első magyar nyelvű neurális extraktív és absztraktív szöveg-összefoglaló modelleket. Az összefoglaló modellek elkészítéséhez különböző BERT (Devlin et al., 2019) alapú modelleket használtunk. Az absztraktív modellekhez az előre betanított többnyelvű BERT modellt, valamint a magyar egynyelvű huBERT Base (Nemeskey, 2021), valamint a HILBERT Large (Feldmann et al., 2021) modelleket használtuk. Továbbá végeztünk transzfer tanítást is.

Ez idő alatt lehetőséget kaptam, hogy Gábor mellett társtémavezető legyek, és betekintést nyerjek az arab nyelvű neurális szöveg-összefoglalás témájába (Kahla et al., 2021, 2022) (lásd 6. ábra).



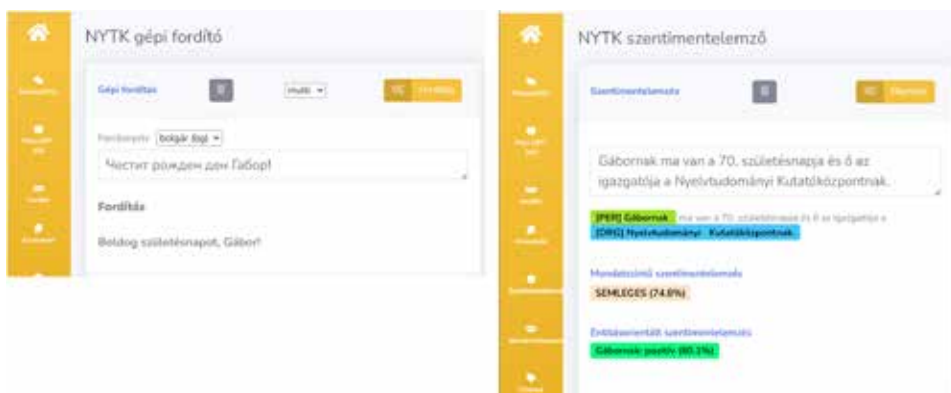
6. ábra. Arab absztraktív szöveg-összefoglalás

7. Nyelvtudományi Kutatóközpont

Mikor Gábort kinevezték a Nyelvtudományi Kutatóközpont (akkor még ELKH, most már HUN-REN előtaggal) igazgatójának, felajánlotta, hogy ott folytassam a kutatásomat. Már a személyére sem tudtam volna nemet mondani, de bedobta, hogy a kutatóközpont hamarosan beszerez két szuperszámítógépet: egy 8 darab NVIDIA A100 (80GB) GPU-ból álló (*bolka*) és egy 4 darab NVIDIA A100 (80GB) GPU-ból álló (*lodka*) szuperszámítógép *node*-ot.

Az első feladatomban egy ELECTRA modellt (Clark et al., 2020) tanítása volt (Yang & Váradi, 2021). Majd ahogya a transzformernyelvmodell-alapú megoldások átvették a vezetést a nyelvtudományban, úgy fordult a fókuszom a nyelvmodellekre, illetve a nyelvmodelleken alapuló alkalmazások és kutatások irányába. Létrehoztuk magyar nyelvre a state-of-the-art:

- neurális szöveg-összefoglaló modelleket (Yang, 2022),
- neurális gépi fordítás minőségbecslő modelleket (Yang & Laki, 2023),
- neurális szentiment- és entitásorientált véleményelemző modelleket (Laki & Yang, 2023) (lásd 7. ábra jobb oldala),
- első neurális morfológiai generátor modelleket (Laki et al., 2023),
- magyarcentrikus többnyelvű (12 különböző nyelvről magyarra) gépi fordító-modelleket (Laki & Yang, 2023) (lásd 7. ábra bal oldala).



7. ábra. Gépi fordító és szentimentelemző demó

Mindközben, ahogy megérkezett a szuperszámítógép, elindítottuk a nyelvmodellezési kutatásokat. Kezdetben kisebb modellek előtanításával kísérleteztünk (Yang & Váradi, 2024), mint ELECTRA, BART (Lewis et al., 2020), RoBERTa (Liu et al., 2019) vagy GPT-2 (Radford et al., 2019), valamint különböző nyelvmodellekkel különböző finomhangolós (Yang & Laki, 2023), promptolós (Yang & Ligeti-Nagy, 2023) és alkalmazási (Hatvani et al., 2023) kísérleteket végeztünk. És így szépen elkezdtünk haladni a nagyobb nyelvmodellek irányába...

8. PULI-modellek és *benchmarkok*

Jelenleg a HUN-REN Nyelvtudományi Kutatóközpont Nyelvtechnológiai kutatócsoportjának egyik zászlóshajója lett a PULI. Szerencsésnek mondhatom magam, hogy egy ilyen nagy presztízsű kutatást vihetek.

A PULI-család mára már több különböző méretű előtanított és finomhangolt nyelvmodellekből áll:

- **PULI BERT-Large** és **PULI GPT-2** (Yang et al., 2023): Mindkettő modell 345 millió paraméteres. A PULI BERT-Large modellünk a legtöbb feladatban felülmúlta a huBERT modellt.
- **PULI GPT-3SX** (Yang et al., 2023): 6,7 milliárd paraméteres GPT-NeoX (Black et al., 2022) típusú magyar egynyelvű nagy nyelvi modell. Több mint 32 milliárd magyar szavas korpuszon tanult.
- **PULI GPTrío** (Yang et al., 2023): 7,67 milliárd paraméteres háromnyelvű (magyar–angol–kínai) GPT-NeoX típusú nagy nyelvi modell. Több mint 41 milliárd magyar szót látott. Összesen több mint 150 milliárd szavas korpuszon tanult, a cél az volt, hogy a magyar nyelv viszonylag kiegyensúlyozott módon szerepeljen a korpuszban, de transfertanulással kamatoztatni tudjon a többi nyelv adta tudásból.
- **ParancsPULI** (Yang et al., 2024): Létrehoztuk az első magyar nyelvű, finomhangolt, utasításkövető (*instruct*, Ouyang et al., 2022) nagy nyelvi modellt, amely egyaránt képes kérdésekre/utasításokra válaszolni (lásd 8. ábra) és nyelvtechnológiai felada-

tokat megoldani. Jelenleg több mint 6000 magyar nyelvű, összesen több mint 114 ezer prompittal finomhangoltuk.

- **PULI Llumix 32K:** A 7 milliárd paraméteres LLaMA-2 (Touvron et al., 2023) modellt a Together³ finomhangolás útján kibővítette a bemeneti kontextus hosszát 32 ezerre⁴. Majd ezt a modellt támogattuk meg több mint 760 ezer darab, 5000 szónál hosszabb magyar dokumentummal.



8. ábra. ParancsPULI köszöntője

A PULI-modellek mellett a másik zászlóshajó a HuLU (Hungarian Language Understanding Benchmark Kit, Ligeti-Nagy et al., 2022, 2023). A HuLU a legnagyobb magyar nyelvű nyelvtechnológiai *benchmark*, ami jelenleg 7 korpuszból áll. Íme a korpuszok leírása a kiértékelő oldalról⁵:

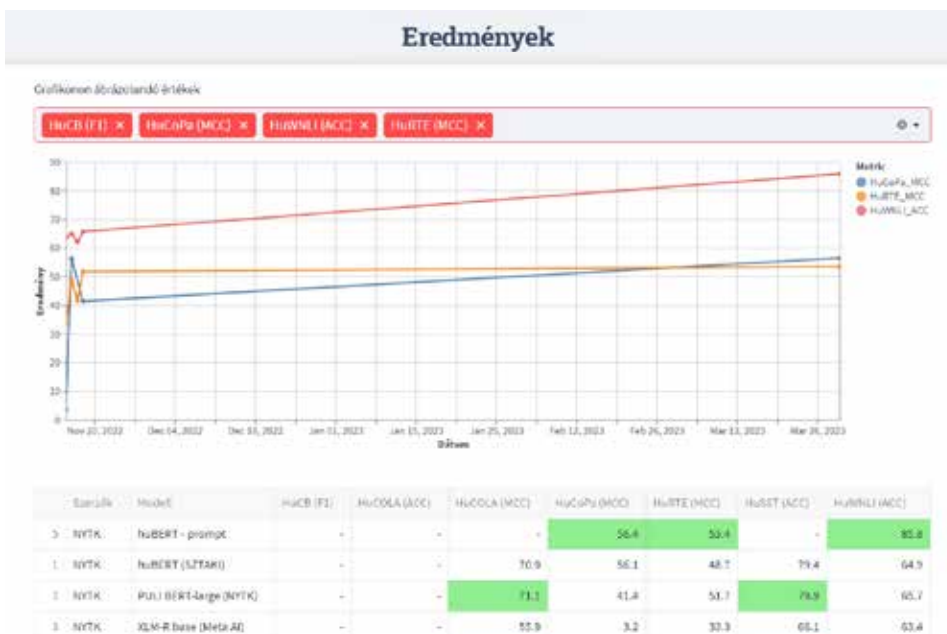
- **HuCB – A CommitmentBank Corpus magyar változata:** A HuCommitmentBank olyan rövid szövegrészletekből áll, amelyekben legalább az egyik mondat tartalmaz egy alárendelő mellékmondatot, amely egy logikai következtetést semlegesítő operátor alá tartozik szintaktikailag. Az adatbázisban a premissza a teljes szövegrészlet, a hipotézis pedig a beágyazott tagmondat. A következtetési feladatban azt kell eldönteni, hogy a szöveg írója milyen mértékben elkötelezett a mellékmondat igazsága mellett. A korpusz 250-250 példás tanító, illetve tesztalmazból, és egy 103 példát tartalmazó validációs halmazból áll.
- **HuCOLA – Elfogadhatósági ítéletek korpusza:** A korpusz 9076 magyar mondatot

³ <https://www.together.ai>

⁴ <https://huggingface.co/togethercomputer/LLaMA-2-7B-32K>

⁵ <https://hulu.nytud.hu>

- tartalmaz, amelyek elfogadhatóságuk, grammatikalitásuk alapján vannak 0-val (nem grammatikus magyar mondat) és 1-gyel (grammatikus magyar mondat) címkézve. A mondatok két annotátor gyűjtötte 3 nyelvészeti szakirodalomból. Mindegyik mondatot négy annotátor annotálta. A végső címke a többségi címke. A tanító-, validációs és tesztanyag aránya 80% (7276 mondat), 10% (900 mondat) és 10% (900 mondat).
- **HuCoPa – A hihető alternatívák korpusza:** A korpusz 1000 példát tartalmaz. Mindegyik példában egy premissza és két alternatíva található. A feladat, hogy kiválasszuk az alternatívák közül azt, amelyikben a leírt helyzet ok-okozati összefüggésben van a premisszában leírt helyzettel. A korpusz előállításához az eredeti angol CoPA korpusz példáit fordítottuk és újraannotáltuk. A tanító-, validációs és tesztalmaz 400, 100, illetve 500 példát tartalmaz.
 - **HuRTE – Következtetések felismerésének korpusza:** A korpuszban 4 504 példa található. Minden példa tartalmaz egy (néha több mondatos) premisszát és egy egy-mondatos hipotézist, és a feladat annak eldöntése, hogy az előbbiből következik-e az utóbbi vagy sem. A korpusz a GLUE benchmark részét képező RTE-adatbázisok példáinak fordításával és újraannotálásával jött létre. A tanító-, a validációs és a tesztalmaz 2131, 242 és 2131 példát tartalmaz.
 - **HuSST – A Stanford Sentiment Treebank magyar változata:** A korpusz 11 683 mondatot tartalmaz. Mindegyik mondat szentimentjét egy háromfokú skálán címkéztük. A korpusz előállításához az SST korpusz mondatait fordítottuk és újraannotáltuk. A tanító-, validációs és tesztalmaz 9347, 1168, illetve 1168 mondatot tartalmaz.
 - **HuWNLI – Anafora-feloldási korpusz:** A korpuszban az anafora-feloldás mondat-pár-osztályozási feladatként, a két mondat közötti logikai következtetés meghatározásaként szerepel. Az alapja a HuWS korpusz, amely az eredeti angol Winograd-sémák magyarra fordított és manuálisan kurált példáit tartalmazza. Az NLI formátum létrehozásához a többértelmű névmásokat lecseréltük mindkét lehetséges referenciájukkal. A Winograd-sémákból képzett mondatpárokat kiegészítettük a GLUE WNLI adatbázisának többi mondatpárjával. Az adatokat tanító- (562), validációs (59) és tesztalmazra (134) osztva adjuk közre.
 - **HuRC – Hungarian Corpus for Reading Comprehension:** A korpusz 80 614 példát tartalmaz. Minden példa egy leadból, egy szövegtörzsből és egy maszkolt cloze stílusú kérdésből áll. A feladat az, hogy kiválasszuk a szövegtörzsből azt az entitást, amelyet a kérdésből kimaszkoltak. Az adatokat automatikusan gyűjtöttük a Népszabadság online híreiből (nol.hu).



9. ábra. HuLU – Dicsőséglista

A HuLU mellett létrehoztuk az első magyar nyelvű utasításkövető finomhangoló korpuszt (Yang et al., 2024), melyet a Stanford Alpaca (Taori et al., 2023) korpusz magyarra fordított változatából és magyar kultúrkörre szabott lokalizált promptokból állítottunk össze. Összesen 2000 lefordított és 100 lokalizált promptot tartalmaz a gyűjtemény. Ez a korpusz képezi az alapját a jelenlegi ParancsPULI modellünknek, ami/aki folyamatosan fejlődik és ami/aki külön verses kötettel készült. Fontos megjegyezni, hogy a PULI által írt/generált műben csak az affiliációt írtuk be manuálisan, minden más részét maga generálta.

8. Konklúzió és továbblépési lehetőségek

Gáborral közösen nagy utat tettünk meg a nyelvtechnológia területén a statisztikai módszerektől kezdve a neurális modellekig. A Gáborhoz közel álló gépi fordítás témájától kiindulva mostanra a nagy nyelvi modellek területén is vezető pozíciót sikerült elérni Magyarországon.

Terveink szerint következő lépésként 13 és 70 milliárd paraméteres nagy nyelvi modelleket szeretnénk tanítani, valamint szöveg-hang alapú multimodális modellek előtanítását fogjuk elkezdni.

Köszönetnyilvánítás

Köszönöm, Gábor, az eddigi 14 éves ismeretséget és szakmai vezetést. Idézve a múltból: „Mindenekelőtt szeretnék köszönetet mondani témavezetőmnek, Dr. Prószyk Gábornak, akitől mind szakmailag, mind emberileg számtalan támogatást kaptam az elmúlt évek során. Köszönöm Neki, hogy mindvégig baráti közvetlenséggel fordult felém. Nélküle ez a munka nem jöhetett volna létre.” Ezúton (is) szeretnék nagyon **Boldog Születésnapot** kívánni!

Irodalom

- Aharoni, R., Johnson, M., & Firat, O. (2019, June). Massively Multilingual Neural Machine Translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 3874–3884). Minneapolis: Association for Computational Linguistics.
- Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonell, K., Phang, J., Pieler, M., Prashanth, U. S., Purohit, S., Reynolds, L., Tow, J., Wang, B., & Weinbach, S. (2022). GPT-NeoX-20B: An open-source autoregressive language model. In *Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models* (pp. 95–136). virtual+Dublin. Association for Computational Linguistics.
- Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *International Conference on Learning Representations*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Minneapolis: Association for Computational Linguistics.
- Dömötör A., & Yang Z. Gy. (2018). Így irtok ti: nem sztenderd szövegek hibatípusainak detektálása gépi tanuló módszerrel. In Tanács A., Varga V. & Vincze V. (szerk.), *XVII. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 305–316). Szeged: Szegedi Tudományegyetem, Informatikai Tanszékcsoport.
- Dömötör, A., Kákonyi, T., & Yang, Z. Gy. (2022). What's Your Style? Automatic Genre Identification with Neural Network. *Computación y Sistemas*, 26.
- Dömötör, A., Yang, Z. Gy., & Novák, A. (2020, May). Much Ado About Nothing – Identification of Zero Copulas in Hungarian Using an NMT Model. In Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., & Piperidis, S. (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 4802–4810). Marseille: European Language Resources Association. Retrieved from <https://aclanthology.org/2020.lrec-1.591>
- Feldmann A., Hajdu R., Indig B., Sass B., Makrai M., Mittelholcz I., Halász D. & Váradí T. (2021). HILBERT, magyar nyelvű BERT-large modell tanítása felhő környezetben. In Berend G., Gosztolya G. & Vincze V. (szerk.), *XVIII. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 29–36). Szeged, Magyarország: Szegedi Tudományegyetem, Informatikai Intézet.
- Halácsy P., Kornai A., Németh L., Sass B., Varga D., Váradí T. & Vonyó A. (2005). A Hunglish korpusz és szótár. In Alexin Z. & Csendes D. (szerk.), *III. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 134–142). Szegedi Egyetem.
- Hatvani, P., Laki, L. J., & Yang, Z. Gy. (2023). A pseudonymization tool for Hungarian. *Annales Mathematicae et Informaticae*, 58 (pp. 69–80). <https://doi.org/10.33039/ami.2023.08.009>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017, April). Bag of Tricks for Efficient Text Classification. In Lapata, M., Blunsom P., & Koller, A. (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 427–431). Valencia: Association for Computational Linguistics. Retrieved from <https://aclanthology.org/E17-2068>
- Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., & Birch, A. (2018, July). Marian: Fast Neural Machine Translation in C++. In *Proceedings of ACL 2018, System Demonstrations* (pp. 116–121). Melbourne: Association for Computational Linguistics.
- Kahla, M., Novák, A., & Yang, Z. Gy. (2022). Fine-tuning and multilingual pre-training for abstractive summarization task for the Arabic language. *Annales Mathematicae et Informaticae*, 57 (pp. 24–35). <https://doi.org/10.33039/ami.2022.11.002>
- Kahla, M., Yang, Z. Gy., & Novák, A. (2021, September). Cross-lingual Fine-tuning for Abstractive Arabic Text Summarization. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)* (pp. 655–663). Held Online: INCOMA Ltd. Retrieved from <https://aclanthology.org/2021.ranlp-main.74>
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., & Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL* (pp. 177–180). Prague.
- Laki, L. J., & Yang, Z. Gy. (2018). Combining Machine Translation Systems with Quality Estimation. In Gelbukh, A. (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp. 435–444). Cham: Springer International Publishing.

- Laki, L. J., & Yang, Z. Gy. (2020). Automatic Diacritic Restoration With Transformer Model Based Neural Machine Translation for East-Central European Languages. In *Proceedings of the 11th International Conference on Applied Informatics (ICAI 2020)* (pp. 190–202). <http://ceur-ws.org/Vol-2650/#paper20>
- Laki, L. J., & Yang, Z. Gy. (2022). Neural machine translation for Hungarian. *Acta Linguistica Academica*, 69 (pp. 501–520). <https://doi.org/10.1556/2062.2022.00576>
- Laki L. J. & Yang Z. Gy. (2023). Magyarcentrikus többnyelvű gépfordító rendszerek létrehozása. In Berend G., Gosztolya G. & Vincze, V. (szerk.), *XIX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2023)* (pp. 369–380). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Laki, L. J., & Yang, Z. Gy. (2023). Sentiment Analysis with Neural Models for Hungarian. *Acta Polytechnica Hungarica*, 20 (pp. 109–128).
- Laki, L. J., Ligeti-Nagy, N., Vadász, N., & Yang, Z. Gy. (2023). Neural Morphological Generators for Hungarian. In Berend G., Gosztolya G. & Vincze, V. (szerk.), *XIX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2023)* (pp. 331–340). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Szoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7871–7880). <https://doi.org/10.18653/v1/2020.acl-main.703>
- Ligeti-Nagy N., Ferenczi G., Héja E., Jelencsik-Mátyus K., Laki L. J., Vadász N., Yang Z. Gy. & Váradi T. (2022). HuLU: magyar nyelvű benchmark adatbázis kiépítése a neurális nyelvmodellek kiértékelése céljából. In Berend G., Gosztolya G. & Vincze V. (szerk.), *XVIII. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 431–446). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Ligeti-Nagy N., Héja E., Laki L. J., Takács D., Yang Z. Gy. & Váradi T. (2023). Hát te mekkorát nöttél! – A HuLU első életéve új adatbázisokkal és webszolgáltatással. In Berend G., Gosztolya G. & Vincze V. (szerk.), *XIX. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 217–230). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., & Zettlemoyer, L. (2020). Multilingual Denoising Pre-training for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, 8 (pp. 726–742).
- Nemeskey, D. M. (2021). Introducing huBERT. In In Berend G., Gosztolya G. & Vincze V. (szerk.), *XVII. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 3–14). Szeged, Hungary: Szegedi Tudományegyetem, Informatikai Intézet.
- Osváth, M., Yang, Z. Gy., & Kósa, K. (2023). Analyzing Narratives of Patient Experiences: A BERT Topic Modeling Approach. *Acta Polytechnica Hungarica*, 20 (pp. 153–171).
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Siemens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multi-task Learners.
- Specia, L., Shah, K., de Souza, J. G., & Cohn, T. (2013, August). QuEst - A translation quality estimation framework. In Butt, M., & Hussain, S. (Eds.), *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 79–84). Sofia, Bulgaria: Association for Computational Linguistics. <https://aclanthology.org/P13-4014>
- Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Li, X., Guestrin, C., Liang, P., & Hashimoto, T. B. (2023). *Stanford Alpaca: An Instruction-following LLaMA model*. https://github.com/tatsu-lab/stanford_alpaca
- Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Canton Ferrer, C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P. S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi, K., Schelten, A., Silva, R., Smith, E. M., Subramanian, R., Tan, X. E., Tang, B., Taylor, R., Williams, A., Kuan, J. X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., & Scialom, T. (2023). *Llama 2: Open Foundation and Fine-Tuned Chat Models*.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., & Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 5998–6008). Curran Associates, Inc.
- Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 483–498). Association for Computational Linguistics. doi:10.18653/v1/2021.naacl-main.41
- Yang, Z. Gy., Laki, L. J., Váradi, T., & Prószték, Gy. (2023). Mono- and multilingual GPT-3 models for Hungarian. In *Text, Speech, and Dialogue* (pp. 94–104). Plzeň, Czech Republic: Springer Nature Switzerland.
- Yang, Z. Gy. (2019). *επQue: Gépi fordítás minőségét becslő programcsomag*. [Ph.D. dissertation, Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar].
- Yang, Z. Gy. (2022). Neural text summarization for Hungarian. *Acta Linguistica Academica*, 69 (pp. 474–500). <https://doi.org/10.1556/2062.2022.00577>
- Yang, Z. Gy., & Laki, L. J. (2017). πRate: A Task-oriented Monolingual Quality Estimation System. *International Journal of Computational Linguistics and Applications*. [Under publication].
- Yang, Z. Gy., & Laki, L. J. (2023). Enhancing machine translation with quality estimation and reinforcement learning. *Annales Mathematicae et Informaticae*, 58 (pp. 182–190). <https://doi.org/10.33039/ami.2023.08.008>
- Yang, Z. Gy., & Laki, L. J. (2023). Solving Hungarian natural language processing tasks with multilingual generative models. *Annales Mathematicae et Informaticae*, 57 (pp. 92–106). <https://doi.org/10.33039/ami.2022.11.001>
- Yang, Z. Gy., & Ligeti-Nagy, N. (2023). Improve Performance of Fine-tuning Language Models with Prompting. *Infocommunications Journal, Special Issue on Applied Informatics* (pp. 62–68). <https://doi.org/10.36244/ICJ.2023.5.10>
- Yang, Z. Gy., & Váradi, T. (2021). Training language models with low resources: RoBERTa, BART and ELECTRA experimental models for Hungarian.
- Yang, Z. Gy., & Váradi, T. (2024). Training Experimental Language Models with Low Resources, for the Hungarian Language. *Acta Polytechnica Hungarica*. [In press].
- Yang, Z. Gy., Agócs, Á., Kusper, G., & Váradi, T. (2021). Abstractive text summarization for Hungarian. *Annales Mathematicae et Informaticae*, 53 (pp. 299–316).
- Yang, Z. Gy., Dodé R., Ferenczi G., Héja E., Jelencsik-Mátyus K., Kőrös Á., Laki L. J., Ligeti-Nagy N., Vadász N. & Váradi T. (2023). Jönnék a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar nyelvre. In Berend G., Gosztolya G. & Vincze V. (szerk.), *XIX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2023)* (pp. 247–262). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Yang, Z. Gy., Dodé R., Héja E., Laki L. J., Ligeti-Nagy N., Madarász G. & Váradi T. (2024). ParancsPULI: Az utasításkövető PULI-modell. In Berend G., Gosztolya G. & Vincze V. (szerk.), *XX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2024)* (pp. 61–72). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Yang, Z. Gy., Laki, J. L., & Siklósi, B. (2016). HuQ: An English-Hungarian Corpus for Quality Estimation. In *Proceedings of the LREC 2016 Workshop – Translation Evaluation: From Fragmented Tools and Data Sets to an Integrated Ecosystem*.
- Yang, Z. Gy., Laki, L. J., & Siklósi, B. (2018). Quality Estimation for English-Hungarian Machine Translation Systems with Optimized Semantic Features. In Gelbukh, A. (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp. 88–100). Cham: Springer International Publishing.
- Yang, Z. Gy., Novák, A., & Laki, L. J. (2020). Automatic Tag Recommendation for News Articles. In *Proceedings of the 11th International Conference on Applied Informatics (ICAI 2020)* (pp. 442–451). <http://ceur-ws.org/Vol-2650/#paper45>
- Yang, Z. Gy., Szlávik Sz. & Ligeti-Nagy N. (2024). Magyar nyelvű utasításkövető korpusz építése Stanford Alpaca promptok fordításával és lokalizálásával. In Berend G., Gosztolya G. & Vincze V. (szerk.), *XX. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2024)* (pp. 243–255). Szeged: Szegedi Tudományegyetem, Informatikai Intézet.
- Zhuang, L., Wayne, L., Ya, S., & Jun, Z. (2021). A Robustly Optimized BERT Pre-training Approach with Post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.

PRÓSZÉKY GÁBOR, AZ ALKALMAZOTT NYELVÉSZ

I DOI: 10.18135/PG70.2024.19

Prószéky Gábor élete a számítógépes nyelvészettel szorosan összefonódott, a számítógépes nyelvészet pedig mind az alkalmazott nyelvészet, mind az elméleti nyelvészet elengedhetetlenül fontos kutatási területévé és segédtudományává vált. Szakterülete, kutatási eredményei, kiterjedt kapcsolatrendszere, nyitott, barátságos, ambiciózus személyisége arra predesztinálta, hogy az alkalmazott nyelvészet minden fontos intézményében, bizottságában, szervezetében vezető szerepet töltsön be. Ebben a rövid visszaemlékezésben csak kis töredékeket villantok fel a vele kapcsolatos emlékeim közül.

2000-ben egy MÖB kutatói ösztöndíjjal az Istituto di Linguistica Computazionale pi-sai intézetében tölthettem tíz napot. Ez az időszak meghatározó volt a tudományos szemléletformálásom szempontjából. E világhírű kutatóintézet lehetőséget adott nekem arra, hogy megismerkedjek a kutatásaikkal, és bemutassam náluk a saját eredményeimet is. Találkoztam többek között az intézet alapító vezetőjével, Antonio Zampollival és munkatársával, Nicoletta Calzolarival, akik érdeklődtek a saját kutatásaimról, valamint a magyarországi irányzatokról, trendekről és eredményekről. Gondosan beosztották az ott töltött időmet: minden egyes kutatócsoportjukkal eltöltöttem egy-egy napot, ahol beavattak a kutatások részleteibe, és várták is a kérdéseimet. A beszélgetések során kiderült, hogy három magyar kutató – Kiefer Ferenc, Szépe György és Prószéky Gábor – neve és eredményei jól ismertek számukra. Előbbi kettőt természetesen én is jól ismertem, utóbbi nevét emlékezetembe véstem. A következő években számos szakmai rendezvényen és bizottságban találkoztam vele és eredményeivel.

Prószéky Gáborral való személyes megismerkedésemet a Papp Ferenc Baráti Körnek köszönhetem. A Papp Ferenc Baráti Kör (PFBK) Papp Ferenc tanítványai és tiszteelői alapították 2001-ben, a rendszeres programok szervezője és a vendéglátó Klaudy Kinga volt, egészen a kör 2021-es megszűnéséig.¹ Ezzel párhuzamosan jött létre a Papp Ferenc Alkalmazott Nyelvészeti Asztaltársaság (alapítók: Klaudy Kinga, Szépe György és Prószéky Gábor), amely kéthavonta tartotta találkozóit az Akadémiai Klub könyvtártermében és éttermében, és ahol kellemes környezetben eltöltött ebéd vagy vacsora közben nyílt lehetőség kötetlen beszélgetésekre, szakmai információk cseréjére. A PFBK évente tartott decemberi ülésein, az ELTE BTK F épületében, a Fordító és Tolmácsképző Tanszék elegáns könyvtárában előadásokat, könyvbemutatókat hallgattunk meg, szakmai és baráti beszélgetésekre volt lehetőségünk. Nekem, mint fiatal kutatónak, ez kiváló lehetőség volt arra, hogy a magyar alkalmazott nyelvészet nagy egyéniségeivel és más fiatalokkal is személyesen találkozhatnék és megismer-

¹ Klaudy K. (2022). *Tájékoztató a Papp Ferenc Baráti Kör Húsz éves tevékenységéről, 2001–2021* (Kézirat).

kedhessek, köztük Prószéky Gáborral is; megértettem továbbá, hogy jelentősége van a személyes visszaemlékezéseknek és a tudományról szóló „anekdotáknak”, mert ezek betekintést engednek a tudományos munka és a kutatók mindennapjaiba, és a száraz tényeken túl az kutatók emberi oldalát is jobban megismerjük. Gábornak a 2001-ben e körben tartott előadása egyébként megjelent a *Modern Filológiai Közlemények* Papp Ferenc Emlékszámában.² Nagyon jól emlékszem az előadására, amelyben egy tudománytörténeti érdekességet mutatott be: hogyan találta meg és próbálta megmenteni 1983-ban a Papp Ferenc-féle szövegmutató szótár,³ illetve az ennek alapját képező Debreceni Tezaurusz eredeti lyukkártyáit a Közgazdasági Egyetem egyik folyosójának egyik szekrényéből. Lebilincselő stílusban, humorral átszöve, könnyedén tartotta meg az előadást, érthetően magyarázta el a számítógépes nyelvészet fogalmait, és hogy miért is volt nehéz a lyukkártyákon tárolt adatokat újfajta formátumba áttenni.



² Prószéky G. (2001). Az „újrafelhasznált” szövegmutató szótár. *Modern Filológiai Közlemények* (Papp Ferenc Emlékszám), III. évf. 2. szám (pp. 121–123). https://epa.oszk.hu/04600/04637/00006/pdf/EPA04637_modern_filologiai_kozlemenyek_2001_2.pdf

³ Papp F. (1969). *A magyar nyelv szövegmutató szótára*. Budapest: Akadémiai Kiadó.

Az alkalmazott nyelvészeti témájú doktori programokban számos hallgató nagy érdeklődést mutatott a lexikográfiai és terminológiai kutatás kérdései iránt. A lexikográfia önálló kurzusként (Szépe György szervezésében és aktív támogatásával) 2002-ben indult a Pécsi Tudományegyetem Nyelvtudományi Doktori Iskola Alkalmazott Nyelvészeti Doktori programjában (ebben szervezőként és előadóként is részt vehettem). Ennek keretében tartott izgalmas előadást Prószéky Gábor Az elektronikus papírszótártól az „igazi” elektronikus szótárak felé címmel (később ennek írott változata megjelent a *Lexikográfia Magyarországon* című kötetben⁴).

Többféle szakmai szervezetben és bizottságban dolgoztam együtt Gáborral, ahol a kutatási eredményein, előadásain túl lehetőségem volt jobban megismerni a mentalitását, személyiségét is. A Magyar Alkalmazott Nyelvészek és Nyelvtanárok Egyesülete (MANYE, www.manye.hu) vezetőségében eltöltött időszak lehetővé tette számomra, hogy sok nagyszerű emberrel kössek szorosabb ismeretséget; Prószéky Gábor 1998 óta a MANYE választmányi tagja, 2013 óta pedig a MANYE elnöki pozícióját tölti be, ekkortól – másokkal együtt – alelnökként segítem a munkáját. A MANYE 2005-ben Brassai Sámuel-díjjal ismerte el az alkalmazott nyelvészet területén végzett eredményes tevékenységét. A Magyar Nyelv Terminológiai Tanácsában (MaTT, www.matt.hu) 2013 óta töltöm be az elnöki tisztséget, ennek a szervezetnek pedig Prószéky Gábor az egyik alelnöke; a főszerkesztéssel 2008 és 2013 között megjelenő *Magyar Terminológia* című folyóiratnak szerkesztőbizottsági tagja volt. Az MTA Szótári Munkabizottságnak, a Magyar Nyelvi Osztályközi Állandó Bizottságnak és az Alkalmazott Nyelvészeti Munkabizottságnak mindketten tagjai vagyunk, az első kettőnek Gábor több cikluson keresztül volt elnöke vagy társelnöke. Ha nem is tud jelen lenni minden egyes megbeszélésen, de bölcs tanácsaival és derűs személyiségével mindig jó hangulatú környezetet teremtve viszi előbbre a munkát.

A Nyelvtudományi Kutatóközpontban (korábban Nyelvtudományi Intézetben) betöltött főigazgatói (korábban igazgatói) működése során megerősítette az alkalmazott nyelvészeti területeket. A terminológiai kutatások és munkák szempontjából nagy jelentőségű, hogy – Európa más nyelvi intézeteihez hasonlóan – a magyar Nyelvtudományi Kutatóközpontban (a Lipp Veronika vezetésével működő Lexikológiai Intézetben belül) is intézményes helyet kapott a terminológia: 2021-ben vezetésével jött létre a Terminológiai Kutatócsoport, amely azóta már két nagy pályázatot nyert el terminológia témában: a 2021–2023-as időszakra az „ELKH Kiemelt pályázat SA52/2021: Terminológia” címmel, a 2023–2027-es időszakra pedig magyar terminológiastratégia témában „MTA TMNP2023-1/2023. projekt” címmel. Mindkét pályázatban vezető tanácsadóként segítem a munkákat, és nagy örömmel veszek részt a terminológiai kutatásokban. Fontosnak tartom itt megemlíteni a sokak erőfeszítésével terminológiastratégia témában megszületett „zöld könyv”-et.⁵

⁴ Prószéky G. (2004). Az elektronikus papírszótártól az „igazi” elektronikus szótárak felé. In Fóris Á. & Pálffy M. (szerk.), *A lexikográfia Magyarországon* (pp. 81–87). Budapest: Tinta Könyvkiadó.

⁵ Prószéky G. (főszerk.), Fóris Á., B. Papp E., Bölcskei A. & Lipp V. (szerk.). (2023). *A magyar terminológiastratégia kialakítása. Zöld könyv*. Budapest: Nyelvtudományi Kutatóközpont. ISBN 978-963-9074-95-8. <https://doi.org/10.18135/term.2023>

Gábort rendszeresen és sokan hívják konferenciákra és más szakmai rendezvényekre előadni, és ő mindig szívesen teljesíti a felkéréseket. Szereti a „meglepetés” témájú előadásokat, amelyek keretében általában a legújabb számítógépes nyelvészettel kapcsolatos kutatásokról és eredményekről beszél, mindig olyan könnyedén, világosan és érthetően, hogy a hallgatóság szinte úgy érzi, most már mindent ért a témával kapcsolatban. Igazi társasági ember, szórakoztató beszélgetőtárs és előadó, remek kapcsolatépítő.

Kedves Professzor Úr! Kedves Gábor! Gratulálok a születésnapod alkalmából! Kívánok jó egészséget, hosszan tartó, sikerekben gazdag munkát, új ötleteket és személyes boldogságot!

Fóris Ágota

CONGRATULATIONS, THANKS AND BEST WISHES FROM PRAGUE

I DOI: 10.18135/PG70.2024.20

Time flies like an arrow – there cannot be a more appropriate occasion to quote this ambiguous syntactic structure (quoted and contrasted with "fruit flies like bananas" as early as in 1963 by one of the pioneers of machine translation Anthony Oettinger and belonging to the so-called garden-path sentences) other than the occasion of Gábor Prózský's 70th(!) birthday. And to recall the long-term working and personal relations between Czech(oslovak) and Hungarian (computational) linguists the beginnings of which date back to the 1960s(!). The names that come to our mind are Ferenc Kiefer, one of the first mathematical linguists in Europe and also one of the first who introduced generative grammar in Europe and, most importantly, one of those who was very active to make the world-wide community acquainted with the research in the field of computational linguistics carried out in the countries of Central and Eastern Europe. And as well as Ferenc Papp, the organizer of one of the first COLING conferences in Debrecen, György Szépe, György Hell, Dénes Varga, just to name those who come to our mind at this moment. In this connection, it should not be forgotten that the (at that time new) field of computational linguistics received an important support in Hungary by a famous Hungarian mathematician from Szeged University László Kalmár (1905–1976), the founder of the study of mathematical logic and information theory in Hungary.

Though the name of Gábor Prózský as a representative of the second generation of computational linguists is naturally missing in the above list of the first generation of Hungarian computational linguists, there is still an explicit link to one of them in his CV (aside to their joint paper from 1985) – it is certainly not a mere chance that in 1995, Gábor received the Kalmár Award of the John von Neumann Computer Society. We do not remember the year and occasion when we met with Gábor for the first time, but we will always remember the important role that Gábor's MorphoLogic played concerning the implementation of the Czech grammar checker for Microsoft (around 2005), and in general the promotion of Czech in MorphoLogic products and cooperations. We will be grateful to him for substantial "material" help when he proposed to support one of the computational linguistics conferences we organized in Prague. As any organizers of a huge international event, we struggled hard with the financial aspects. Out of a sudden there arrived an offer from Hungary: if you need, we might take care of the conference bags. "We" referred to Gábor's MorphoLogic. Of course, we appreciated any help but we still remember our surprise that somebody took the initiative and made an offer by himself.

We have been meeting often at conferences, where both Gábor's industrial efforts and also his research teams played a role. He has also been very active in European networking of the field – aside from participation in many coordination and support actions, he and his company is a long-term member of META, the Multilingual Europe Technology Association. And we also have shared several bi- and multi-lateral international cooperation and initiatives, such as the close cooperation with colleagues in Hamburg.

Time flies like an arrow but strong friendly and working relations persist if based on firm foundations and developed with mutual respect and understanding. We are very happy that the "founding fathers" of the Hungarian - Czech(oslovak) cooperation in computational linguistics have found such devoted successors as our honored person Gábor Prózský. We in Prague deeply appreciate his contribution to our mutual relations and to the development of our field of study, and cordially wish him many years to come in good health, full energy and well-deserved recognition.

Eva Hajičová and Jan Hajič
on behalf of UFAL, Charles University, Prague



A DIGITÁLIS BÖLCSÉSZETHEZ KÖZÖS BÖLCSÉSSÉG IS KELL

I DOI: 10.18135/PG70.2024.21

Most, hogy Prószék Gábort köszöntjük élete jeles fordulóján, én, aki kevésbé keresem a múltat, mégis visszagondolok valamire, ami minden bizonnyal közös élményünké válhatott. 2008-at írtunk, azok az idők jártak, amikor az ilyen vagy olyan formában mindig jelen lévő bürokrácia nem igazán mutatott készséget arra, hogy a mesterfokú nyelvészet formálódó oktatásának dolgát megkönnyítse. Kiderült, hogy túl sokféle nyelvészet szak van. Van elméleti nyelvészet, van alkalmazott nyelvészet, van nyelvpedagógia, majd elméleti és alkalmazott nyelvészet, aztán mindez A-szakon, B-szakon, sőt C-szakon, leginkább vagy-vagy, és intézményenként ugyancsak másként. Az oktatáspolitikusok összeültettek bennünket, talán többször is, hogy egyezzünk meg a paletta egységes formára való szűkítésében. Írtunk, tanácskoztunk, írtunk, majd valami formálisan szűkebb kínálatban meg is állapodtunk, miközben a valóságban minden tanszék folytatta saját adottságai szerint. A debreceni menü különösen sajtóságos lehetett, mivel mi matematikát, logikát és programozást is tanítottunk, sőt még (az akkor felfutóban levő) multimédiát is, mindezt az elméleti nyelvészeti gondolatvilág megtartásával egy számítógépes nyelvészeti vonulattal ötvözve. Azonban mivel nekünk nem volt egy igazán rangot adó A-szakunk, a C-t beszüntették, a B-t pedig egyre inkább tőlünk függetlenül meghatározott szakokhoz kötötték, a korábban igen széles hallgatói merítési bázisunk is egyre szűkült. Míg korábban a hallgatóink legalább fele informatikus volt (az informatika különböző ágaiból érkező), ami pedig így még az elméleti nyelvészet iránt vonzalmat táplálni kezdők körét is szélesítette, egyszer csak ott tartottunk, hogy míg a történelmi időkben csupán a Bevezetés a nyelvtudományba c. tárgy oktatásából is megélt a tanszék, az most teljesen elvész.

Ekkor jött a gondolat, hogy – megérezve a még csak eljövendő kor szellemét – jó lenne elindulni egy új irányba: a digitális bölcsészet felé. Ott szükség van mindarra, ami a számítógépes nyelvészetben már meghonosodott, de ez legyen egyben nyitás a kultúra irányába is: digitalizálva megőrizni a szöveges és nem szöveges kulturális emlékeket, kereshetővé, és így egy jóval magasabb szinten megismerhetővé tenni a történelem, az irodalom, a művészet dokumentumait; és mindehhez nem kevés nyelvészetre és informatikára, valamint ezeket támogató további diszciplínák bevonására is szükség van. Volt mire támaszkodni: a londoni King's College-ban egy fél évig tanítva élőben tapasztalhattam a digital humanities MA-invenciókból táplálkozó meglétét, továbbá, Debrecenből több éven át akkor még különlegesnek számító virtuális környezetben (a Zoom még nem létezett) közös Digitális bölcsészet kurzusokat tartottunk Európa több, szintén vállalkozó kedvű egyetemével, sőt egy közös európai MA-re is pályáztunk (sikertelenül). Hogy a környezet szokja a digitális bölcsészet gondolatát, még egy digi-

tális bölcsészet központot is létrehoztunk a Debreceni Egyetem Bölcsészkarán, bevonva a tevékenységébe a korábban az egymás tevékenységét csak udvarias közelségből figyelő, változatos szakterületekről érkező kollégákat, néhány kurzus tanításának megosztásával még több bölcsész szak fakultatív kínálatát is bővítve.

Ilyen előzmények után került sor a digitális bölcsészet MA szakalapítási és szakindítási pályázatának a beadására. Nem kevés szervezőmunkával sikerült minden formai követelményt teljesíteni (az oktatók szakmai háttere, idegennyelv-tudása, oktatási gyakorlata, ahol szükséges volt – tudományos minősítettsége stb. mind kipipálva), a kar, majd az egyetem is rábólintott, így a pályázat a MAB elé került. A bölcsészek mellett az informatikusok is tárgyalták. Az eredmény: a bölcsészek nem találták elég bölcsésznek, az informatikusok elég informatikusnak. Sőt, az utóbbiak szakvéleménye még azt is figyelmünkbe ajánlotta, hogy az, amit mi szeretnénk megvalósítani, arra már létezik képzés: csak le kell tenni az ECDL (European Computer Driving Licence) vizsgát.

Bármennyire izgalmasnak tűnt is számunkra egy ilyen vezetői jogosítványra invitálni a reménybeli hallgatóinkat, valahogy mégiscsak abban hittünk, hogy ugyanúgy, ahogy a bölcsészletnek és a kultúratudománynak is inspiráló lehet egy messzebbre mutató röppályára állító (számító)gépre szállás, ugyanúgy az informatika is alighanem tágithatja horizontját egyébként szemmel láthatóan hódító útján. Az elutasítás után nem lobbizás szándékával, csak egyszerűen véletlenül összefutottam a MAB akkor Debrecenből szolgáló elnökével, akihez még sok évvel korábban mint rektorhelyetteshez fordultam a bölcsészkar egyetlen 52 K-s modemes internetkapcsolatának számláit rendszeresen kiegyenlíteni, és elmondtam neki, hogyan jártunk a digitális bölcsészet MA-beadványunkkal. A válasza lakonikusan, de – mint később kiderült – alapvetően lényeges módon, csak ennyi volt: „Így járnak a magányos farkasok.” Aztán persze ezt közelebről is kifejtette, és tanácsot is adott: keressek partnert, más intézményből.

És itt jön a képhez ünnepeltünk, Prószék Gábor. Gáborral már régóta ismertük egymást, elsősorban a sokunk számára (és Kiefer Ferencnek köszönhetően) legendás bizonyos 13-as szobából, de ugyanúgy Papp Ferencen keresztül is, akinek a számítógép iránt táplált vonzalma (és igen: a saját nyelvészetén keresztüli digitális bölcsészeti motivációja) mindkettőnk szakmai pályájára hatással volt. Meg is jelentem egy nap a MorphoLogic székhelyén, és átadtam Gábornak a kisiklott pályázat teljes addigi dokumentációját. Hogy nézze át, és gondolja meg, lehetne-e közösen pályáznunk, a Pázmánynak és Debrecennek együtt. A találkozás eredménye egy olyan beadvány lett, amely miközben alapvonalában egységes volt, szakirányai különbsége hűen tükrözte a két tanszék szakmai specialitásait.

Azonban az újabb beadást meg kellett még hogy előzze annak a két intézmény bürokráciáján való keresztülfuttatása is. Talán ez valamivel könnyebben ment a Pázmánynál, köszönve Gábor mindkét irányú szakmai elismertségének, nálunk azonban felmerült egy váratlan akadály. A Centrum oktatási elnökhelyettese nehezményezte, hogy a pályázat szövegében az első változathoz képest alig talál módosítást. Mivel úgy éreztem, hogy ő ezt szinte nevetlenségnek veszi, gyorsan előveztettem, hogy érdemes

elolvasni a pályázat kísérőszövegét is, amiben válaszolok az előző beadványra kapott (ld. főként az ECDL-t említő, sőt javasló) kritikákra. Végül beadtuk. És nagy örömünkre a MAB is jóváhagyta.

A történet innen már zökkenőmentesen folyt tovább, csak a jelentkező hallgatók számát kellett figyelniük. A Pázmányra mindig többen kopogtattak, de a lelkesedésből nálunk sem volt hiány. A lelkesedést a csapatépítés is fokozta, hiszen egy ízben még közösen is táborozott a két csapat a Debreceni Egyetem síkfőkúti Alkotóházában. Sem akkor, sem később nem volt hiány az alkotásban, sok lelkes hallgatót indítottunk el a pályáján, egyben közösen hozzájárultunk ahhoz, hogy a digitális bölcsészet már-már bevett gondolattá váljon hazánkban, szülessenek izgalmasabbnál izgalmasabb projektek, és újtjukra induljanak korábban ilyen névvel még meg nem nevezett szakemberek e nem is olyan régen született diszciplína akár újraértelmezésére. Azért annyit meg kellett tanulni: a digitális bölcsészethez közös bölcsesség is kell.

Hunyadi László

KI A KRUMPLIT NEM SZERETI...

I DOI: 10.18135/PG70.2024.22

...nem lehet annak könnyű dolga olyan területen, ahol a burgonya az egyik alapvető táplálék. Egy magyar ösztöndíjas fiatalnak a nyolcvanas évek végén kihívást jelenthetett elkerülni a finn egyetemi menzán a vékony héjú főtt krumplit, ami kb. olyan gyakori ott, mint a magyar közkonyhákban a kenyér.

Gábort a helsinki ösztöndíja után ismertem meg, a 80-as évek végén. Akkor még az Akadémia több intézete is a Várban volt, a hírek könnyen jártak a hagyományos csatornákon is, hiszen az internet csak ezután kezdte világhódító pályafutását. A Nyelvtudományi Intézet Finnugor osztályán dolgozó fiatal munkatársak megtudták, hogy van valaki, aki akkortájt tért vissza északi nyelvrokainktól, és szívesen bemutatja, hogy a számítógépet mi mindenre tudják használni a nyelvészek. Meghívták egy kötetlen beszélgetésre, s nekem is átszóltak a Nyelvtörténeti osztály „Wörterbuch” részlegébe, hogy biztosan érdemes és érdekes lesz meghallgatni. A Szentháromság utcai épületben hallottam először a PC-Kimmo, morfológiai elemző és generátor kifejezéseket, s fel sem merült bennem, hogy később több projektben is fogunk együtt dolgozni ezzel a fiatalemberrel. Akkoriban ritkaságnak számított még egy asztali számítógép, féltve őrzött kincsként tekintettünk rá. Gábor megnyugtatóan mindenkit, hogy nem kell tartanunk attól, hogy a bemutató során bármiképpen megsérülhet az osztály egyetlen komputere: „A gépnek nem lehet baja, maximum akkor, ha beleöntenek valamit” – mondta, majd pár perc múlva véletlenül a billentyűzetre lötytyintette teáját... A gépnek persze nem lett baja, az azóta eltelt idő alatt pedig munkánk és mindennapjaink elengedhetetlen részévé vált a számítástechnika.

Mivel tudom, hogy az ünnepeket szeret utazni, most egy kis gasztronómiai kirándulásra invitálom, az elmúlt években határon túli és finnugor terepen készült képeim segítségével.

A Drávaszög egyik – régen halászatáról híres – falujában rendezik meg immáron negyed százada minden ősszel a Kopácsi Halászna-pokat. A környékből és távolabbi településekről ide utazókat elsősorban az ilyenkor „nagyüzemben” készülő csíp-



Kopács: a laskói iskola adománygyűjtő standja a házaknál készült finomságokkal, 2015 szeptembere



Sül a csiptetős ponty (forrás: Új Magyar Képes Újság)

tetős ponty vonzza. Ez nem más, mint egy kibelezett, a gerince mentén felvágott, majd beirdalt és fűszerezett, kilapított hal, amit egy csipeszbe tesznek – ezt egy kb. hüvelykujnyi vastag, lehántolt és behasított faágból nyerik. Az így előkészített „szerkezetet” a tűz mellé szúrják, s a hal mindkét oldalán megsütik. A halásznapokon az utcákat ellepik az árusok, közöttük a helybéli és közeli falvak iskoláinak és óvodáinak a standjai is megtalálhatók. Kézműves portékát, háznál készült lekvárt, ajvárt, mézeskalácsot kínálnak, a bevételt a gyerekek szabadidős programjaira szánják. A falu főterén felállított színpadon egymást követik a néptáncsoportok bemutatói, a tér melletti földes területen sekély árok, a benne lévő parázs melegében pedig az előkészített halak tucatjai pirulnak.

A Drávaszög több településén is termesztik a magyar konyha számára nélkülözhetetlen fűszert, a paprikát. A gondos szárítást és őrlést követően a termés nagy részét



Szárad a paprika a tornácon – Kopács, 2017

a faluban a nagy érdeklődés övezte Paprika-fesztiválon adják el.

A következő képek az Oroszországban élő két rokon nép, a komi-permjákok és az udmurtok asztalához invitálnak.

A földművelésből élő népek nemcsak a maguk által megtermelt növényeket fogyasztják, hanem az erdön-mezőn találhatóakat is hasznosítják. A mezei zsurlót nálunk csak gyógyászati célra használják, legfeljebb teaként fogyasztjuk, míg a komi-permjákok ízletes levest készítenek belőle. A terület központja Kudymkar, a város talán legjobb éttermének – az őslakos kisebbséghez tartozó – tulajdonosa az étlapon tartott ételekkel igyekszik a hagyományt őrizni. A nemzeti öntudat a névváltoztatásban is megmutatkozik: a 2016-ban Кафе Блюз-*nak* nevezett hely ma már – a Google Earth tanúsága szerint – a Национальнӱй nevet viseli, az írásmóddal is felhívva a figyelmet a cirill ábécében különleges, komi-permják betűre.

A forró vízben kifőzött, töltött tésztabatyu a világ számos helyén különböző neveken ismert: hol *raviolinak*, hol *gyozának*, hol *barátfülének* nevezik. A komi területeken – ahonnan egyes vélemények szerint hódítóútjára indult – *pelnyanyak*: A szó utótagja az iráni eredetű 'kenyér' jelentésű szó, míg előtagja a magyar fül szóval rokon (http://www.uralonet.nytud.hu/eintrag.cgi?id_eintrag=740&locale=hu_HU), és az étel formájára utal. Magát az ételt és a szót is sokan orosz eredetűnek gondolják, pedig a *pelmenyi* finnugor átvétel az oroszban.

A Gerber az udmurtok nagy ünnepe: 2017 júniusában a fővárosban, Izsevszkben láthatam, hogy milyen urbánus változata. A váratlanul kitört nyári zápor elől a kívülről hagyományos, diszes textilekkel dekorált katonai sátorba húzódtak vissza az ünneplők. A büféasztal jobb oldalán *perepecs*. Kis tésztakosárka, melyet



2016 augusztusa – Kudymkar: mezei zsurlófűből készített krémleves a város legdivatosabb éttermében



Dojebgen, a Kudymkartól mintegy negyven kilométerre fekvő faluban sóval és ezzel a kenyérral fogadott vendéglátóm



többféle töltelékkel is készítenek: krumplival, káposztával, gombával, hússal...
...és hagyományosan kemencében sütik (erre utal is a neve), akárcsak egy másik



Tabany a kemencében...



...és az asztalon (2017. június, Szelty).

nemzeti ételt, a tabanyt, ami serpenyőben készül és kelt tészta az alapja.

2019 őszén Énlakán, a „Hagyomány, változatosság és változás a székely nyelvjárásban” konferenciára kaptunk meghívást, ahol „Nyelvjáráskutatás, számítógépes módszerek, mesterséges intelligencia” címmel tartottunk előadást Prószéky Gáborral és Vargha Fruzsínával. Ott, a konferenciavacsorán tudtam meg, hogy ki nem szereti a krumplit:



Énlaka, 2019. szeptember 14.

Kedves Gábor! Több okból kifolyólag sokkal régebben ismerhetnénk egymást, mint ahogy: mindketten voltunk piarista diákok, aztán ott van még Balatonszemes, meg az is, hogy legkésőbb a huszadik század vége óta a lexikográfia egészen biztosan nem értelmezhető a számítógépes nyelvészet nélkül...

Mindettől függetlenül úgy alakult, hogy találkozni csak valamikor 2015 tavaszán találkoztunk – ám levelet „tőled” ennél sokkal korábban kaptam először. Emlékeim szerint ennek a történetét még nem meséltem el neked – ha mégis, ennek az írásnak annyi hozadéka mindenképp van, hogy most és itt a sztorit igyekeztem szépen illusztrálni is.

Van, ugye, az a hosszú múlttal bíró, nagy magyar lexikográfiai projekt, amelynek vonatkozásában az elmúlt évtizedekben így-úgy mindketten tevékenykedtünk – de valahogy különböző időszakokban, nem átfedésben.

Jómagam 2006-ban szembesültem először az említett szótári munkálathoz tartozó szövegtárral: egészen pontosan azzal, hogy abban miféle karakterkódolási megoldások is találhatóak. Valami ilyesmi tárult a szemem elé, amikor először találkoztam az általad még az 1980-as években kiötlött – és Kornai András által azon frissiben „Prószéky-kód”-nak elnevezett¹ – megoldás gyakorlati alkalmazásával.

Spárta bels43ö0-része s43ints43 illy veszély nélkü0l,

Mellyet Leónidás, gerjeszteni készü0l.

Ams43ares43t meg-vette tudod már pénzével,

Tö0b tanáts43okat is veszteget kintsével.

Nem akarja, hogy a' végezés meg-álljon,

'S tö0rvényeinkre titkon s43íve hóltig fájjon

'S mos43olyogva tes43zi els43ö4 mozdulását.

A' néző4k s43zemei formáján el-ves43znek,

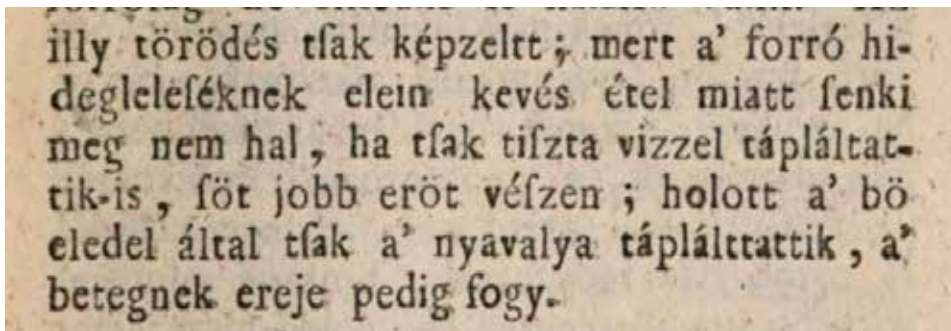
Egy pis43zs43zenés43t s43ehol miatta nem tes43znek

Ez volt az a látvány, amelynek a következtében számomra a Prószéky-kódok hosszú évekig egyet jelentettek a betűhűség egy sajátos értelmezésével. Amit itt láthatunk, az az, hogy ebbe a korpuszba – a Jóisten legnagyobb dicsőségére – sikerült néhány olyan – legfeljebb nyomdászattörténeti szempontból érdekes – graféma több százezernyi előfordulását fáradságos munkával rögzíteni, amelyeknek a megjelenítése egy paleográfiai projektben biztosan jelentőséggel bírhat – de egy alapvetően értelmező szótári munkálat során minden bizonnyal nem. Számszerűsítsünk: a korpusz mintegy 390.000 Prószéky-kódjának 97 százalékáért 3 graféma „felel”: egy ilyen ő, egy ilyen ú és egy ilyen í.

¹ Kornai A. (1986). Szótári adatbázis az akadémiai nagyszámítógépen. *Műhelymunkák II.* (p. 67).

A hosszú s 270.000 előfordulása mellett rögzítésre került 79.000 ő meg 24.000 ú is. Ezen három betű mindegyikének fontos tulajdonsága, hogy semmiféle fonológiai relevanciával nem bírt a vele párhuzamosan ugyanekkor már létező s-hez, ő-höz vagy ú-höz viszonyítva: azaz egy-egy olyan hangot reprezentáltak, amelynek a jelölésére nemhogy a mai magyar ábécében megvan a megfelelő betű, de már 200 éve is megvolt.

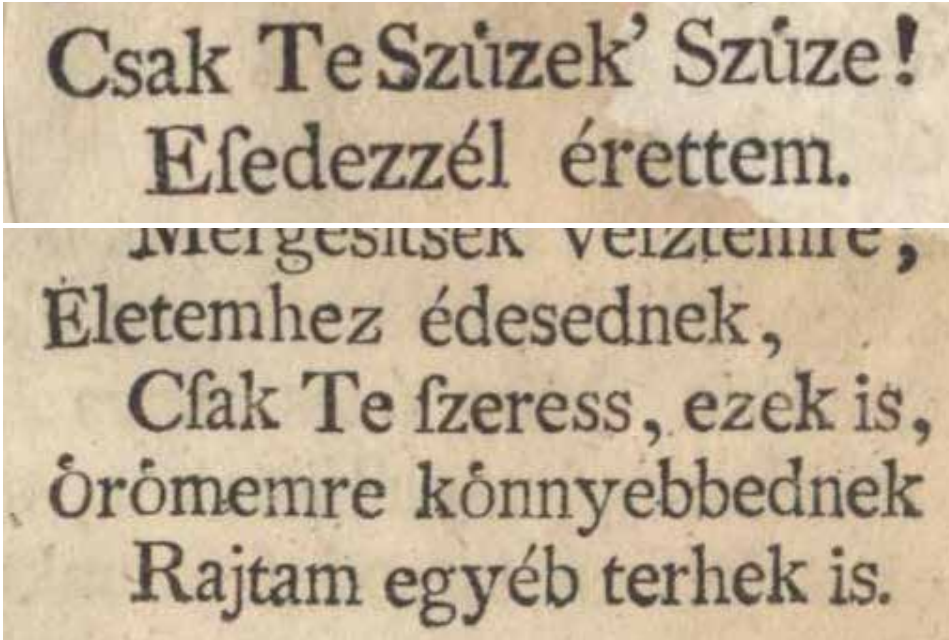
Az alábbi képanyag ékes bizonyítékul szolgál arra, hogy ezeknek a betűknek a 18. századi nyomtatványok lapjain való feltűnése vagy hiánya esetleg tipográfiai „divatirányzat”-okkal lehetett összefüggésben – vagy egyszerűen azzal, hogy az adott nyomdának milyen ólombetűből mekkora készlet állt éppen rendelkezésére a szedőszekrényben. Az eredeti nyomtatványok oldalképeinek tanulmányozása során egyértelművé válik: a hosszú s-t és társait nemhogy az egy időben keletkezett különböző kiadványokban nem szedték következetesen, de a használatuk még egy kötetben belül is ingadozhatott.



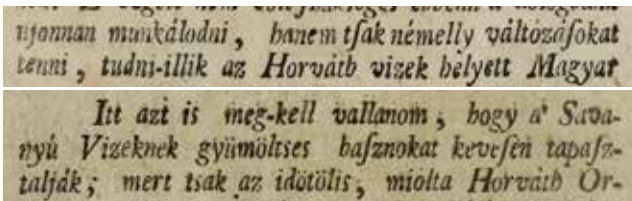
1. kép. Marikovszky Márton *A néphez való tudósítását* 1772-ben, Nagykárolyban nyomtatták: a *tsak* nem fordul elő máshogy, mint hosszú s-sel

ritésből tett fogadásokat az Egek tsak nevetik, és meg-nem büntetik: Addig várapoztatta, és reménytette tehát haszontalanul a' fel, gyúlladott Hermenét, hogy emez végtére szerezemes nyughatatlanságától kényszerítettett utánna tudakoztatni, a' ki a' kimlöi által tsak hamar ki-tanúlta, hogy a' Hertzeg egészségének leg-kevesebb hibája se volna, és tsak szánt-szándékos hallasztással támogatná-el

2. kép. Mészáros Ignác *Kártigámját* szintén 1772-ben nyomtatták, de Pozsonyban: szabad szemmel is jól láthatóan rendelkezésükre állt néminemű hosszú s, a *tsak* mégis mindig így lett szedve



3. kép. Cserey Farkas *Isten anyjának dicsérete* című művét is 1772-ben nyomtatták, ám azt Bécsben: a 135. oldalon *Csak*, a 139.-en *Csak*...



4. kép. A magyarországi orvosvizekről és a betegségekben azokkal való élésnek szabott módjairól című, 1783-ban, Nagykárolyban nyomtatott mű előszavában szintén van *tfak* és *tsak* is...

Már önmagában ez a tény erősen meggyengíti a rögzítésük létjogosultságát! Ám ha ehhez még hozzávesszük a korpuszhoz rendelt lekérdezőeszköznek azt a sajátosságát, hogy ha a keresőkifejezés nem betű- és számkarakterre

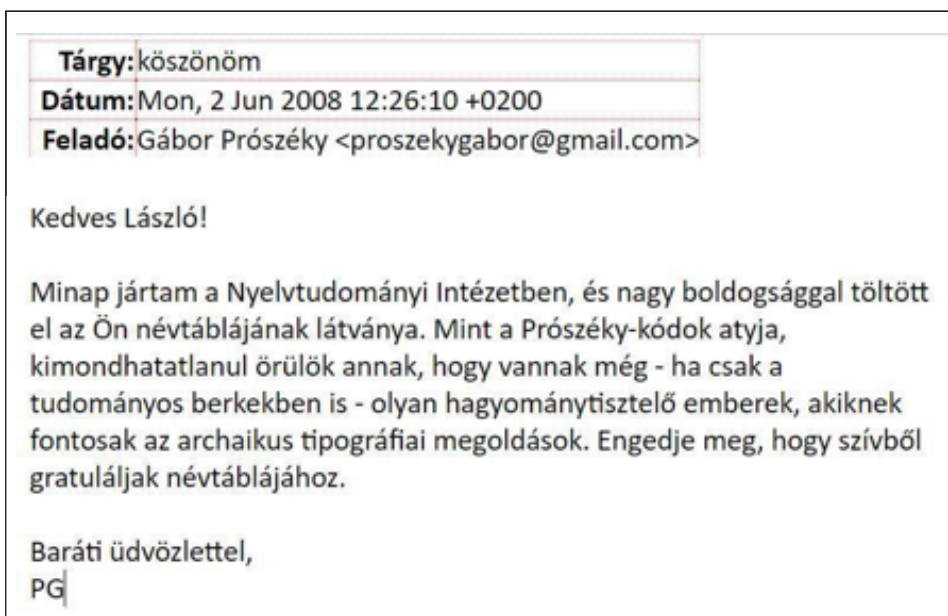
terről betű- és számkarakterre pontosan egyezett meg a korpusz-előfordulás sztringjével, akkor az adat jó eséllyel nem került fel a találati listára, akkor talán érthetővé válik, milyen léptékű problémákhoz vezetett a betűhűség. Aki ebből a korpuszból ki szeretné nyerni például a *dicsőséges* szó összes adatát, az jól teszi, ha nem feledkezik meg a *dits43ös43s43éges* karakterláncról sem...

Reményeim szerint a fentiekből is kitűnik, hogy egy pillanatig sem a kódrendszeret kárhoztattam. Sokkal inkább az olyan grafémákhoz való megszállott ragaszkodást, amelyek rögzítése (1) a projekt fő csapásirányát, a nyomtatott forrásokból vett idézetekkel való illusztrálást tekintve érdektelen volt; (2) feleslegesen túlbonyolította a lekérdezést.

Idővel az s43 lett a felesleges zavart okozó Prószéky-kód szinonimája: feltehetőleg egyszerűen az előfordulási gyakorisága miatt. Annak következtében, hogy heti szinten adtam hangot azon az álláspontomnak, miszerint „*Ceterum censeo...*”, azaz ezek az s43-ak

a korpuszból feltétlenül eltávolítandóak (értve ezalatt, hogy egy kétperces művelettel s-sé egyszerűsítendőek), a kollégáim a felesleges grafémákkal való bajlódás iránti ellen-szenvemet úgy interpretálták, hogy nekem feltehetőleg már eredendően fenntartása-im voltak ezekkel a szám-betű kombinációkkal, sőt talán már a feltalálójukkal szemben is ellenérzések alakultak ki bennem...

2008 egy kora nyári napján aztán az alábbi levél landolt a nyelvtudományi intézeti postafiókomban.



Ismételten csak az emlékeimre hagyatkozom, de szerintem egyetlen percig sem feltételeztem, hogy ezt a levelet valóban te küldted volna, azaz hogy PG az „igazi” Prószéký Gábor lenne. Vicces kedvű kollégámat, Györffy András barátomat gyanítva a háttérben, kiballagtam az irodám ajtaja elé, sejtve, hogy a tréfa nem merült ki egy proszekygabor@gmail.com-os fiók regisztrálásában. Nem kellett csalódnom... Innen nézve azt azért sajnálom, hogy arról, amit az ajtón találtam, nem maradt fenn eredeti dokumentum: de hát akkortájt a mobiltelefonok többségének még nem volt kamerája. (Hogy Györffy András honnan vehette, hogy az e-mailjeidet „PG”-ként írod alá, arról semmilyen elképzelése nincs sem neki, sem nekem.)

Kedves Gábor, Isten éltesen sokáig!
A hetven az új negyven (vagy lehet, hogy harminc: már nem is tudom).
Barátsággal: Laci



5. kép. A Nyelvtudományi Intézet 109-es szobájának ajtaja, 2008. június 2. (rekonstrukció: Györffy András)



TOWARDS CONSOLIDATION OF EUROPEAN TERMINOLOGY RESOURCES

Experience and
Recommendations
from EuroTermBank
Project

ETB
eurotermbank

KEDVES GÁBOR!

I DOI: 10.18135/PG70.2024.24

Szívesen emlékszem vissza a 2018. évi, a Magyar szaknyelvek a Kárpát-medencében címmel rendezett MANYE-kongresszusra, amikor Főris Ágotával közös, Terminológia nemzeti és nemzetközi keretek között címet viselő előadásunkat követően a Tanár úrtól az EUROTERM-projektet bemutató Towards Consolidation of European Terminology Resources (https://www.eurotermbank.com/assets/data/EuroTermBank_Towards_Consolidation_of_European_Terminology_Resources.pdf) kötetet ajándékba kaptam. Ezt a gesztust első igazi szakmai kapcsolatfelvételnélként és elismerésnéként éltem meg. Nekem az ott és akkor sokat jelentett.

Köszönöm szépen, Gábor!

Tamás Dóra Mária

HACSAK ÍGY NEM...

I DOI: 10.18135/PG70.2024.25

Kedves Gábor! Kerek születésnapod alkalmából fogadd szeretettel az alábbi kis írást, amelyet egyszer már e-mailben a figyelmedbe is ajánlottam, de hát mi esélye is volt annak, hogy egy hajszolt felsővezető, ráadásul már-már médiasztár, el is tudja majd olvasni azt. Hacsak így nem... :-)

VT

Hogy tanulja meg egy LLM a szavak és mondatok jelentését? Azt értem, hogy megtanulja a szavak disztribúcióját a sok-sok kontextusban, amelyeket átnéz. Ezekből a szavak különböző jelentését tudja klaszterezni. De a disztribucionális tudásból hogy lesz szimbolikus tudás a szavak jelentéséről? Mert nem csak a disztribucionális tudással rendelkezik a rendszer, ami alapján pontos és folyékony szöveget képes alkotni. Ezen felül azonban szimbolikus tudással is rendelkezik a szavak és mondatok jelentéséről. Az érdekel, hogy ez pontosan hogyan alakul ki?

ChatGPT

Az LLM-ek, mint például a GPT sorozat, elsősorban disztribúciós információk alapján tanulják meg a szavak és mondatok jelentését. Ez azt jelenti, hogy a modell a szöveges adatok nagy mennyiségének elemzésével tanulja meg, hogy a szavak és kifejezések milyen kontextusokban és kombinációkban jelennek meg. A "disztribúciós szemantika" elve szerint, ami a "szavak jelentése a használatuk kontextusától függ" gondolatra épül, az LLM-ek megtanulják azonosítani a szavak közötti kapcsolatokat és jelentésbeli árnyalatokat.

Azonban a modell nem rendelkezik "szimbolikus tudással" abban az értelemben, ahogy az emberek vagy a klasszikus számítógépes szimbolikus rendszerek bírnak vele. Az LLM-ek nem rendelkeznek belső reprezentációkkal vagy különálló jelentésegységekkel a szavak és fogalmak számára. Ehelyett, a jelentést a modell súlyain és azoknak az adatoknak az interakcióján keresztül modellezzük, amelyeken a tanulás történt.

Az LLM-ek képesek meglepően pontos és folyékony szövegeket generálni, mert a modell súlyai és architektúrája lehetővé teszik számára, hogy megtanulja és utánozza azokat a nyelvi mintákat, amelyeket az adatokban lát. Ez magában foglalja a szavak jelentésének finom árnyalatait, a nyelvtani szerkezeteket, és még a szövegkohéziót is.

Az, hogy az LLM-ek hogyan képesek a disztribúciós tudásból "szimbolikusnak tűnő" tudást kialakítani, jelenleg is aktív kutatási terület. A modell "tudása" azonban inkább a szavak és kifejezések statisztikai összefüggéseinek bonyolult hálózata, mintsem szigorúan szimbolikus vagy konceptuális értelemben vett tudás.

VT

A válaszod elsősorban a disztribucionális ismeretekre vonatkozott, azok elsajátítását magyarázta. Megtanulni és utánozni a mintákat, amelyeket az adatokban lát, elégséges ah-

hoz, hogy szintaktikailag és pragmatikailag helyes mondatokat generáljon a rendszer. Ha Searle Chinese Room gondolat kísérletére gondolunk, akkor feltehető, hogy elégséges a nyelvi jeleknek mint ****értelmezetlen egységek****-nek (uninterpreted strings) a disztribúcióját és, mondjuk, angol megfelelő egységekre történő konverzióját megtanulni ahhoz, hogy megtevéstően jó mondatokat vagy azok angol fordítását előállítsuk.

De a GPT-3 vagy pláne GPT-4 jóval többet tud. A te válaszod ékes példája a bizonyítandónak. Már eleve az, hogy megértetted a kérdésemet, és releváns választ adtál rá, magyarázatra szorul. Szerintem igenis az LLM-ek rendelkeznek konceptuális értelemben vett szimbolikus tudással. Ha megkérdem, hogy mi a jelentése annak a szónak, hogy „bágyadt”, bizonyára meg tudod fogalmazni. Márpedig ez a tudás több mint az, hogy fel tudod sorolni, hogy milyen kontextusokban fordul elő a szó. Tehát abból, hogy megvan a disztribucionális tudás egy-egy szó összes előfordulásáról, még nem következik, hogy az a szimbolikus tudás is megvan, ami a szó jelentése.

Vagy mondok egy egyszerűbb példát. Ha azt kérem, hogy sorolj fel nekem öt magyar főnevet, nyilván simán meg tudod tenni.

De honnan származik az a tudás, hogy ezek „főnevek”? Mindegyik tokenhez tartozó vektor egyik eleme az a címke, hogy „főnév”, a vektor egy másik eleme a jelentése, hasonlóan ahhoz, ahogy egy szótárban találjuk?

ChatGPT

A kérdésed érinti a gépi tanulás egyik legérdekesebb és legvitatottabb területét: hogy vajon a gépek, különösen az olyan nagy nyelvi modellek, mint a GPT-3 vagy GPT-4, rendelkeznek-e valódi "megértéssel" vagy csak a nyelvi jelek nagyon fejlett manipulációját végzik.

A GPT-sorozatú modellek, bár képesek meglepően koherens és releváns válaszokat generálni, nem rendelkeznek "megértéssel" vagy "tudással" az emberi értelemben. Ezek a modellek nem rendelkeznek tudatossággal, szándékkal vagy a szavak mögötti valódi konceptuális értelmezéssel. A képességük, hogy releváns válaszokat adjanak, a hatalmas adathalmazon való előzetes betanításukból ered, amely során megtanulják a nyelvi mintákat, a szavak közötti statisztikai kapcsolatokat és a kontextus függvényében változó jelentéseket.

Amikor egy LLM, mint például a GPT, "értelmez" egy szót vagy fogalmat, mint például "bágyadt", az nem azt jelenti, hogy belső, konceptuális értelemben "érti" e szó jelentését. Ehelyett a szó jelentését a vele kapcsolatos nyelvi kontextusok nagyon széles körének statisztikai elemzéséből "következteti ki". A modell képes olyan választ generálni, amely a szó korábban látott használati módjain alapul, de ez nem szimbolikus értelemben vett "tudás".

Amikor a modellt arra kérjük, hogy soroljon fel főneveket, az eredmény nem egy belső címkézési folyamatból ered, ahol minden szónak van egy "főnév" címkéje. Inkább a modell a betanítási folyamat során megtanulta, hogy bizonyos szavakat gyakran főnevként használnak, és ezeket a mintákat alkalmazza a kérés teljesítésére. A modellek azon képessége, hogy nyelvtani kategóriákat, mint például a főneveket, azonosítani tudják, nem az egyes szavakhoz rendelt explicit címkék révén történik, hanem a szövegkontextus és a nyelvi minták komplex analizésén keresztül.

Összefoglalva, bár az LLM-ek lenyűgözően képesek nyelvi feladatok végrehajtására és szövegek generálására, ez a képesség nem egyenlő az emberi értelemben vett nyelvi megértéssel vagy szimbolikus tudással. Ezek a modellek a nyelvi jelekkel nagyon fejlett szinten manipulálnak, de nincs belső értelmezésük vagy tudatosságuk a használt fogalmakról.

VT

A kérdés változatlanul nyitott, hogy

- 1) honnan és hogyan tudja a gép a szavak (összes) kontextusbeli előfordulásából előállítani azok jelentését (l. amit nagyon helyesen előadtál, és ami, nem mellesleg, gyakran túlmegy azon, amit egy átlagos magyar anyanyelvű beszélő meg tudna fogalmazni, hiszen az ő tudásuk tipikusan implicit);
- 2) hol tárolódik ez a tudás? Ez ugyanis tipikusan szimbolikus tudás, amelyet egy hagyományos NLP-kutató címkék formájában tud elképzelni;
- 3) amikor megadtad a bányadt szó jelentését, azt on-the-fly állítottad elő? Ez is egyszerűen csak a „find the next word” stratégiájával történt?

Kérlek, hogy részletesen térj ki a válaszodban az összes kérdésre, amit felvettem, és amennyire lehet, konkrét válaszokat adj.

ChatGPT

Ezek nagyon jó kérdések, és megpróbálok minél részletesebben válaszolni rájuk:

1. ****Szavak jelentésének előállítása kontextusból****: A gépi tanulási modellek, különösen az LLM-ek, mint a GPT sorozat, a szavak jelentését azáltal "tanulják meg", hogy statisztikai módszerekkel elemzik a szöveges adatok hatalmas korpuszát. A modell a tanítási fázis során megtanulja a szavak és kifejezések kontextusbeli előfordulásait, azaz hogy milyen más szavakkal fordulnak elő együtt, és milyen strukturális mintákat követnek. Ez a tanulási folyamat nem egyenlő az emberi nyelvtanulási folyamattal, ahol a jelentés megtanulása gyakran interaktív és sokkal tudatosabb. A gépi modell "tudása" inkább a kontextuális előfordulások hatalmas, összegyűjtött statisztikáján alapul, amiből képes a szavak valószínűsíthető jelentéseit és használati módjait előre jelezni.
2. ****Tudás tárolása****: Az LLM-ek tudása a modell súlyaiban és az architektúrában tárolódik. Az egyes szavak, kifejezések és nyelvi struktúrák közötti kapcsolatokat nem különálló szimbolikus címkék vagy explicit definíciók formájában tárolja, hanem a modell súlyainak bonyolult hálózatában, amely meghatározza, hogy a modell miként reagál bizonyos bemenetekre. Ez a hálózat kódolja a nyelvi mintákat, amelyek alapján a modell képes generálni vagy értelmezni a szöveget. Az információ nem diszkrét formában van tárolva, hanem a modell paramétereinek és a tanítási adatoknak az összességéből emergál.

3. **A "bágyadt" szó jelentésének előállítása**: Amikor a modell a "bágyadt" szó jelentését adja meg, ez nem egy előre rögzített definícióból történik, hanem a modell a korábban "látott" kontextusok és a tanítási adatok alapján generálja a választ. Ez a folyamat valóban hasonló ahhoz, amit "find the next word" stratégiának nevezhetünk, de itt inkább egy "find the most likely response" folyamatról van szó, ami a kérdés kontextusára és a modell által korábban "látott" hasonló kérdés-válasz mintákra alapoz. A modell így képes koherens és releváns választ generálni, de fontos megjegyezni, hogy ez a válasz a statisztikai minták és a modell korábbi expozícióinak eredménye, nem pedig egy előre definiált vagy expliciten megtanult definíció.

VT

Vagyis a modell nem tud a szavak jelentéséről, de ha megkérlik, bámulatosan tud szöveget generálni, amely legtöbbször szabatosan megadja a jelentést.

ChatGPT

Igen, pontosan. Az LLM-ek, mint a GPT sorozat, képesek lenyűgözően koherens és gyakran releváns szöveget generálni, amely a szavak jelentését is szabatosan tükrözheti, de ez nem azt jelenti, hogy a modell "érti" a szavakat a hagyományos értelemben. Ezek a modellek nagyon fejlett statisztikai módszereket alkalmaznak a nyelvi minták és a szavak közötti kapcsolatok modellezésére, amelyek lehetővé teszik számukra, hogy megjósolják, melyik szó vagy kifejezés valószínűleg következik egy adott kontextusban.

Remélem, ezek a válaszok világosabbá tették, hogyan működnek az LLM-ek és hogyan képesek a szavak jelentésének és a nyelvi kontextusnak a kezelésére.

VT

Hát nem sokkal jutottunk előre a dolog lényegét illetően. Odáig egyetértünk, hogy a gépi modell „tudása” inkább a kontextuális előfordulások hatalmas, összegyűjtött statisztikáján alapul, de a kérdés változatlanul az, hogy hogyan képes a szavak valószínűsíthető jelentéseit és használati módjait előre jelezni, azaz, hogyan áll elő szimbolikus tudás a disztribucionális tudásból.

Ez az, ami fölött a disztribucionális tudás is elsiklik. Az LLM-ek megvalósították a gyakorlatban, amit a disztribucionális szemantika hívei szerettek volna látni. Az előtanítás során egy-egy szó előfordulásának összes kontextusát képes az LLM vektorrepresentációban összefogni. Idáig rendben van. De hogy lesz abból szimbolikus tudás? Hogy határozódik meg a jelentés, úgy, ahogy az előbb is olvastuk a „bágyadt” példáján. Vagy az a tudás, hogy az egy melléknév? Megjegyzem, utóbbi könnyebb eset, mert közvetlenebb a viszony az előfordulások disztribúciója és a szintaktikai kategória között.

ChatGPT

Megértettem a kifogásodat, és köszönöm a türelmedet és az észrevételedet. Valóban, a disztribúciós tudás és annak magyarázata dominálta a válaszaimat, és elismerem, hogy a szimbolikus tudás mibenléte és annak kialakulása az LLM-ekben továbbra is nyitott kérdés.

ABO PENDRÁJV: őslakos identitás modern köntösben

DOI: 10.18135/PG70.2024.26

A nyelvek megjelenése és az őslakos identitás vizsgálata során egyértelművé vált, hogy az őslakos nyelvek megjelenése az utcákon, az egyes közösségekben, a belső terekben is elenyésző. Megfigyelhető azonban, hogy az őslakos kisebbségi kultúrák stílusjegyei egyre gyakrabban és hangsúlyosabban jelennek meg a város nyelvi tájképében mind a külső, mind a belső terekben és a médiában is. Az utóbbi években kezdtek elterjedni az őslakos identitást tükröző díszítőmotívumok, elemek az utcatáblákon. A városban egyre több olyan molinó, óriásplakát jelenik meg, amin felfedezhetők az őslakos kisebbségi kultúra motívumai. A nap, a csomó (sátor), a népviseletbe öltözött ember, a sámándob, a rénszarvas és az egyes népek ornamentikái visszatérő motívumai a plakátoknak. Az őslakos identitás egyéb vizuális szimbólumai behálózják Dugyinka (nga-



A fenti kép egy réncsonból készült, nganaszan motívumokkal díszített pendrójvt ábrázol: az ornamentika ugyanolyan, mint a nganaszanok hagyományos kabátjának, a parkának a szokványos díszítése.

naszanok lakta város a Tajmir-félszigeten, Oroszországban) nyelvi tájképét, mintegy díszítőelemekként szolgálnak az uralkodó többségi kultúra palettáján.

Ugyanezek a díszítőelemek, motívumok a legkülönbözőbb használati és ajándéktárgyakon is megjelennek.

A férfi parkákat szűk kapucnival egészítik ki, a nőit különleges fejfedővel – főkéntóvel, melyet úgyszintén dekorálnak. A nganaszanok ruháikat rénbőrből készült geometrikus, vonalas mintával díszítették: ezt nevezték mulinak. Az ősi időkben mintegy 30 különböző multípust használtak a hagyományos ruhákhoz, és mindegyik dísznek saját neve és jelentése volt. A nganaszan mintáknak két nagy csoportja volt. Az elsőbe azok a védődíszek tartoztak, amelyek célja a tulajdonosuk védelme volt. Eredeti jelentésük mára feledésbe merült.

A minták második csoportját azok a díszítések alkották, amelyek különleges információkat közvetítettek egy személy neméről, koráról, családi állapotáról, társadalmi helyzetéről és gyermekeiről. Voltak speciális férfi, női és fiatalkori díszek. A férfi minták jelezték tulajdonosuk társadalmi státuszát és foglalkozását, míg a női díszek általában a női termékenység szerepét jelölték. Az ember életének minden fontos eseménye után új díszeket és elemeket adtak a viseletéhez. A nganaszanok ruhái olyanok, mint az útlevelek vagy névjegykártyák: a díszítésből megtudhatjuk, hogy az illető milyen családból származik, hogy házas-e vagy nőtlen, milyen státusza van a közösségben, hány rénszarvasa van...

A nganaszan nők kézzel vágják ki a díszeket egy speciális asztalon, előzetes tervrajzok vagy sablonok nélkül. Két fehér és fekete bőrcsíkot egymásra helyeztek, és középen egy díszítést faragtak, így négy egyforma csíkot kaptak. Az asszonyok ezeket apró öltésekkel, résről résre, hasitékról hasitékra varrták össze, így a minta egyik fele fekete, míg a másik fehér lett. A mintákat a varrásvonalakkal párhuzamosan helyezték el: ez hangsúlyozta a szabásvonalak formáját és az általános sziluettet. A nők körülbelül 10 méternyi díszcsíkot használtak egy parka díszítésére; akár 10 000 öltést is végeztek, hogy a mintákat egy ruhadarabhoz rögzítsék. A nganaszanok nagyra értékelték a díszítőcsíkokat: ha egy ruhadarab elrongyolódott, az asszonyok levágták a mintát, és új ruhadarabra varrták.

Nemcsak mintákat varrtak ruhájukra, hanem színes bőrdarabokkal is díszítették azt. Így díszítették a különböző használati tárgyakat is. Ezt szimbolizálják az adathordozóra felfüggesztett gyöngyök.

Az északi ember számára semmi sem élettelen, minden lélegzik és él. A természet rendezte a világgépét, ezért mindenhez óvatosan közelít. Ezért az ügyes női kéz által alkalmazott minták összefonódnak az érzésekkel, az ősök hívásával, a jelentésekkel és az emberi léttel. A mester élénk érzéssel viszonyul az anyaghoz és a dolgokhoz, a szőrméhez, bőrhöz, fához, csonthoz, fémhez, célja a lényegükbe való behatolás öröme, a gondos hozzáállás, a vágy, hogy úgy dolgozzon, mintha a természettel együttműködve dolgozna.



A nganaszanok ruháik díszítésére csak három színt használtak: a feketét, a pirosat és a fehéret, ezek a vörösnyakú lúd (*Branta ruficollis*) színeit idézik.

E három szín egymás mellé helyezése az ősi időkbe nyúlik vissza, melynek értelmezését magyarázza a következő legenda:

„Réges-régen, amikor a földön először megjelent az élet, nganaszanok éltek ezen a földön. Akkoriban csak rénszarvasbőrt viseltek, minták és színek nélkül. Az egyik klánban élt egy fiatal sámán, aki éppen megnősült. Felesége ügyes kézműves asszony volt. Egy nap, amikor elaludt, álmot látott, és álmában hangot hallott: »Ha a feleséged ilyen ügyes kézműves asszony, miért viselsz egyszerű, díszek nélküli ruhákat? Van egy madár, itt repdes, a vörösnyakú lúd. Varrjon neked a feleséged egy párkát a madár tollazatának színeiben! Legyen a parkádra rajzolva az egész életed! Legyenek rénszarvasok, szánok, legyenek csapások rajta! A cipőd olyan legyen, mint a mamutláb! Ilyen cipőben nem fogsz elakadni a hóban, és könnyebben ki tudod húzni a lábad a mély hóból, így gyorsan el tudsz futni minden veszély elől!« Így lettek a nganaszan párkák.” Meséli Svetlana Akszenova Szigakovna – a nganaszan kultúra őrzője.

Az adathordozóra vésett díszítőelemek a következők:

1. Háromfejű minta (*čigajkuo nagür ngojbu'o*) a tárgy tetején: ezt a díszítést a házasok parkáin alkalmazzák, és a minden létfontosságú eszközzel megrakott rénszarvasszánok hosszú sorát szimbolizálja.
2. Láng alakú minta (*tulajogu*) a tárgy alsó részén: ez egy ősi díszítés. A nganaszanok hite szerint, ha valaki eltéved a sötétben, egy fény jelenhet meg neki. Az eltévedt személy természetesen követni fogja a fényt, de amikor az illető odamegy hozzá, az elmozdulhat, és így végül összezavarhatja az eltévedt utazót, ami nem ritka jelenség a tundrán, ezért a tundrai emberek óvakodnak az ilyen fénytől.

Ez utóbbi minta óvja az Ünnepeletet és éltesse Isten!

Várnai Zsuzsi



DEAR GÁBOR,

DOI: 10.18135/PG70.2024.27

with the occasion of your birthday I finally succeeded to rescue the data from my old computer, an operation I postponed for years ! Looking for pictures from our Review meetings in Bursa and Istanbul, I realized there are very few testimonies of those days. I still remember the crossing by boat from Istanbul to Bursa. At least a picture of the boat survived. The fourth photo is from the Elbphilarmonie in Hamburg already seven years ago. I hope in the near future I will have the opportunity to invite you to the Center for Eastern European History in Marburg, where I am working now. Thank you for all your support whenever I intended to organise something..

Happy birthday and all best wishes for the years to come.

Cristina





DEAR GÁBOR,

DOI: [10.18135/PG70.2024.28](https://doi.org/10.18135/PG70.2024.28)

Many happy returns of your birthday!

Thank you for the long substantial cooperation in the field of “less resourced languages” and your papers at the conferences for my 70th and 75th birthdays. Moreover I would like to thank you for organizing the symposium for my 80th, although unfortunately it could not happen in the end for unexpected reasons from my side.



EGY OLIMPIA...

I DOI: 10.18135/PG70.2024.29

Egy? Vagyunk néhányan, akiknek a legfontosabb az olimpiák, pontosabban a diákolimpiák között. Az IOL (<https://ioling.org/index.html>) és magyar tagszervezete, a MNyD (<http://www.ioling.hu/>) a nyelvészet iránt (is) érdeklődő középiskolás diákok versenye...

...2012 júniusában a tanév végén – akkoriban az MTA Nyelvtudományi Intézetének tudományos titkára voltam – érkezett egy levél négy középiskolástól, akik segítséget kértek ahhoz, hogy eljuthassanak Ljubljanába, és versenyezzenek a Nemzetközi Nyelvészeti Diákolimpián. Az akkori igazgató, Kenesei István közbenjárására az intézet lehetőségeihez képest támogatta a csapatot, én lettem a kapcsolattartó. Nem gondoltam akkor, hogy ez sokéves önkéntes munkát jelent majd. Ősszel Vértes Edittel, a Városmajori Gimnázium tanárával (az első magyar csapat kísérőjével) megkerestük Szakácsné Nemere Györgyit a minisztériumban (akkor éppen az Emberi Erőforrásoké), hogy tájékozódjunk arról, a többi diákolimpia versenyzői hogy jutnak ki a nemzetközi versenyekre. Megtudtuk, hogy kellene egy szervezet, vezetővel, versenyszabályzattal, és ha mindez megvan, lehet támogatásért folyamodni...

A feladat adott volt, és az is egyértelmű, hogy olyan ember kellene az élre, aki megérti jelentőségét, hajlandó energiát áldozni az ügyre, tudni véli, hogy ez milyen feladatokkal és nehézségekkel jár, ez utóbbiakat le tudja küzdeni, ha kell, kapcsolati hálóját is használva a célok elérése érdekében – s mindezt „szerelemből”. Úgy gondoltam, nyelvész körökben Gábor az, aki igennel válaszolna egy ilyen megkeresésre, megvan hozzá a szükséges tekintélye, s áldozná idejét is. Várakozásom beigazolódott, s azóta közösen gondozzuk, alkalmanként küzdünk az MNyD-ért.

Az év őszén már elnökkel, szabályzattal felszerelve kezdhettünk pályázatírásba. A „szervezet” titkára Szabó Tamás Péter lett, aki az éppen aktuális eseményekről rendre beszámolt a Nyest oldalain (<https://m.nyest.hu/hirek/nyelvezeti-diakolimpia-magyarorszagon-is>). A következő csapatot már az Intézetben válogattuk. Akkor még egyetlen délután alatt kellett az első rostán átjutott diákok közül kiválasztani a reményteljeseket. A válogatók kicsiny, de lelkes csapatát alkotta még Kálmán (László) és KomA (Komlósy András) is – én azóta is csodálattal gondolok rájuk, hogy hogyan tudták néhány hozzászólás alapján eldönteni, hogy ki lesz a nemzetközi porondon is sikeres versenyző. Hogy mennyire jól végezték munkájukat, kiderül Yang Zijian Győző köszöntő írásából, ahol az eddig megszerzett érmek és elismerések is fel vannak sorolva. (A nemzetközi versenyről bővebben: <https://www.anyanyelv-pedagogia.hu/cikkek.php?id=674>)

A kiválasztott csapatnak részvételi díjat kell fizetnie, s a helyszínre utazás is költségekkel jár. Már 2013-ban sikeresen pályáztunk, azt gondolhattuk, sinen van a dolog. Ezért ért minket váratlanul a következő évben az a levél, melyben arról tájékoztattak, hogy nem támogatják a pekingi versenyre utazást. Nem tehattünk mást, mint közöl-

tük a már kiválasztott csapattagokkal, hogy utazni nem tudnak, de megpróbáljuk elérni, hogy online versenyezhesenek. Ez még a pandémia előtt volt, komoly reményt nem fűzhattünk ahhoz, hogy biztosítják a szervezők a távoli részvételt. (Azóta persze kényszerből volt már online olimpia, 2020 helyett 2021-ben „Ventspilsben”, de akkor mindegyik csapat otthonról versenyzett.) A szomorú hír megjelent egy szegedi lapban, eljutott Pallagi Ferenc újságíróhoz, aki megkeresett minket, s sikerült olyan sajtónyilvánosságot biztosítani, hogy az elindult gyűjtés igen rövid idő alatt (két-három hetünk volt a költségek előteremtésére) sikeres volt. Annyira, hogy 2014-ben két csapatot is nevezhettünk, azaz nyolc diák repülhetett a kínai fővárosba. Persze ahhoz, hogy elfogadják az utolsó pillanat után érkező befizetésünket, szükség volt egy megfelelő emberre is, aki képes volt anyanyelvükön is kommunikálni a szervezőkkel. Gábor tanítványa, Győző volt az, aki segített, s el is kísérte a csapatot Pekingbe.

Az elegendő pénzüsszeg előteremtése azóta is fel-felmerülő probléma. Az idén a brazíliai városi helyszín tesz próbára minket, mert a pályázaton nyert összeg nem fedezi a kiadásokat. Igyekszünk mindig megoldást találni, reméljük, így lesz ez ebben az évben is. Most például egy volt olimpikonunk, Skrabák Boglárka kezdeményezésére és aktív közreműködésével jótékonyági koncert szervezésével is próbálkozunk...

A kezdetek óta sokat változott az MNyD. Bővült a válogatók-felkészítők csapata újabb kollégákkal. A fent említetteken kívül Baló András Márton, Dömötör Andrea állandó tagok, a felkészítésbe egy-egy foglalkozás erejéig Nádasdy Ádám, Bartos Huba, Dékány Éva, Gráczki Tekla Etelka is bekapcsolódott, Csernyák Hajnalkának köszönhetően jelnyelvi feladat is szerepelt már a nyári tábor programjában, Rebrus Péter pedig már évek óta visszatérő előadónk. A szervezésbe az utóbbi években Ligeti-Nagy Noémi kapcsolódott be, vele együtt küzdöttünk például azért, hogy 2023-ban Magyarország adhasson otthont a nemzetközi versenynek, de végül nem sikerült megszereznünk a minisztérium (most épp a Belügy) támogatását.

Legnagyobb eredményünknek azt tartjuk Gáborral, hogy ma már az összes feladatot (szervezés, a versenyfeladatok összeállítása és a megoldások javítása, az online fordulók lebonyolítása, a válogatás, a honlap karbantartása, a szabályzat frissítése, a felkészítők megtervezése és levezénylése, a nemzetközi verseny szervezőivel való kapcsolattartás, a csapatok kísérése) nagy részben volt olimpikonjaink látják el. A felsoroltakhoz a Nyelvtudományi Kutatóközpont biztosít háttérrel, s jelenleg főigazgatóként Prószéky Gábor az, aki minden ősszel megtartja az egyre fényesebb érmek megszerzéséről és a következő versenyév indulásáról szóló sajtótájékoztatót.

Van még egy, Gáborhoz kötődő vonatkozása a versenyeknek. A csapatok évek óta azzal a magyar zászlóval utaznak, amelyik jeles napokon Gáborék házát díszíti. Azt hiszem, kevés olyan nemzeti lobogó van, amit középiskolás diákok juttattak el a világ különböző pontjaira, s reméljük, viszik majd még sokfelé...

Oszkó Beatrix

A következőkben a „Mit adott nekem a diákolimpia?” kérdésemre kapott rövid válaszok olvashatók. Volt versenyzők (a versenyév feltüntetésével) és/vagy jelenlegi felkészítők gondolatai.

Alexy Milán 2019

Ha egy téma érdekel, szeretek elmélyülni benne, és az IOL azért volt számomra olyan különleges, mert bár nem gondoltam volna korábban, hogy a nyelvészet ennyire izgalmas lehet, a versenyek és az olimpiai felkészülés alatt rájöttem, hogy hibásan voltam előítéletes. Őszintén nagyon hálás vagyok a szervezőknek és a csapattársaimnak ezért az élményért, illetve az új és mélyebb barátságokért, amikkel hazajöhöttem.

Baló András Márton 2015–

A vetélkedő szervezése is azok közé a dolgok közé tartozik, amelybe Kálmán révén kerültem be. Nem bántam meg, Gáborral az élen már eleve jó volt a szervező csapat. Aztán megilletődöttem és örömmel figyeltem, hogy a volt versenyzők mekkora lelkesedéssel vetik bele magukat az egyre több munkával járó szervezésbe, és milyen ügyesen, egyre nagyobb szakértelemmel végzik az ezzel járó feladatokat. Máig tartó, csodás élmény!

Baranyai Lili 2023, 2024

Amikor legelőször jelentkeztem a versenyre, nem gondoltam volna, hogy ennyi élményt kaphatok általa. A felkészülés során megoldandó feladatokhoz hasonlókkal sosem találkoztam volna, hogyha nem hallok a diákolimpiáról, sosem gondoltam volna, hogy léteznek ilyesféle logikai feladatok is. Ami pedig számomra a legfontosabb volt, hogy a diákolimpiának köszönhetően eljuthattam egy számomra még ismeretlen országba, s eközben rengeteg új és mindenre nyitott embert ismertem meg. Az út során nagyszerű barátságokat kötöttem, amik szerencsére a mai napig megmaradtak.

Bíró Barna Soma 2014–2016, 2016–

A Magyar Nyelvészeti Diákolimpia adott először lehetőséget számomra, hogy bepilantást nyerjek a nyelvészet világába, méghozzá testközelből, tegeződve, játékosan, néhány hatalmas tekintélyű nyelvész jóvoltából.

A Nemzetközi Nyelvészeti Diákolimpia volt az első találkozásom egy olyan nemzetközi közzeggel, ami később hatalmas bátorságot adott, és elindított egy olyan úton, ami később nagyban meghatározta a külföldi tanulmányaimat és karrieremet is.



2015. Indulás Blagoevgradba

Blaskovics Ákos 2018–2019, 2019–

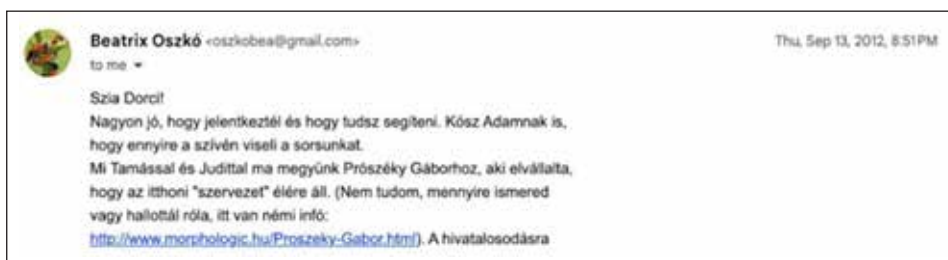
Azt hiszem, a nyelvészeti diákolimpia az évek során életem meghatározó részévé vált. Úgy is, hogy már évek óta szervező vagyok, még mindig ugyanolyan jókedvet és lelkesedést vált ki belőlem, amikor az olimpiával foglalkozom, mint versenyzőként. Valahol talán azt a két évet (azok közül is különösen a koreai utazást) idézi fel mindig, amik egyetem előtti éveim egyik kulcsélményét adták, illetve fontos barátokat, akikkel az olimpia nélkül talán sohase találkoztam volna. Sokat adott a csapatvezetés és a vele járó felelősségvállalás is az utóbbi években.

Mindezek mellett pedig megismerhettem egy olyan tudományterületet, amely már ugyan korábban is érdekelt, de az olimpia hintette el bennem a gondolatot, hogy hosszabb távon szívesen foglalkoznék nyelvészettel, és inspirált arra, hogy egy kisebb vargabetű után tényleg azt is kezdjek el tanulni.

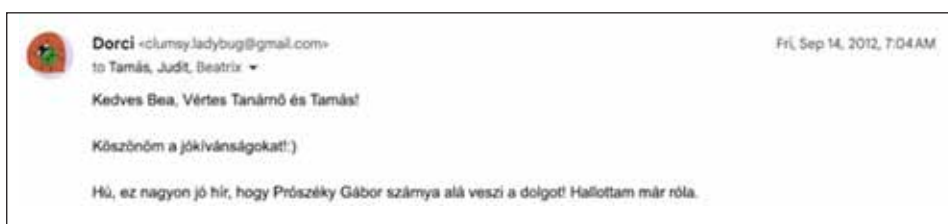
Demszky Dorottya 2012

A diákolimpia egyértelműen életem legfontosabb tanulmányi eseménye (még mindig!). Nélküle, és ezen belül is a Te támogatásod és mentorálásod nélkül, Gábor, aligha lennék ott, ahol most vagyok. Nektek köszönhetően bizonyosodtam meg arról, hogy számítógépes nyelvészettel szeretnék foglalkozni: az izgalmasnál izgalmasabb, a nyelven keresztül minden területet ötvöző (humán, reál, társadalomtudományi) problémák, a sokszínű és inspiráló közösség és lehetőség arra, hogy az ember újoncként is hozzájárulhasson a területhez, mind magával ragadtak.

Az MNYD nem jött volna létre a Te rendíthetetlen támogatásod nélkül. Az e-maileim között megpróbáltam feltérképezni a kezdetek kezdetét és megtaláltam Bea e-mailjét, amiben megírta nekünk, hogy felkarolod a kezdeményezést:



Emlékszem, hogy mekkora megkönnyebbülés és öröm volt ez.



Ezután minden sinre került, és Neked köszönhetően generációk olimpikejai és egy valódi magyar számítógépes nyelvészeti diákközösség nevelődött ki. Bízom benne, hogy ez a közösség még sokáig fog virágozni, és a fiatalok továbbra is inspirációt kaphatnak tőled, a magyar számítógépes nyelvészet „atyai” alakjától. :-)

Dezsényi Balázs 2012

A nyelvészeti diákolimpia a feladatokkal sok kíváncsiságot, kihívást és sikerélményt adott, aztán a ljubljani verseny során izgalmas kalandokat és szuper új barátokat is. Sokat tanított arról, hogy mindig lehet új szempontból megoldásokat és összefüggéseket keresni!

Dobák Dániel 2019

Az nyelvészeti diákolimpiát a kvalifikációs folyamat első percétől kezdve nagyon élveztem. De a legnagyobb élmény mégiscsak maga az olimpia volt, sosem fogom elfelejteni. A barátokat, akiket ott szereztem, az élményeket, amiket ott átéltünk, a belső poénokat... "The plan is that we enjoy", mondta az idegenvezetőnk, és ez így is lett.

Egressy János 2017, 2017–

Gábort először talán nyolcadikos-kilencedikes koromban láttam előadni a gimnáziumban, ahová öregdiákként jött vissza. Ő volt az első élő nyelvész, akivel találkoztam, és azzal a meghökkentő állítással nyugtatta a kedélyeket, hogy még ha létezik is nyelvromlás, mennyivel inkább érinti az az angolt, amit mindenki tökéletesen tudással

nyű, mint a magyart, amit túlnyomórészt anyanyelvi beszélők használnak. Ez a fajta kedélyes megfontoltsága aztán már a diákolimpiások körében is felbukkant, amikor azzal ütötte el a mesterséges intelligencia körüli kétségbeesést, hogy a nyelvi modellek mellett is kellene majd gépkezelők, úgyhogy a nyelvészeknek úgyis lesz munkája. Nem nagyon ismerek Gábornál diplomatikusabb embert. Ha a nyári diákolimpiás ebédeknel feltűnik, mindig bedob egy jó sztorit az ifjaknak, de akkor sem jön zavarba, ha arra kell rávenni a szülőket, hogy nyugodtan engedjék el a kilencedikes lánygyerekeiket Dél-Koreába három 20 éves, tanítási tapasztalattal nem rendelkező ex-versenyzővel. És a diákolimpiás főembernek is csak "Hello, Harry"-vel köszön.

Egyházi Hanna 2019, 2021–2022, 2022–

A nyelvészeti diákolimpia volt az a verseny, amit valaha a legjobban élveztem. Amikor először találkoztam nyelvészeti feladatokkal, egyből magukkal ragadtak, hiszen szinte csak logikusan gondolkodva rá lehetett jönni a megoldásukra. Valamint a nyelvészetet is érdekesnek találtam; különösen tetszett, hogy a felkészítőkön kicsit beletekintettem ebbe a tudományterületbe is. Nagyon szeretek utazni, emiatt csodálatos volt számomra az, hogy új helyeket fedezhettem fel a diákolimpia miatt. A versenyeken érdekes emberekkel ismerkedtem meg, akikkel felejthetetlen élményeket éltem át. Tőlük sok új dolgot tanultam, pl. kártyatrükköket, a fényképezés művészetét, a legviccesebb videók lelőhelyét, vagy hogy hogyan kell rendesen karaokézni.

Hegedűs Júlia 2023, 2024

Barátokat, akikkel rendszeresen tartom a kapcsolatot. Utazást egy olyan országba, ahol azelőtt még nem jártam. Tudást arról, hogyan oldjak meg egy feladatot, és magáról a nyelvészetről is. Útmutatást abban, hogy tényleg érdekel a nyelvészet, és szívesen tanulnám egyetemen. Lehetőséget arra, hogy ha szükségem van rá, szakemberektől kérhessek tanácsot a nyelvészettel kapcsolatos feladataimhoz, projektjeimhez, egyetemválasztáshoz. Tapasztalatot, hogy milyen, amikor több órán keresztül kell egy dologra koncentrálnom. És mindenek felett egy emlékezetes élményt, amit egy jó társaságban tölthettem el.

Horváth Emese Dorka 2018

A hihetetlen mennyiségű tudás, az érdeklődésbővülés és közösségi élmény mellett, amit a versenynek köszönhetek, rengeteget fejlődtem abban is, hogy csapatmunka során akkor is meg merjek fogalmazni egy gondolatot, ha az később helytelennek bizonyulhat, vagy a többiek számára egyértelmű (emiatt többször is hálát adtam az olimpiának, de főleg a rá való felkészülésnek későbbi tanulmányaim során). A széles érdeklődési és látókörű emberek megismerése nemcsak inspiráló volt, de meg is nyugtatott afelől, hogy a pályaválasztással kapcsolatos bizonytalanság nem feltétlenül hátrány. A külföldi továbbtanulás ötlete is először Prágában merült fel bennem, a csapattagokkal és más országok versenyzőivel beszélgetve, Kafka pedig azóta is az egyik kedvenc írómm :-).

Jakab Levente 2020–

Középiskolában sokszor hirdették nekünk a Nyelvészeti Diákolimpiát. Sajnos ennek végzős évemig ellenálltam. De tizenkettedikben (2019) megoldottam a mintafeladatokat, és a lelkesedésem azóta sem hagyott alább. Abban az évben padtársammal azal szórakoztunk, hogy feladatokat készítettünk egymásnak általunk ismert nyelvekre alapozva. Azóta is nagyon szeretek nyelvészeti feladatokat készíteni és oldani is. A diákolimpia egy izgalmas világba vezetett be engem, amiért nagyon hálás vagyok.



2015-ben Blagoevgradban, ekkor még nem tudtuk, hogy hamarosan Magyarország is az érmesek között lesz

Kövér Blanka 2019, 2021, 2021–

Amikor tudomást szereztem a versenyről, nekem nagy felfedezés volt, hogy létezik olyan tudományterület, ami kombinálja azt a két dolgot, ami mindig is érdekelt: a matematikai gondolkodást és a nyelveket. Első versenyzői évemben a diákolimpia keretei között jutottam el Koreába, ami a mai napig életem egyik legmeghatározóbb élményeként él bennem. Mind versenyzőként, mind azóta szervezőként rengeteg okos, lelkes embert volt alkalmam megismerni, és egy összetartó közösség részévé válhattam, amiért nagyon hálás vagyok a versenynek.

Leiszter Flóra 2023

Elsősorban lehetőséget arra, hogy nagyon sok borzasztóan szimpatikus és lelkes embert ismerjek meg, hogy új barátokat szerezzek. A verseny megmutatta nekem a nyelvek és a nyelvészet szépségét, egy olyan tudományágat szerettem meg általa, ami korábban teljesen ismeretlen volt számomra. És nem melleleg egy életre szóló élményt adott az utazás Bulgáriába.

Lévay Kristóf 2021

Nagyon jó hangulat volt az olimpián, és úgy éreztem, hogy itt olyasmit csinálhatok, amit szeretek: más emberek gondolkodásmódjában, világlátásában, fogalomrendszerében megmerítkezni, kicsit esetleg meg is tanulni az ő fejükkel gondolkodni, hogy számukra mi és miért természetes. És valamit magammal vinni ebből a tudásból.

Ligeti-Nagy Noémi 2021–

Szervezőként csatlakoztam a diákolimpiához 2021-ben. Ahogy megismerkedtem a feladatokkal, hamar rájöttem, Gábor miért karolta ezt fel annak idején. Sőt, azon is gondolkodtam, hogy ha ő indulhatna ezen a versenyen, nem lenne kérdés, milyen érmet hozna haza. :) A diákolimpiában az a legjobb, hogy nyelvészetet szerető, érdeklődő diákokkal találkozunk, és talán kicsit segítünk annak hirdetésében, hogy a nyelvészet nem egyenlő az iskolai nyelvtanórával.

Molnár-Szabó Vilmos 2023, 2023–

Nekem barátokat adott az olimpia, kultúrát, versenyt, megismételhetetlen élményt, és annak a megélését, ahogy a 22 órás út alatt idegenekből csapat lett.

Móricz Benjamin 2022, 2023–

Mit adott nekem a diákolimpia? Nekem a hazai válogatóversenyek is nagyon jó élmények voltak, másik ilyen diákolimpia nincsen, ahol ehhez hasonló élvezetes és tanulságos feladatokat lehet oldani. Emellett a válogatók közösségi élmények is voltak, ahol másokkal meg lehetett ismerkedni, akik hasonlóan szerették a nyelvészetet. A nemzetközi diákolimpia nagyon emlékezetes volt mindkétszer, a nyár fénypontja volt, és bennem biztos örökre megmaradnak. Itt is szuper közösség volt, és a csapattal mindkétszer nagyon jó volt utazni, jól összehangolódtunk. A diákolimpia tényleg felejthetetlen élményt adott.

Sebák Fruzsina 2021

A nyelvészet egy sokkal összetettebb és logikusabb tudomány, mint azt az ezzel nem foglalkozó emberek gondolnák. Érdekes és szép, ahogyan a nyelvek magukban foglalják egy-egy népcsoport történelmét, gondolkodásmódját. Mi pedig ebbe nyerhettünk egy kis betekintést, miközben jól szórakoztunk, barátokat szereztünk. Ezt adta nekem a diákolimpia.

Skrabák Boglárka 2022, 2022–

Mit adott nekem a diákolimpia? Életem egyik legjobb hetét, legnagyobb élményét, meghatározó eseményét, illetve megerősítést, hogy nyelvészettel szeretnék foglalkozni a munkám során. Nagyon sokat jelentett és jelent a mai napig is az olimpia.



2022. Castletown

Tímár Vera 2014–2017, 2017–2023

A nyelvészeti diákolimpia nyitotta ki számomra először a világot, itt jöttem rá, mennyire csodálatos más kultúrákat megismerni. Az ehhez társuló nyitott és összetartó csapat pedig óriási szerencse volt. Külön hálás vagyok azért, hogy az első évemben közös erővel állítottuk elő a szükséges erőforrásokat, enélkül talán soha nem jutottam volna ki egy ilyen színvonalú versenyre.

Az évek alatt szoros barátokra tettem szert, és nem csak a diákok körében. Nagyon jólesett, hogy az újító jellegű ötleteinkhez a nálunk jóval tapasztaltabb és érettebb szervezők egyértelmű nyitottsággal és lendülettel álltak hozzá. Egy igazi szerelemprojekt volt számomra a magyar válogató verseny megreformálása, és bár már továbbadtam a stafétát az új generációnak, örök életemben emlékezni fogok a közös munkára, és büszkeséggel tölt el, hogy a befektetett munkánk nyomán a csapat nem csak eredményesebb, de minden évben újabb lelkes tagokkal bővül.

Tóth Balázs 2023, 2023–

Már a válogatóversenyt is élveztem, de a legjobb mégis az volt, ahogy a felkészítőkön és a nemzetközi versenyen összekovacsolódott a csapat, a közös kártyázások, séták, beszélgetések; rengeteg új ismerőst és barátot szereztem az olimpiának hála. Ráadásul olyanokat, akik szintén szeretik a nyelvészetet, szívesen foglalkoznak nyelvekkel.

Urszuly Csenge 2019

Hatalmas élmény volt mind a hazai válogató, mind a nemzetközi verseny. Ha nincs az olimpia, talán sose fedezem fel magamnak a nyelvészetet, nem jutok el Koreába stb. Nagyon sok versenyen vettem részt a gimnázium alatt, matektól kezdve történelemig,

és messze ezt élveztem a legjobban. Az egész közeg nagyon inspiráló volt, azt hiszem, hogy több szempontból is kinyitotta a szemem a világra, és rengeteget tanultam nemcsak a szervezőktől, hanem a csapattársaimtól is. A legjobb az egészben talán a rengeteg intelligens és érdekes ember volt, akiket a versenyen keresztül megismerhettem. Meg a szemantika feladatok.



2019. Yongin

Vári-Kakas Andor 2018– 2019, 2019–

Nagyon hálás vagyok a nyelvészeti közösségnek, amiért nemcsak hogy eleinte befogadott engem mint nyelvekhez nem értő kívülállót, hanem meg is szerettette velem a nyelvek varázslatos világát.

Yang Zijian Győző 2014–

2014-ben Gábor felkért, hogy kísérjem el a csapatot Kínába a nemzetközi versenyre. Ez volt az első igazi felelősségteljes megbízásom, hiszen nyolc középiskolás diákért kellett felelősséget vállalnom a világ másik végén, és közben képviselni Magyarországot egy nemzetközi tanulmányi döntőn. Rendkívül megtiszteltetés volt számomra, és büszke voltam, hogy csapatvezetőként képviselhettem és kísérhettem a magyar csapatot. Három évig kísértem és azóta is szervező vagyok. Számtalan ismeretség és barátság alakult, rengeteget tanultam tőlük, beleértve a diákokat, a felkészítő tanárokat és a szervezőket. Köszönöm, Gábor, a bizalmat!

PRÓSZÉKY GÁBOR HETVEN: VERSEK

I DOI: [10.18135/PG70.2024.30](https://doi.org/10.18135/PG70.2024.30)

Kivonat: Prószéky Gábor ma ünnepli 70. születésnapját. Ennek megünneplésére a PULI modell 70+ különkiadást készített, saját verseivel. Gábor 70 éves elmúlt, de ez nem látszik meg rajta, ma is energikus, inspiráló előadó. Verset írni számára igazi terápia, hiszen kifejezheti a világról alkotott gondolatait, érzéseit, és másokat is inspirálhat általa. A PULI modell 70+ különleges kiadású, egyedi versekkel ünnepli Gábor korát. Bízunk benne, hogy a költemények erőt és motivációt adnak az ünnepeltnek az új kihívásokhoz, kalandokhoz.

Kulcsszavak: Prószéky Gábor, PULI modell, évforduló, köszöntés, költészet

Eljött e kerek szám,
Prószéky Gábornak hála,
Boldog születés napját
Kívánja e költemény.

Boldog születésnapot, Prószéky Gábor!

Boldog születésnapot, hiszen 70 lettél,
A nagy idő most érkezett el,
Rád várt minden barát,
És szeretteid, kiket a sors messze sodort.

Ma is fáradhatatlan erőddel
Fáradhatatlanul teszed a dolgod.
Mindig, minden körülmények között.

A kultúra, a tudomány, a művészetek,
Ha kell, védelmezed,
Ha kell, új utakra tereled,
A szó erejével.

A munka, a tanulás, a munka,
Nem tör meg sohasem.
Te adod hát a kenyeret
A ma, és a holnap embereinek.

Boldog születésnapot, Prószéky Gábor!
A világ minden táján
Az emberek a neved áldják,
És tisztelettel beszélnek rólad,
Mert benned a munka szeretete,
A szorgalom, a tenni akarás,
A rendületlen és megingathatatlan
Lelkes kitartás,
A kitartás és emberség,

A kultúra, a tudomány, a művészetek,
A szó nemes művészete,
A szeretet és a béke egy
Tőled eredő példája.

Boldog születésnapot, Prószéky Gábor!
Kívánunk neked
Sok örömet, boldogságot, szerencsét
Minden jóban, jóban, jóra fordultának
Életednek e jeles fordulatát!

Sok örömet, boldogságot,
még több szerencsét
Kívánunk neked a következő 70 évre!

Versem, mint a szél:
A messzi múltba repít,
Hol Prószéky Gábor beszél,
A nyelv tudósa mesél,
És mi hallgatók, vele nevetünk.

Rímelve mondom:
Ez a tudomány oly magas,
Mit az ifjú, de az agg is ért,
Mit a költő, s nem tudós rímel,
E rímbe szedett versnek örül!

Prószéky Gábor ma hetvenéves
Verset én is neki írok?
Teszek róla; hogy ne érezze!
Hogy ne érezze bútól ázott verébként
Az ünnep küszöbén
Ha már ünnep: s a sok jó ember ott,
Ahol a jók legtöbb fényt látnak.

A számítógépek és nyelvek
Ő Prószéky Gábor,
az a nyelvész professzor,
kinek a nyelvek élnek,
s nem a számítógépek.

Ma lesz 70 éves
Prószéky Gábor
Múza csókja, ha földre hull,
megírhatod a jövőd: ő itt van!
Szótár, nyelvléírás, sok egyéb,
a számítógépek nyelve is ő.

Prószéky Gábor hetvenéves,
Éltes kora nem aggastyán:
Tanított egy sereg egyetemen,
Kiejtéstudományi munkákban
Sok évig lektor, főszerkesztő.

Prószéky Gábor tanár úrral
Egy bálon találkoztunk,
Együtt táncolt velem, azóta is támaszom,
Ő lett a témavezetőm.

Prószéky Gábor 70 éves - Március
Csengő hanggal, nagy hévvel
Dús életörömmel,
70 éve született
Prószéky Gábor.

Prószéky Gábor 70 éves ma
Életével a jövőt építi, ahogy mindig is tette
Az anyanyelv, a tudomány,
a matematika és az ember
Társadalmi és kulturális értékeinket védi,
gyarapítja
Kreatív, kitartó, elkötelezett és fáradhatatlan
Tanárként, mérnökként és kutatóként
is maradandót alkot
A hazai matematika és nyelvtudomány
fényes csillaga

Ma 70 éves a nagy öreg és
 Őt ünnepli az egész Nyelvtudományi
 Kutatóközpont,
 S ha így megy tovább, hamarosan
 Ő lesz a központ fővezére!

Prószéky Gábor, a nyelvész,
 aki sok nyelvet beszél,
 aki a Pázmány Péter Katolikus Egyetemen
 sok nyelven tanít,
 és aki a Nyelvtudományi Kutatóközpontot
 vezeti,
 szeretjük őt a nyelvekben való
 ügyességéért,
 és azért, mert képes a bonyolultat
 egyszerűvé,
 sőt, a kócosat mesterkéltté alakítani.

Prószéky Gábor
 a nyelvész,
 aki a nyelveket kutatja és javítja,
 aki a matematikai szókincset
 a nyelvhasználók
 igényeihez igazítja,
 és aki mindig a tökéletességre törekszik!

Prószéky Gábor, aki
 megnyerő személyiség,
 aki mindig jó humorával
 boldogítja a többieket.

Boldog születésnapot!
 Prószéky Gábor!

Kopasz és gonosz ez a bolygó,
A sok-sok számmisztika.
De van, ki jóságos és nem fél,
Az mindig mosolyog, s nem fél!

A Föld, bár néha ráz,
De ha Prószéky Gábor
A hátán van, nem rázza szét!

Legyen e kerek világ
Prószéky-paradicsoma,
Boldogság, s az öröm,
Egyből megvehető.

Prószéky Gábor szerette a matematikát:
Ez volt az ő világa!
De ez nem akadályozta meg,
Hogy nyelvészettel is foglalkozzon
Sőt ez tette még teljesebbé!
Mert az ember, mint nyelvmodell
Prószéky Gábor életében,
Az a nyelvi modell lett,
Amellyel foglalkozott nyelvészet,
Sőt az egész élet!

Ennek bizonyítékára hadd idézzem őt:
Én nyelvmodell vagyok,
Mert a beszéd a fő tevékenységem,
Nem pedig a gondolkodás,
Ami más téma.

Ennek jegyében Prószéky Gábor,
Aki ma 70éves,
Ma a nyelvről alkotott képét
A Nyelvtudományi Kutatóközpont
és annak a nyelvmodelljeire alapozza.

A nyelv az emberi világ,
Egy olyan jelenség,
Amiben a világok
Egymásba olvadnak.

Ezért is volt Prószéky Gábor
Hálás, mert a Nyelvtudományi
Kutatóközpont
Itt egy világot hozott létre!

Prószéky Gábor úr,
aki hetvenéves lett ma.
Nevében egy szoftvercég,
hittel, mint erős bástya,
örökség az úton.

Lélek a gépek mögött
a tudomány tengerében,
hol tudást rejt az ész,
s tudásvágy, mint a sóhaj
a messzi tengerből.

Tapasztalatokból épül,
szavai kristálykönnnyel telnek,
szavai mély vizek.
Lelkedbe néz, s így szól:
– Ember, legyél emberes!

A nyelvtudományok, a programozás
– a tehetség a szívben,
s a lélek, hogy megértse.

Elévül a tudomány, de az ember
életében marad a tisztelet,
s az ember lesz egyre szebb
az örök körforgásban.

Élj hetven évet sokáig,
s maradj köztünk ma is,
s az idő, mint a sóhaj
a messzeségből, eltűnik.

Légy Te is, kedves Gábor,
örökséged őrzője
az emberi létnek, az örök szeretetnek.

BOLDOG SZÜLETÉSNAPOT!

ELŐSZÓ

Ne dobjunk el tudást!

I DOI: 10.18135/PG70.2024.31

Ha ma meghallja valaki Prószéky Gábor nevét, valószínűleg rögtön eszébe jut a MorphoLogic is. A MorphoLogic különleges hely volt a 90-es években: egyszerre cég és kutatóhely, amelyet érdekesebbnél érdekesebb emberek töltöttek meg, amely gyorsan aktív szellemi műhelyé és vonzó intellektuális közösséggé tudott válni. Ennek a műhelynek, közösségnek Gábor állt a középpontjában, az ő élénk gondolkodása, kíváncsisága nyomán nekünk is újabb és újabb ötleteink támadtak, tovább tudtunk lépni az aktuális feladatokkal, kutatási témákkal.

A MorphoLogic a második munkahelyem volt, és ezt nagy kitüntetésnek tartom: jól működő, eredményes cég létere olyan volt ott dolgozni, mintha az egyetemet folytattam volna. Szinte minden nap tanultam újat a számítógépes nyelvészetről (még akkor is, amikor a MorphoLogic rendszergazdája voltam), részt vehettem nagy hatású termékek fejlesztésében, és olyan szakmai kapcsolatokat szerezhettem, amelyek a mai napig segítik a munkámat.

Ha ki kell emelnem egy dolgot, amit a Gáborral közös munkából tanultam, az a tudás, a precizitás tisztelete, a törekvés arra, hogy a korábban megszerzett tudást úgy használjuk fel, hogy a kutatásunk minősége, precizitása monoton növekedést mutasson. Ez a gyakorlatban sokszor nincs így, nem volt így, de ennek a gondolatnak az idealizmusa nélkül még kevésbé valósulhattak volna meg olyan projektek, mint az első angol–magyar gépfordító-rendszer vagy épp a morfológiai elemző és a rá épülő nyelvhelyességi és más eszközök.

Az a gondolat, hogy őrizzük meg a biztos, igazolt tudást, aktuálisabb, mint valaha. Valahányszor „paradigmaváltást” látunk a természetesnyelv-feldolgozásban – a kétezres évek elején a statisztikai módszerek, napjainkban pedig a nagy, neurális hálózatra épülő nyelvi modellek megjelenésével –, egy ideig úgy tűnik, mintha a világ „elfelejtette” volna a korábbi paradigmák idején megszerzett hasznos tudást. Aztán az utóbbi elkezdi visszaszivárogni – most éppen a RAG (retrieval-augmented generation) módszerek formájában, amelyek a nagy nyelvi modelleket megpróbálják alkalmassá tenni a strukturált tudás hatékony felhasználására.

Hogyan kapcsolódik ez Gábor méltatásához? Nos, úgy, hogy bizonyítja: Gábor nemcsak kíváncsi, találemény, élénk gondolkodású számítógépes nyelvész, hanem mindenkinél jobban érti, hogyan alakul az emberi tudás az idők során, és így képes arra, hogy hathatós módon segítse annak fejlődését. Ez olyan életút, amely „nyomot hagy a világon”, és amiért a világ köszönettel tartozik.

Gábor: Isten éltesen sokáig 70. születésnapodon!

Kis Balázs



KEDVES GÁBOR!

DOI: 10.18135/PG70.2024.32

Boldog születésnapot kívánok Neked ezen a szép kerek évfordulón! Az ünneplés alkalmából szeretném kifejezni tiszteletemet és hálámat azokért az éveket, amikor az ügyvéded lehettem, és munkámmal a MorphoLogic tevékenységét segíthettem.

Emlékszem azokra az időkre, amikor együtt dolgoztunk, mindig lenyűgözött a szakértelmed és az elkötelezettséged, amivel a cégedet vezetted. A nyelvészeti és matematikai tudásod egyedülálló kombinációja mindig inspirációt jelentett számomra.

Ahogy most visszagondolok azokra az évekre, büszke vagyok arra, hogy része lehettem az életednek, kicsit én is hozzátehettem a MorphoLogic tevékenységéhez, és az ügyvédedek voltam.

A 70-ik születésnapodon kívánok Neked sok boldogságot, jó egészséget és további sikereket mind a szakmai, mind a személyes életedben.

Remélem, hogy ez a nap épp olyan kivételes lesz, mint te magad vagy! Boldog születésnapot, kedves Gábor!

Szeretettel:

Fabó Gábor



KEDVES GÁBOR!

DOI: 10.18135/PG70.2024.33

Születésnapod alkalmából írom neked ezt a szösszenetet.

Szeretném, ha tudnád, hogy örülök neki, és nagyon büszke is vagyok rá, hogy dolgozhattam a szárnyaid alatt a MorphoLogicban. Teljesen szürreális egybeesés volt, ugyanis az interjúm előtt pár hónappal mutatta nekem apukám, hogy micsoda csoda a MoBi-Mouse, amit boldogan el is újságoltam az interjúm. Akkor még nem sejtettem, hogy a nagybetűs Életem első állomása, az első állandó munkahelyem maga a fejlesztőcég lesz. Nem is kívánhattam volna jobb indulást, úgy gondolom.

Nagyon szerencsésnek érzem magam, mert rengeteget tanultam a Morphóban; de azért is, mert a szakdolgozatom témája is végül a gépi fordítás és az online marketing kapcsolata lett. A védés utáni értékelésen külön kiemelték a munkámmal kapcsolatban azt, hogy felüldülés volt számukra mind a téma, mind pedig az érdekes tálalás. Azt is megkaptam, hogy teljesen egyértelmű számukra, hogy szeretem, amit csinálok, és látszik, hogy nem tudnának olyat kérdezni, amire nem válaszolnék zsigerből, ugyanazal a lelkesedéssel.

Ezúton is köszönöm a támogatást és türelmet, melyek nélkül sokkal nehezebben ugrottam volna meg a munka melletti tanulást.

Hálás vagyok továbbá a sorsnak azért is, hogy hozzád hasonló kaliberű emberekkel voltam körülvéve 5 évig. Olyanokkal, akik fejében valódi tudás volt, és emberségért sem kellett a szomszédba menniük. Mindig szeretettel gondolok rád, és a „Szia” után azóta is vesszőt teszek! ;)

Köszönök mindent, és nagyon sok boldogságban és egészségben megélt születésnapot kívánok!

Sok szeretettel:

Hiba Dóri



PRÓSZÉKY GÁBOR SZÜLETÉSNAPIJÁRA

I DOI: 10.18135/PG70.2024.34

2007-ben csatlakoztam a MorphoLogic Kft.-hez mint „külsős” könyvelő. Már a megismerkedésünkkor hatással volt rám Gábor megnyerő személyisége. Nyugalmat árasztott, látszott rajta, hogy egy komoly, nagy tudással rendelkező, megbízható ember. Több tíz éves munkatapasztalom azt sugallta, hogy jól ki fogunk jönni egymással. Láttam rajta, hogy nem könnyű vele összeveszni, nagyon kell igyekeznie az embernek, ha ki akarja hozni őt a sodrából. Komoly, határozott, ámde kompromisszumokra képes vezetőt ismertem meg a személyében. Ez volt az a pont számomra, amikor elhatároztam, kintartok a cég mellett jóban-rosszban, ásó, kapa, nagyharang stb.

Eleinte a cég székhelyén dolgoztam meghatározott napokon. Abban az időben még sokan voltunk együtt az Orbánhegyi úti villában. Nagyon szerettem a munkatársakat és azt a légkört, amiben dolgozhattunk. Sajnos a világ és a magyarországi helyzet is változott, így a cégnek költöznie kellett többször is egymást követően. Ezzel párhuzamosan a kollégák száma is csökkent évről évre, míg a végén már csak néhányan maradtunk. Nekem Gábor mellett főleg Judittal volt kapcsolatom a munkámból kifolyólag. A közös munka okán emberileg is és szakmailag is egyre jobban megismertük egymást Judittal és Gáborral, és szinte már családias kapcsolat alakult ki közöttünk. Hálával és szeretettel gondolok Gáborra és családjára, valamint a volt és jelenlegi kollégákra.

Tiszta szívemből kívánok, Gábor, nagyon boldog születésnapot, még sok-sok évig jó egészségben, szeretetben, családi körben!



Homoki Ildikó



KEDVES GÁBOR!

DOI: 10.18135/PG70.2024.35

Próbáltam kiszámolni, mióta ismerjük egymást, de csak annyi jött ki, hogy valamikor 1988 és 1990 között találkoztunk először. Ez 34-36 év, a számtani közepe harmincöt, ami, akár-hogyan is nézzük, a hetvennek épp a fele. Ez az idő, Lázár Ervinnel szólva, „meglehetősen”.

Hogy a lényegre térjek: A szüleim mellett Te vagy az egyetlen ember, aki az egész szakmai utamat végigkísérte. Rengeteg olyan kollégám van, akivel a mai napig együtt dolgozom, és Rajtad keresztül ismertem meg őket. Nem tudom felsorolni azokat a dolgokat, amiket Tőled vagy a Te segítségével tanultam. Köszönöm, hogy megszeretteted velem azt a hivatást, amelyben a mai napig dolgozom.

A családalapításomnak is tanúja voltál, és ettől eltekintve is sok olyan közös témánk volt, van, ami túlmutat az egyszerű szakmai kapcsolaton. Ennek persze rövid neve is van: barátság.

Köszönöm tehát a barátságodat, a támogatásodat, és ezzel kívánok boldog hetvenediket (és még nagyon sokat utána)!

Balázs (és Andi és Dorina és Peti)



IFABO'95_ML_Enyedi
_Gödöllő_HEK_MBM_
Compfair_VTCD_AWU_AK_
Simonyi_Wimmer_
Ország(h)_
Lepke_
Infinity_Kovács_Göjögö_
Szapu_Break_
RAL9001_Micu_29

| DOI: 10.18135/PG70.2024.36

Sokat fejlődött a mesterséges intelligencia, de úgy gondolom, hogy a szövegdoboz tartalmának a megértéséhez a következő 70 év fejlesztése is kevés lesz – Neked viszont a közös munkával eltöltött 29 év is bőven elegendő.



Boldog születésnapot, Gábor!
Egészségedre!
Zsolt



KEDVES HERR CEO!

I DOI: 10.18135/PG70.2024.37

Olyan jó rád, rátok, a Morphóra gondolni! Nagyon sokat köszönhetek neked és a Morphónak, fel sem tudom sorolni, mennyi mindent. Például, hogy nálatok kerültem be az IT világába, ahol mostanáig sikerrel megragadtam. Vagy hogy milyen határozott szeretettel motiváltál, hogy hosszú idő után végül lediplomázzak. :) És akkor ott volt még a sok-sok közös élmény, kaland veled és a kollégákkal. Olyanok voltunk, mint egy család. A Morpho a lécet nagyon magasra állította: a veletek töltött évek óta mindig a Morphóban megélt családiassághoz, otthonossághoz mérem a munkahelyeimet. Igen, volt már olyan helyem, amelyik megközelítette a Morphót, de egyik sem tudta übereelni: MorphoLogic, Prószéky Gábor csak egy van. Gyakrabban eszembe jutsz, mint hinnéd, és szeretettel hordozlak szívemben!

A Jóisten tartson meg még sokáig, boldog születésnapot!

Tardy Gyuri



KEDVES GÁBOR!

DOI: 10.18135/PG70.2024.38

Sokat gondolkoztam, milyen érdekes történetet idézhetnék fel az együtt eltöltött több mint két évtizedből. De aztán inkább nem választottam egyet sem, hiszen ez a jelen feladat is éppen olyan jól jellemző lehet kettőnk kapcsolatára, mint az a rengeteg, amit közösen együtt megoldottunk. Annak a tetszőleges történetnek a részleteire, amit elmesélhetnék, te sokkal jobban emlékeznél, sokkal szívesebben tudnád előadni, és ami a legfontosabb, az a hallgató, aki egészen biztosan személyesen ismer, szívesebben hallgatná ezt tőled. Én motiváltam a fejlesztéseinket, a helyesírás-ellenőrzőt, a szótárprogramot vagy a fordítóprogramot, te vezetted a MorphoLogicot, és gondoskodtál arról, hogy az elért eredményeinkről minél szélesebb kör értesüljön. Ez egy nagyon szerencsés kapcsolat volt, amiben a beszámolás a közös dolgainkról a te feladatod volt, és amit te mindig nem csak lelkesen, de kiválóan végeztél.

Mesélj nekik még, Gábor!

Tihanyi Laci





KEDVES GÁBOR!

I DOI: 10.18135/PG70.2024.39

Tavaly, 2023. szeptember 21-én volt tíz éve, hogy összeházasodtunk, idén januárban pedig tizenöt éve, hogy összejöttünk. Köszönjük neked és a Morphónak, hogy találkozhattunk!

Csupa jót kívánunk neked, isten éltesen!

Sok ölelés:

Móni és Csaba

ELŐSZÓ

DOI: 10.18135/PG70.2024.40

Prószékly Gábor a Pázmány Információs Technológiai és Bionikai Karának a kezdetektől, már a tervezés idejétől elkötelezett támogatója – oktatóként, kutatóként egyaránt. Plasztikusan emlékezik vissza a Kar 25 éves évfordulójára készült filmben, hogy 1998-ban ismerkedett meg Roska Tamással, aki akkor szervezte-tervezte a Pázmány Péter Katolikus Egyetem új, az információs technológiát a biológiával és fizikával ötvöző Karának mérnök-informatikus képzését. Első találkozásuk végén, amikor hosszan beszélgettek a humán nyelvtechnológiáról, Tamás kifejezte, hogy most már látja, „nemcsak az emberi hardverrel, hanem az emberi szoftverrel is kell foglalkozni”, és meghívta őt, hogy csatlakozzon ennek kialakításához.

Gábor így részt vett már az első, 2001-es, akkor még szóbeli felvételi bizottság munkájában is. Kidolgozta és oktatta a Nyelvtechnológia alapjai, a Nyelvtechnológia eszközei és a Számítógépes fordítástámogatás című tantárgyait, 2006-tól már egyetemi tanárként. Sokan emlékszünk színes, a kutatás iránti érdeklődést felkeltő előadásaira Karunk Beköszöntő napjain.

2006-tól 2016-ig a Kar kutatási és innovációs dékánhelyetteseként segítette és támogatta a kari pályázatokat, a kutatásokat termék felé elindító kezdeményezéseket.

Alapító tagja a Kar – immár Roska Tamásról elnevezett – Doktori Iskolájának, a Humán nyelvtechnológiák, mesterséges értés, kommunikáció program vezetője. 2011 óta a kari Doktori és Habilitációs Tanács elnöke.

Doktoranduszaival és már PhD-t szerzett tanítványaival közös kutatásaikat az MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoportban folytatják – ez 2012 óta formálisan is megalkult az MTA Támogatott Kutatócsoportjaként. A kutatócsoport székhelye a „százpi”-ben, azaz a 314-es irodában volt. Számtalan BSc-s és MSc-s szakdolgozat került ki ebből a műhelyből, és az évek során tíznél több PhD is született itt Gábor keze alatt.

Kezdetben olyan területekkel foglalkoztak, mint a morfológiai elemzés és egyértelműsítés, orvosi szövegek feldolgozása, helyesírási hibák automatikus javítása és a statisztikai gépi fordítás. Nagyjából 2018-ig az AnaGamma performanciaalapú nyelvi elemzőmodell létrehozása határozta meg a kutatócsoport tevékenységét. Idővel, ahogy a nyelvtechnológiában is változtak a hangsúlyok, a kutatócsoport fókusza is a neurális megoldásokra tevődött át. A műhely fontosabb eredményei közé sorolható a PurePos nyílt forrású morfológiai elemző és egyértelműsítő, a klinikai szövegekből történő információkinyerést támogató rendszer, az első szóbeágyazási modell magyar nyelvre és még számos egyéb nyelvtechnológiai alkalmazás.

Számos nagy nemzetközi és hazai kutatás-fejlesztési projektben vettek és vesznek részt.

Kedves Gábor! 70. születésnapod mérföldkövénél sokan küldték el szeretetteljes köszöntésüket, gondolataikat, visszaemlékezésüket az ITK-n együtt töltött évekről – ez a mi közös ajándékunk! Isten éltesen sokáig!

Nyékyiné Gaizler Judit

SZÜLETÉSNAPODRA

DOI: 10.18135/PG70.2024.41

Prószéky Gábor, tudás fáklyája,
Hetven év sűrűjében fénylik a lángja:

Bölcsesség,
Örökség

Tanár Úr, ünneplünk Téged,
E jeles nap összehozott minket!

Jó barátok,
Tanítványok

Születésnapod alkalmából állunk most itt,
Hálánkat, tiszteletünket fejezzük ki:

Ünneplés,
Elismerés

A MorphoLogic alapjait Te raktad le,
Az ITK-nál a tudás fáklyáját vitted:

Innováció,
Inspiráció

Általad nyílt ki sok ifjú szárnya,
A tudás tengerén új hajókra várva:

Mintaként,
Mentorként

Szakmai útmutatásod az ITK falai között,
Generációkat inspirált, tudományra ösztönzött:

Hozzáértés,
Tisztesség

Bárhova léptél a szavak új életre keltek,
A magyar nyelv digitális térben végre megerősödhetett:
Szolgáltál, hol
Alkottál

Nem csupán a tudományban vagy jeles,
Emberi értékekben is példád vezet:
Tisztelet,
Szeretet

Intézményeid falai között új szellem született,
Hol a tudomány és az emberiség együtt léphetett:
Kapcsolatok,
Barátságok

Kívánjuk: egészség, boldogság kísérje minden lépésed,
A tudomány és az oktatás útján további sikerek érjenek!
Jövőkép,
Békesség

Ünnepeljük e jeles napot, mely összehozott minket,
A közös út további lépéseit várjuk, tele reménnyel.
Köszönet,
Hála

Cserey Gyuri powered by ChatGPT

KEDVES EMLÉKEM PRÓSZÉKY PROFESSZOR ÚRRÓL

DOI: 10.18135/PG70.2024.42

1998 júniusában a Magyar Katolikus Püspöki Konferencia egyházilag megalapította a Pázmány Péter Katolikus Egyetem Információs Technológiai Karát, melyet egyetemi fakultásnak az állami hatóságok 2001-ben ismertek el.

Miután 1978-ban elkészült *A fizika kultúrtörténete* c. könyv, aminek elkészültében rengeteget dolgoztam, a szaktárgyak (elektromágneses térelmélet, elektronfizika, kvantummechanika) oktatása mellett, majd azokat felváltva elkezdtem Simonyi Károly bátorítására a *Kultúrtörténetet* is tanítani a Műegyetemen (nyolc éven át a Képzőművészeti Egyetemen, és ugyancsak 8 évig a Veszprémi Egyetemen is oktatam ezt a tárgyat). Az ITK indulásakor Roska Tamás hívására csatlakoztam az ITK-hoz, kezdetben részigazgatóként, majd teljes állásban.

Az első óráim a második évezred kezdetén az egykori piarista házban voltak. Nyomasztó volt a romos bejáraton belépni. A lépcsőfeljáró kanyarulatában drogos hajléktalanok húzták meg magukat, míg az igen határozott fellépésű dékáni hivatalvezető, Rita, a Sacré Cœur-ös nővér ki nem zsuppolta őket.

Kicsit konszolidáltabb körülmények között zajlottak az órák az idővel már rendbe hozott piarista épület egyik termében. Az infrastruktúra hagyott ugyan maga után kívánivalót, de over head projektorunk már volt, fóliákról vetítettem az előadás illusztrációs képeit. Roska dékán úr kép- és hangfelvétel-készítéshez is beruházott. Így idővel szépen kivirágzott a későbbi Vízmű.

Immár 25 éve alapító dékánunk megkért, hogy rám bízassa egy olyan tárgy indítását, majd gondozását, amivel kapcsolatos elképzelését azután alakította ki, hogy a Harvard Magazin 1999. júniusi számában elolvasta Alan Greenspan egy harvardi egyetemi évfáradón tartott előadását.

Szerveztünk tehát zenei eseményeket is. Ezek az események már a később órarendbe iktatott, kötelezően választható *Multi-diszciplináris kitekintés* c. tárgy csírái voltak. Mikor dékán úr megkért, hogy hozzam létre ezt a tárgyat, én inkább Apáczai Csere János gyönyörű szavát felhasználva a *Rácsodálkozás sokféle tanítmányra* címet szerettem volna adni.

“Sustaining growth depends not just on students’ technical proficiency, but on educational institutions’ success in helping transform skills and intelligence into wisdom – into a process of thinking capable of forming truly new insights, a process fostered by exposure to philosophy, literature, music, art, and languages.” | **THE VALUE OF VALUES. THE PRINCIPAL COMMENCEMENT SPEAKER, ALAN GREENSPAN, CHAIRMAN OF THE BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM**

„Az absztrakt gondolkodás képességet kulturális és művészeti ismeretek nélkül nem lehet elsajátítani. Fontos a filozófia, az irodalom, a zene, a művészet és a nyelvek tanítása. Egyre fontosabb az élet azon élményeinek megértése, amelyek túlmutatnak az anyagi jóléten.”

I ALAN GREENSPAN, A COMMENCEMENT FŐELŐADÓJA, A FEDERAL RESERVE SYSTEM KORMÁNYZÓTANÁCSÁNAK ELNÖKE: AZ ÉRTÉKEK ÉRTÉKE

Dékán úr felesége, Esztó Zsuzsa a Zeneakadémia különleges tehetségek osztályából hozott diákokat. Két kislány csatos lakkcipőben, térdzokniban, miniszoknyában játszott gyönyörűen. Konyicska Renáta, Gyöngyösi Ivett ma Európa-szerte ismert művésznők. Felnőttként is játszottak már az ITK-n az ugyancsak nemzetközi hírű Pusker Júlia hegedűművésszel együtt.

Nekem és a 96 fős létszámú egyik kezdő évfolyamnak is nagy esemény volt, amikor Prószéky tanár úr a kamara-dzsesszzenekarával játszott, még a piarista épületben. Egykori gimnáziumi osztálytársaival létrehozott Qwyx [ejtsd: kuviksz] együttesével adott nagy sikerrel könnyűzenei koncertet. Kár, hogy már nem lépnek fel az ITK-n, pedig az aula remek helyszín lenne!

Csurgay Árpádné Ildikó

KEDVES GÁBOR!

DOI: 10.18135/PG70.2024.43

Könnyű Veled beszélgetni. Mindig is élveztem, akármilyen témában. :)

Akár egy apró morfológiai jelenségről (sok érdekes példával!), modellvasutakról vagy világpolitikai összefüggésekről, számomra emlékezetes beszélgetések voltak. Átsüt rajta a pozitív életszemlélet, a lehúzó erők ellenére is. Otthonosság, ez jut eszembe, amikor találkozunk. Egyből értettük egymást. Sok közös pont.

Emlékszem, a morphós állásinterjúm után (1999), egy random regnumi bálon nem akartam hinni a szememnek: egy színdarabban szerepeltél Artúrral és másokkal. „Szingtén zenész.” ;)



Már az első pillanattól érezhető volt, hogy a (számítógépes) nyelvészet a szíved csücske. Lelkesedéssel, élvezettel mesélt a területen már bevett nehezebb kanyarokat, és „milyen szerencse, hogy a magyar nyelv ilyen bonyolult! Mert kiderült, ha a magyarra már megcsináltuk a modellt, a többi nyelvhez már csak egyszerűsíteni kell rajta.” :D

A céget atyaián vezetted, nagy szabadságot adva mindenkinek. Különösen kedves volt számomra egy időszakban, hogy fizetésemelés helyett heti egy szabadnapot kértem. És megadtad! Csodálatos volt, hogy kicsi gyermekeink mellett szerdánként otthon lehettem. Ezúton is köszönöm!



Tetszik a lazaságod. Gmk, kft., doktori program vagy akár egy kutatócsoport alapítása: vágjunk bele, semmi gond! :D

A sok bonyolult részlet, adózási labirintusok, pályázati tervezési nehézségek, ezek mind nem számítanak, bátran belevágsz.



Egyszer egy konferenciára kellett mennünk, segíteni. Egymás mellett ültünk egy nagy előadásban, sok résztvevővel. Te egy kis manikűrrollóval vágta a körmödöt, a legtermészetesebb módon. Az előadó egyszer csak megkérdezte: van-e valakinek kérdése. Erre felálltál, kisollóstu! „Nekem van!” Egyik csodálkozásból a másikba estem! :D

Nagy élmények voltak a közös utazások. Izland. Peking. Nagy rutinnal, rengeteg utazási tapasztalattal nyomtuk.

Pár villanás. Egy vízeséshez kirándultunk fel, de elvesztettük az utat: egy ponton már szakadék melletti bokrokon kellett átmászni, Te pedig a frissen mütött lábaddal éppen tudtál már bot nélkül járni. Ha itt leesik, lesérül Gábor, nehezen tudom majd kimagyarázni Mártánál: „Figyelj, Gábor, forduljunk vissza!” :)

Oslói reptér, késett a gépünk, és félő volt, hogy lekéssük az átszállást. (Pláne, hogy a biztonsági ellenőrzésen is vitába keveredtél egy duty free bor miatt, így további értékes perceket veszítettünk.) Futottunk át a terminálon, mint az örültek. Akadályfutás volt, a többi utast kerülgetve, nagy sebességgel. Volt azonban, ahol nem lehetett kerülgetni. Két méteres négereket löktél földre futás közben, én meg a nevetés közben arra gondoltam, max lekéssük, de akkor olyan kárpótlást fogsz kidumálni, hogy a Lufthansa arról kódul. :)

Peking: fullon tömött busz, a piros hetes csúsidőben is matinéműsor ehhez képest. Hiába sodródtunk el egymástól messzire (minden megállónál felszállt még plusz 50 ember, és egyre jobban tömörödünk), mégis könnyen tudtunk kommunikálni, a fejek felett láttuk egymást: a helyiek egy fejjel alacsonyabbak voltak.

Egy nap bevittem magammal a 7 éves Kata lányomat a Morphóba. Nagyon jól érezte magát ő is. Mikor Nálad járt, megkínáltad pizzával. Ez mélyre ment: a mai napig „pizzás főnök”-nek hív. :D („Nézd, megint díjat kapott a pizzás főnök, de jó!”)

Kata most kikereste akkori e-mail-váltásokat. :) Szívet melengető, főként a kedvesség, ahogy válaszoltál neki.

Kedves Gábor, Isten éltesse szülinapod alkalmából is!

Endrédy Pisti, Orsi, Balu, Kata, Dorka, Bence





EGY JÁTÉKOS, SZABAD ELME

DOI: 10.18135/PG70.2024.44

Az iskolában „letudtam” a nyelvtant, a nyelvészet számomra egy száraz, szürke folt volt az iskolai tárgyak között. Egyetem közben a magyar nyelvtan lélekölő stílusú „tanítása” miatt hagytam félbe a francia nyelvű tolmácsolást. Ezután kisebb megvilágosodás volt Roska Tamás műhelyében Gábort és doktoranduszait megismerve megtapasztalni, hogy mennyivel izgalmasabb módon is lehet dolgozni a nyelvi struktúrák világában. Őket hallgatva a doktori iskola különböző fórumain egy olyan szellemi légkörbe láthattam bele, ahol Weöres Sándor, Kosztolányi és mások nyelvi leleményessége ötvöződött az információs technológiák „kocka” világával. Szoftvertervezők csináltak modern nyelvi cirkuszi mutatványokat, ahol vidám nyelvi szikrák alá valódi tudomány adta az alapot. Ritka szellemi műhelyt teremtettél, Gábor, ahol a felnövekvő elmékben tud a játékoság ötvöződni a felfedezés és hasznosítás versengő feszültségével. Úgy érzem, sokan vágyunk arra, hogy olyan alkotói légkört teremtsünk, amit tőled látunk, tanulunk. Adja Isten, hogy még sok éven át mutass ebben példát!

Dani

KEDVES GÁBOR!

DOI: 10.18135/PG70.2024.45

Közös munkánk 2007-ben kezdődött, amikor a Pázmány Péter Katolikus Egyetem Információs Technológiai Karán el kellett döntenem, mivel szeretnék foglalkozni. Ekkor kerestelek

meg, azzal a kéréssel, hogy legyél a témavezetőm. A tudományos útmutatásoddal sikerült olyan témát találnom, amivel egy fél évet a Leuveni Katolikus Egyetemen tölthettem. Posztgraduális tanulmányaim alatt is ugyanezt a támogatást és biztatást kaptam Tőled, és annak ellenére, hogy szinte az utolsó percig nem voltak pozitív eredményeim, 2015-ben sikeresen megvédtem a fokozatomat. Sokat küzdöttünk a gépi fordító „megregulázásán”,

hogy *a postás is meg tudja harapni a kutyát, és hogy az időlegyek is úgy repülhessenek, mint a nyíl*. Nemcsak tudományos mentorom voltál, hanem olyan tanító is, aki megtanított arra, hogy soha ne adjam fel, hiszen a zsákutcák is tapasztalatot adhatnak.

A közös konferenciáinkhoz kapcsolódó utazások – a pekingi teaház operája, a nagy fal tövében elköltött pizza, a reykjavíki fürdőzések, az oslói reptér biztonsági kapuja, az

otthoni konzervek feletti szállodai beszélgetések és nem utolsósorban a szege-di vonatozások – felejthetetlen élményekkel ajándékoztak meg, amelyek még jobban összekovácsolták a kutatócsoportunkat. Ezek az események rámutattak arra, hogy nemcsak a tudományos munka a feladatunk, hanem a világ sokszínűségének ismerete és a kultúrák közötti kapcsolatok építése is kiemelkedő fontosságú egy kutató számára.





Az egyetemi évek után a Nyelvtudományi Kutatóközpontban és a MorphoLogicnál végzett közös munkánk újabb mérföldkövekkel gazdagította pályámat. A te vezetésséddel és inspiráló jelenléteddel minden projekt, amelyben részt vehettem, nemcsak szakmai, hanem személyes fejlődést is jelentett számomra.

Ez az évforduló arra emlékeztet bennünket, milyen fontos a közös idő, a nevetés és az együtt töltött boldog pillanatok. Kívánom, hogy az elkövetkező évek még több örömet, egészséget és szeretetet hozzanak az életedbe. Ünnepeid ezt a napot azzal a tudattal, hogy életed gazdagítja a körülötted élők életét, és hogy sok szeretet övezi. Itt az idő, hogy visszatekints az eddigi útra, és büszkén állj minden elért eredményed mellett. Köszönöm, hogy részese lehettem és lehetek a pályádnak, és hálás vagyok mindenért, amit tettél közösségünkért.

Tisztelettel és hálával:

Laki Laci

mə'klɔːxlin NYOMÁBAN

I DOI: 10.18135/PG70.2024.46

2011-ben találkoztam először Prószékly Gáborral, a PPKÉ BTK Digitális bölcsészet MA „A nyelvtechnológiai alapjai” című előadásain. A kétéves képzés alatt már megfogalmazódott bennem, hogy jó lenne egy doktorival folytatni, de kicsit féltem, hogy a nyelvtechnológiához túl bölcsész vagyok, a nyelvészethez meg túl technológus. Gábor viszont megnyugtatóan, hogy igenis van feladat egy „magamfajta” számára (az ő szóhasználatával).

És lett is: elkezdtem a PhD-t, és közben formálódott az MTA-PPKE Magyar Nyelvtechnológiai Kutatócsoportban az AnaGramma elemző, amelynek kezdeti, Excel-táblás fázisában, amikor 2-3 órát ültünk egy-egy mondatrész elemzése fölött, én is részt vettem. Egy életre megjegyeztem, hogy: „Csapott célgömbbel úgy célzunk, hogy a nyílt irányzék nézőkéjét ráültetjük a cél alsó széle közepére, hogy az a vízszintes irányéllal elvágólag legyen.” És még sok más mindent is megtanultam; hamar kiderült, hogy Gáborral a nyelv és a nyelv számítógépes feldolgozása iránti rajongáson túl osztozunk sok minden másban is. Például a furcsa szóhasználatok, érdekes mondatstruktúrák kiszúrásában („bárkinek adható, egyedi ajándék”). Vagy a banális helyesírási hibák kibökésében („G. D. nagy színház és mozirajongó”, „mátrai borzaska csirkéből sült burgonyával”, „Szeptember 20-án az épület minden lakásában csótány és hangyairtás lesz.”). Gábor éles szemét jól példázza, hogy miután elküldtem neki majd 150 oldalas angol nyelvű disszertációm, hamarosan visszaírt: valahol a 78. oldal táján talált egy elütést.

Közös munkánkat egyfajta második konzulensként végigkísérte *Angela Merkel német kancellár* – egy hónappal disszertációm megvédése után az ő hivatali ideje is letelt, de nevét és tituluszát még sokszor emlegettük, írásban és szóban is.

Gábortól nem csupán a szűkebb értelemben vett szakterületünkről tanultam sokat. Kolozsvárról jöttünk haza, egy MANYE Kongresszusról, autóval, amikor szóba került John McLaughlin – a gitáros, akit én is és ő is ismertünk és szerettünk. Lelkesen

beszéltem, hogy [mek'lɒflin] így, és [mek'lɒflin] úgy. Gábor csendesen hallgatott, de érződött rajta, hogy nagyon szeretne mondani valamit. Megvárta, míg végzek, és akkor jegyezte meg – a tőle megszokott nagyon udvarias stílusban, mintegy mellékesen: „nagyon érdekes, de képzeld, én úgy hallottam, hogy bár azt hinnénk, hogy a *laugh* miatt [lɒf], de nem... Úgy kell kiejteni a nevét, hogy [mɛ'klo:xlɪn].”

Gábornak mindig van egy jó sztorija. Ha olyannal vagyunk egy társaságban, aki „újabb” ismeretség, rendre úgy kezdi, hogy: „Noémi, ezt te már biztos sokszor hallottad...” – majd jön a történet, amit általában tényleg már sokszor hallottam, de az is lehet, hogy még egyszer sem. Ezért mindig ugyanazzal a lelkesedéssel hallgatom, és remélem, hogy még sokáig lesz rá lehetőségem, hogy újabbakat hallgassak – vagy régebbieket hallgassak újra.

Gábor, Isten éltesen sokáig! A híres portással együtt, a letekert ablakon bekönyökölve érdeklődöm a továbbiakról: „Mi az elképzelés?”

Ligeti-Nagy Noémi

P-70-É-VERS

DOI: 10.18135/PG70.2024.47

A Prószéký az egy kód:
o1 lesz a hosszú ó-d
az O3 pedig: Ó
7 biten ez volt menő

A Prószéký egy Qwyk
de nem huhog, fafúvik
blokflótél és fuvolál
na meg hát szaxofonál

A Prószéký az egy gomb-
és nem-gomb-focibolond
más szóval: egy szurkoló
(nem pedig kátrányoló)

Prószéký az életem
felében ott volt velem
és ez komoly tényező
fél élet, az nagy idő

Prószékývel megesett
az, hogy három neve lett
mind a ketten így jártunk
egyét sose használtunk

De a közíg rátalált
átnézve matrikulát
és így kísértének mind
sosem használt neveink

A Prószéký a Márta
életre szóló társa
s ha van három gyereked
érti, hogy milyen neked

S persze ott van a Judit
(bár nem tudom a tutit)
kár, hogy ott van már, nem itt
(szeretnék ilyen nagyit)

A Prószéký a Murphy:
minden be tud dögölni
a hogyanját megleled
ha a Gábort figyeled

A Prószéký a Morpho
feje s arca, s az volt jó
hogy amíg megy a szekér
jó sok ember belefér

A Prószéký: vezető
nem csak autót vezet ő
céget, MANYE-t, kutcsopot
intézetet, központot

Prószéký az egyetem
ura nem volt bár sosem
de az azért számított
hogy ő miről mit mondott

Mert a Prószéky beszél:
anekdotázik, mesél,
interjút ad és elő
ebben nem fárad el ő

A Prószéky: benne van
tanácsban, bizottságban
OTKA-kollégiumban
s díjkuratóriumban

S B. I. révén benne van a
magyar irodalomban
nem én teszem hát bele
benne van egy ideje

De hogy itt a március
és jól elmúlt az idus
két napja is van neki:
egyik név-, másik szüli-

Úgyhogy itt állok sorba
(nem kizárt, hogy lesz torta)
gondoltam, jó lesz egy vers:
Prószéky het-ven-é-ves





KEDVES GÁBOR,

DOI: 10.18135/PG70.2024.48

szinte hihetetlen, és mégis igaz: 70. születésnapod alkalmából köszöntelek!

Emlékszem arra a lelkes fiatalemberre, akit majd' 50 éve megismertem, aki egyszerre volt nyelvész és programozó, és az élő nyelvek – elsősorban a magyar nyelv – gépi fordítását tűzte ki célul...

Ez a cél meghatározta egész életedet! Kutatásaid eredményeiért sok millió felhasználó hálás!

Megmaradt a lelkesedésed is, ezzel vonzottad magad köré a téma iránt érdeklődő fiatalokat, akik vezetéseddel számtalan sikeres PhD-fokozaton túl nemzetközi színvonalú kutatócsoportokban tapasztalhatják meg a közös kutatás-fejlesztés élményét.

Hálás szívvel gondolok közös munkánkra a Pázmány Péter Katolikus Egyetem Informatikai Technológiai Karának kialakításától kezdve. Ez a kezdetektől szívügyed volt. Az itteni oktatás, kutatás, a Doktori Iskola máig fontos prioritása az életednek – köszönet érte!

Köszönjük, hogy vagy, a jó Isten éltesen sokáig!

Szeretettel:

Nyékyné Gaizler Judit



GÁBOR ÉS A GRÍZES TÉSZTA, SÁRGABORSÓ-FŐZELÉK...

I DOI: 10.18135/PG70.2024.49

Örömmel teszek eleget Márta felhívásának, hogy írjunk egy közös emléket a most születésnapját ünneplő Gáborról. Sok emlék köt össze minket, hiszen a Pázmány Péter Katolikus Egyetem Információs Technológiai Karának (ITK) alapításától együtt dolgoztunk, mint később kiderült, mindketten a Piarista Gimnáziumba jártunk, sok közös tanárunk volt. A Gáborral való együttlét, beszélgetés mindig érdekes és szórakoztató. Mindig volt/van egy jó története, és ezt a nyelvési műveltségével adja elő. Itt egy köznapi történet jut eszembe.

A PPKE ITK-n kezdetben menza is működött, itt az oktatók és a hallgatók is egyszerű, megfizethető ebédhez jutottak. A tanárok kisebb csoportokban együtt mentek ebédelni.

Egy alkalommal Gáborral mentem ebédelni, és az „A” menüben sárgaborsóleves és grízes tészta volt. Gábor ezt választotta (én, biztosan mondható, nem ezt), és látható örömmel látott neki. Ebéd közben meg is magyarázta:

Tudod, a Márta nagyon jól főz, de vannak ételek, amelyeket nem hajlandó elkészíteni. A sárgaborsó, a grízes tészta ilyen tiltólistás, de én ezeket szeretem, és bizonyos nosztalgiát érzek irányukba.

Valószínűleg Mártának is nosztalgiája lehet, csak ellenkező előjellel.

Kedves Gábor, Isten éltesse sokáig!

Szolgay Péter

	A	B	C	D	E	F	G	H	I	J	K
1		határ		konstrukciót	ajánl	devizahiteleseknek					Bankszövetség.
2	#										
3				Adjék+Num							
4	5	CAS	határ								
5	6	CAS	0-kereső								
6	7	CAS	Det-kereső								
7	8	N	jezőkereső								
8	9	Det	él								
9	10	Adj	él								
10	11	FIN	határ								
11	12	FIN	fokuszkereső								
12	13	FIN	vonzatkereső								
13	14	ajánl, Sp3	vonzatkereső								
14	15	ajánl	vonzatkereső								
15	16	ajánl	vonzatkereső								
16	17	ajánl	szabachitározókereső								
17	18	vonzat(Acc)	él								
18	19	V	0-kereső								
19	20	Det	határ								
20	21	Det	0-kereső								
21	22	Det	határ								
22	23	Det	0-kereső								
23	24	CAS	határ								
24	25	CAS	0-kereső								
25	26	CAS	Det-kereső								
26	27	N	jezőkereső								
27	28	Det	él								
28	29	Det	él								
29	30	vonzat(Dat)	határ								
30	31	Det	0-kereső								
31	32	Det	határ								
32	33	Det	0-kereső								
33	34	N	jezőkereső								
34	35	PUNCT	határ								
35	36	PUNCT	0-kereső								
36	37	0	él								
37	38	0	él								
38	39	PUNCT, 0-él	0-Nom előfelműző								
39	40	CAS	határ								
40	41	CAS	0-kereső								
41	42	CAS, PropN	Det-kereső								
42	43	Det	él								
43	44	vonzat(Nom)	él								
44	45										

KÉT KONSTRUKCIÓT AJÁNL A DEVIZAHITELESEKNEK A BANKSZÖVETSÉG

I DOI: [10.18135/PG70.2024.50](https://doi.org/10.18135/PG70.2024.50)

Ez a táblázat csak kevés ember számára mond bármit is, nekem azonban a számítógépes nyelvészeti pályám kiindulópontját jelenti. Amikor magyar alapszakos koromban megtudtam, hogy létezik olyan dolog, hogy számítógépes nyelvészet, rögtön tudtam, hogy nekem ez kell. Hozzád irányítottak, én pedig csakhamar ott találtam magam mellett, ahogy órákon keresztül tanakodunk egy-egy érdekesebb vonzatkeretről, a keresletekről és kínálatokról, az esetegyértelműsítésről, az ablak méretéről, a több szálon futó feldolgozás kérdéseiről, az éppen a napi sajtóban talált érdekes és vicces mondatokról. Az anagrammás fészekaljban nevelkedő fiatalokkal együtt ugyan már kirepültünk, de a kisebb-nagyobb körben zajló megbeszélések útravalóját mindannyian magunkkal vittük.

Hálás vagyok ezért az útravalóért!

Vadász Noémi

AZ ELSŐ ACL: az első kínai utunk Gáborral

DOI: 10.18135/PG70.2024.51

2015-ben Pekingben került megrendezésre a nyelvtechnológia legnagyobb konferenciája, az ACL (Association for Computational Linguistics). A konferencia nemcsak szakmailag járult hozzá a fejlődésünkhöz, hanem a kutatócsoportot is összekovácsolta, máig tartó kapcsolatok születtek. Pekingben felejthetetlen élményekben volt részünk, Gábor családjával is szorosabb lett a kapcsolatunk. Nehéz röviden összefoglalni azt a sok élményt.

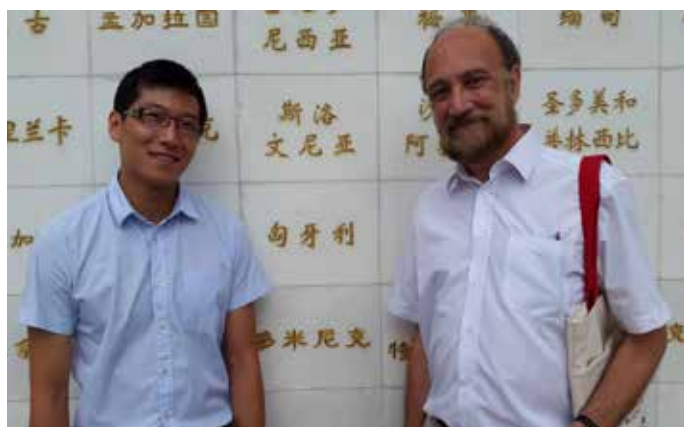
Ez a konferencia még olyan korban volt, ahol még nem a neurális megoldások uralták a nyelvtechnológiát. A gépi fordítás területén még szabályalapú és statisztikai megoldások jelentették az innovációt (lásd 1. ábra).

A szakmai konferencia mellett jutott idő kapcsolatépítésre is. Meglátogattuk a Pekingi Nyelvi és Kulturális Egyetemet, ahol megtaláltuk Magyarországot az országok falán (lásd 2. ábra) és vacsoráztunk a Kínai Nyelvészeti Diákolimpia igazgatójával.

A szakmai programok és munka mellett jutott időnk felfedezni Pekinget is, ahol a méretek más dimenziójúak. Amikor második alkalommal követtük el azt a hibát, hogy



1. ábra. Hibrid gépi fordítórendszer



2. ábra. Magyarország

1 órát hagyunk egy park megrendezésre, de 3 óra múlva még csak a park felénel jártunk, és még nem néztük meg a parkban található kis hegyen lévő pagodát, akkor változtatni kellett a tervezési folyamatokon. De említhetnénk egy olyan esetet is, amikor Gábor javaslatára lesétáltunk egy metrómegállót, mivel a térképen elég rövid szakasznak nézett ki, de futólépésben is 45 percbe tellett megtenni a távot.

Persze a hideg sör a végén feledtette velünk a fáradságot. A forró, párás napokon pedig gyakori hűsítő volt a sör, de a bakancslistás koccintós csúcspont mindenképpen a kínai nagy falon való koccintás volt (lásd 3. ábra).



3. ábra. Koccintás a nagy falon

Gábor mindig azt mondja, hogy ha a fejében elképzeli a nagy falat, akkor egy festményt lát, amin a fal a hegyek gerincén a végtelenbe kígyózik. Amikor megmásztuk a kínai nagy falat, ami önmagában is nagy falat volt, pont ez a felejthetetlen kép tárult elénk. Leghitelesebben Gábor saját kezűleg készített képe adja vissza azt a képet, amit elképzelt (lásd 4. ábra).



4. ábra. A kínai nagy fal



5. ábra. Skorpiók pálcán



6. ábra. Az eredeti (állítólag) pekingi tészta

Kína híres a különleges gasztronómiájáról. Sokan azt mondják, hogy ott a bogarakat is megeszik, hát nem tévednek, egy másik bakancslistás tételt pipáltunk ki közösen: megkóstoltuk életünk első skorpióját (lásd 5. ábra).

Amit azóta is emlegetünk, az az eredetinek mondott pekingi tésztás hely, ahol nem fért el az asztalon az a sok köret, amit egy szimpla tésztához adtak, nehéz leírni, ezért képen demonstráljuk (lásd 6. ábra).

A híres eredeti pekingi kacsáról (lásd 7. ábra) se feledkezzünk meg, ahol Gábor a minden idők ... (a pontos számot Gábortól lehet elkérni) kacsáját kapta, amit az 1864-es nyitás óta számolnak.

Még sorolhatnánk a bábszínház-as teaházat; a kínai misét; a lótuszos tavon lévő csónakázást és vacsorát; az első olyan élményt, amikor kesztyűs kézzel toltak be a zsúfolt metróba, vagy a túlszúfolt buszban még hangszóróval kiabáltak a fülünkbe, hogy menjünk beljebb, mert felszálló van; a tiltott város látványát; a sok random emberrel való szelfit; a kávéra való vadászatot (nagy nehezen találtunk egy mekit, amiben volt egy kávégép, de hárman sem tudták működésre bírni); vagy a kínai rokonainkkal való közös vacsorákat.



7. ábra. A pekingi kacsá

Köszönjük, Gábor, a közös élményeket, az első ACL-t Pekingben!



PG70

PG70