

from 77 data collection sites in total. In order for the project's sound archive (not just the data of the interviews processed so far) to be really available for use, the study emphasizes the need to create an open access, easy-to-use and searchable database based on various criteria.

Keywords: sociolinguistics, language attitudes, language and dialect awareness, New General Atlas of Hungarian Dialects, database.

PARAPATICS ANDREA
Pannon Egyetem Humántudományi Kar

TUDOMÁNYTÖRTÉNET

Prószéky Gábor és a Számítógépes Nyelvészet 35 éve*

1. Miért aktuális a Számítógépes Nyelvészet, 35 évvel megjelenése után? A köszöntéshez azért, mert Gábor (eddigi) pályájának felezőpontjára esik, és ezzel egyben a szakma (eddigi) pályájának felezőpontjára is! A méltatáshoz pedig azért, mert ez egy jó könyv, érényei (melyek egyben Gábor érényei) 35 év távolságból még világosabban látszanak.

Maga a gondolat, hogy a fordítói munka mechanikus, tehát elvben számítógéppel is elvégezhető, egyszerre merült fel WARREN WEAVER méltán híressé vált levelében, melyet a kibernetika egyik atyjához, NORBERT WIENERhez intézett (1947, megjelent WEAVER 1955-ben) és ANDREW DONALD BOOTH – egyebek közt a mágnesdob-memória egyik feltalálója – egy 1947-es memorandumában (BOOTH 2003). Az első működő rendszerek hét évvel később, 1954-ben láttak napvilágot: Amerikában a híres Georgetown-IBM kísérlet orosz → angol irányban, Angliában pedig az APEXC gépen a Bierkbeck College-ban, angol → francia irányban.

2. Előttörténet. Akkoriban hárman voltunk ifjú törökök a Nyelvtudományi Intézet 13-as szobájában: Kálmán László, Prószéky Gábor és én. Maga a szoba Kiefer Ferenc és Szépe György főnökleite alatt az „Általános, alkalmazott, és elméleti nyelvészeti (fő)osztály” névre hallgatott, ahol mi, ifjak a legtöbbet a nálunk 4-5 évvel idősebbektől tanultuk: Komlósi András, Bánréti Zoltán, Szabolcsi Anna, Kenesei István, É. Kiss Katalin, Kontra Miklós, Bakró-Nagy Marianne „mentoráltak” minket. A többi fiatal, Törkenczy Miklós, Bartos Huba, Siptár Péter később kapcsolódtak be, és a következő generációra még 4-5 évet várni kellett.

Gábor a könyvet először 1985-ben zárta, én már Stanfordinból kommentáltam, végül 1989-ben jött ki. Kenesei 1987 október 13-i e-mailjében ezt írta: „Bocs, ha elfelejtettem az intézeti magyar ábécét, de sohase használtam”. Mint ebből is kiderül, a „Prószéky-kód” már akkoriban is nélkülözhetetlen volt. (Ma már, azt hiszem, az egész világon egyedül én

* Előadásként elhangzott Prószéky Gábor 70 éves köszöntésén a Magyar Nyelvtudományi Társaság ülésén, Budapesten, az ELTE BTK-n, 2024. május 14-én.

használok.) Kálmán László 1987. május 6-án ezt írja: „Légyszíves írd a Prószékynek egy levelet, mert komolyan aggódom miatta. Egyszerűen semmire se lehet rávenni. Valahogy lelkesíteni kellene, vagy mit tudom én, rávenni, hogy kevesebb dolgot csináljon (...) van egy állása, amit szokás szerint imád, és vagy négy félállása.” Ebben Gábor, azt hiszem, nem sokat változott.

3. A könyv helye a szakirodalomban. 1989-es megjelenése félúton van az első és az utolsó klasszikus közt (HAYS 1967; JURAFSKY és MARTIN 2009). Ez volt nemcsak nálunk, hanem az egész világon lényegében az első LaTeX-ben írt nyelvészeti könyv. A világban se volt még sok, egy évvel előzte csak meg BARLOW és FERGUSON (1988), de persze Stanfordban könnyű volt, hiszen a TeX megteremtője, DONALD KNUTH ott volt professzor!

Fontos szerepet játszott VOTISKY ZSUZSA, a Latex-et hazánkban elfogadottá tevő TypoTex kiadó megalapítója, és HAVASS MIKLÓS (SZÁMALK). HAVASS-sal Gábornak már volt indirekt kapcsolata, mert ő folytatott egy olyan népdalkincs-kompjuterizációs programot, amit eredetileg HAVASS csinált, aztán KÖSZEGI GYÖRGY, majd SIPOS MIHÁLY, utánuk jött Gábor (HAVASS 2013).

4. Az alkotási módszer. Gábor hihetetlen alapossággal dolgozta fel az irodalmat (30+ oldal irodalomjegyzék, akkoriban nagyon nehezen megszerezhető szakirodalomból), a saját munkájának (ami a GÁZOLAJ-tól (1979) kezdve észrevehető volt, én legalábbis akkor vettem észre) csak pár százalék teret ad az irodalomjegyzékben is és a könyvben is (16. fejezet).

5. A könyv struktúrája

- ÁTTEKINTÉS komoly nyelvészeti bevezetővel, elméleti nyelvészet, formális nyelvészet, „a számítógépes feldolgozás szempontjából legfontosabb természetesnyelvi jelenségek” – 60 oldal
- MÓDSZEREK ÉS FORMALIZMUSOK (függőségi, kategoriális, CF, transzformációk, ...) – 180 oldal
- MŰKÖDŐ RENDSZEREK – 180 oldal
- SZABAD SZÓRENDŰ NYELVEK FELDOLGOZÁSA – 60 oldal
- A MAGYAR NYELV FELDOLGOZÁSA – 50 oldal
- IRODALOM – 30 oldal

Alapos, részletes, és nagyon jó a szelleme! Nem ragad le a szintaxisnál vagy a morfológiánál, végig a szemantikailag (sőt pragmatikailag) használható rendszerek érdeklik. Megelőz egy olyan kutatási irányt (KÁLMÁN–PRÓSZÉKY 1984, KÁLMÁN–PRÓSZÉKY 1985), ami újra aktív, bár ma egész más (nem szabályalapú, hanem ideghálózatos) felfogásban.

6. Ami kimaradt. A beszéd- és írásfeldolgozás, a statisztika sajnos teljesen figyelmen kívül lettek hagyva. Utólag persze könnyű okosnak lenni, de ezek bevételét én már akkor javasoltam Gábornak, amire ő akkor azt mondta, hogy ahhoz még egy ugyanilyen terjedelmű könyvre lenne szükség! Valóban, írtak ilyet előtte is (HERDAN 1964) és utána is (MANNING–SCHUTZE 1999), csak sajnos magyarul nem. Talán említést érdemelt volna

FAHLMAN (1979) mint a QUILLIAN (1968) által megindított hálózati irány akkoriban leginkább kiérlelt változata, továbbá MCCLELLAND és ELMAN (1986), mely az ideghálózatok direkt nyelvészeti alkalmazásával már akkor erős vitát váltott ki (PINKER–PRINCE 1988).

7. A krédó. Ekkoriban írta HAWKING (1988): „Ha csak egy lehetséges egyesített elmélet létezik, az sem más, mint egy csomó egyenlet és szabály. Mi lehel életet az egyenletekbe, mi teremt számukra világegyetemet, amit leírhatnak? A tudomány szokásos módszere, a matematikai modell felállítása nem válaszolhat arra a kérdésre, hogy miért áll a modell rendelkezésére a modellezésre váró világegyetem. Miért vállalja a világegyetem a létezéssel járó összes kényelmetlenséget?” (Molnár István ford.)¹

Gábor úgy gondolta (és ebben osztozott vele a másik két ifjú török is), hogy egy nyelvtani szabályrendszerbe csak a számítógépes implementáció képes életet lehelni. Manapság divatos az ELI5 (explain like I'm 5, 'úgy magyarázd meg, mintha öt éves lennék'), de az implementáció még ennél is butább befogadót tételez, olyat, akinek aztán tényleg mindent maximális szabatossággal el kell magyarázni. Itt nincs helye homályos fogalmazásnak, informális analógiáknak, általános (és a nagy általánosságban teljesen megfoghatatlan) elvekre való hivatkozásoknak, a számítógépnek nem lehet lukat beszélni a hasába.

8. A hatás. Tartós és jelentős: amiképp Kálmán László az ifjabb nyelvészgenerációk sorát nevelte ki, akképp Gábor megindította a számítógépes nyelvészek hazai képzését, megcsinálta a tanszéket a Pázmányon; a könyvből és tőle generációk tanultak. Isten éltesse sokáig!

Hivatkozott irodalom

- BARLOW, MICHAL – CHARLES A. FERGUSON eds. 1988. *Agreement in Natural Language: Approaches, Theories, Descriptions*. Stanford: Center for the Study of Language and Information.
- BOOTH, ANDREW DONALD 2003. Mechanical Translation. *Readings in Machine Translation*. Eds. SERGEI NIRENBURG, HAROLD SOMERS YORICK WILKS. MIT Press, Cambridge, Mass. 19–20.
- FAHLMAN, SCOTT 1979. *NETL, a system for representing and using real-world knowledge*. MIT Press, Cambridge, Mass.
- HAVASS MIKLÓS 2013. Hullámlavaglás. In: *A számítógép-tudományról egyes szám első személyben*. Szerk. KORNAI ANDRÁS. TypoTex, Budapest. 137–182.
- HAWKING, STEPHEN W. 1988. *A Brief History of Time from the Big Bang to Black Holes*. Bantam Books, New York.
- HAYS, DAVID G. 1967. *Introduction to Computational Linguistics*. American Elsevier Publishing, New York.
- HERDAN, GUSTAV 1964. *Quantitative linguistics*. Butterworths, London.
- JURAFSKY, DANIEL – JAMES H. MARTIN 2009. *Speech and Language Processing*. 2nd edition. Pearson.

¹ Even if there is only one possible unified theory, it is just a set of rules and equations. What is it that breathes fire into the equations and makes a universe for them to describe? The usual approach of science of constructing a mathematical model cannot answer the questions of why there should be a universe for the model to describe. Why does the universe go to all the bother of existing?

- KÁLMÁN LÁSZLÓ – PRÓSZÉKY GÁBOR 1985. FMR Grammar. *MTA NYTI Working Papers (Műhelymunkák a nyelvészet és társtudományai köréből)* 1: 31–41.
- MANNING, CHRISTOPHER – HINRICH SCHUTZE 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, Mass.
- MCCLELLAND, JAMES L. – JEFFREY L. ELMAN 1986. The TRACE model of speech perception. *Cognitive Psychology* 18/1: 1–86.
- PINKER, STEVEN – ALAN S. PRINCE 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28: 73–194.
- PRÓSZÉKY GÁBOR 1989. *Számítógépes nyelvészet*. Számítástechnika-alkalmazási Vállalat, Budapest.
- QUILLIAN, M. ROSS 1968. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science* 12: 410–430.
- WEAVER, WARREN 1955. Translation. *Machine translation of languages*. Eds. W. N. LOCKE – A. D. BOOTH. John Wiley & Sons, New York. 15–23.

Gábor Prószéky and 35 years of his textbook *Computational Linguistics*

On the occasion of Gábor Prószéky's 70th birthday we discuss the significance of the textbook *Számítógépes Nyelvészet* (Computational Linguistics) that he wrote 35 years ago.

Keywords: computational linguistics, Gábor Prószéky, 70.

KORNAI ANDRÁS
Budapesti Műszaki Egyetem

SZEMLE

Nyelvelmélet és diakrónia 5. Egy sorozat utolsó darabja

PPKE BTK Elméleti Nyelvészet Tanszék – Magyar Nyelvészeti Tanszék, Budapest, 2023. 424 lap.

2022-ben is megrendezték a Nyelvelmélet és diakrónia konferenciát. Ez három-éves váltakozásban volt a Nyelvelmélet és dialektológia, valamint a Nyelvelmélet és kontaktológia konferenciákkal. Ezzel bezárult az ötödik teljes ciklus, amely 15 évvel ezelőtt vette kezdetét.

A konferencia anyagából készült kötet 16 tanulmányt tartalmaz, melyet a jeles évfordulóra való tekintettel a két alapító, É. KISS KATALIN és HEGEDŰS ATTILA cikke nyit meg. Ezt követően először a magyar, majd pedig más nyelvekre vonatkozó kutatásokról olvashatunk. A tanulmányok ugyanakkor számos ponton kapcsolódnak egymáshoz mind a témák, mind pedig a kutatási módszerek tekintetében. Az ismertetés során a cikkeket nem a kötetben szereplő sorrendjük szerint, hanem tematikus csoportokban ismertetem.

Az első cikk (9–30) egy napjainkban igazán divatos kutatási területet érint: az evidencialitást. É. KISS KATALIN írása egy tipológiailag ritka jelenségre hívja fel a figyelmet