

Stroblon vs Stroblön, avagy a korpusz és a fonológus

Az elmúlt évtizedek nyelvészeti kutatásainak fontos eszköze lett a nyelvi adatok korpuszokból való kinyerése. Ugyanakkor ez nem feltétlen használható fel a nyelvészet minden területén. például a fonológiában, kivált az újabb kölcsönszavak fonológiájában, amit egy olyan kutatás példáján kívánok példázni, amely a szillabikus *l*-t tartalmazó idegen szavak magyarba történő adaptálását vizsgálta (Blaskovics & Ittész 2022).

Az említett kutatás egy laboratóriumi kísérleten alapult, amelynek során 28 beszélőtől gyűjtöttünk adatokat arról, hogy milyen harmonicitású toldalékalternánst választanak szillabikus *l*-t tartalmazó töveknél. (A vizsgált tövek közül hátsók: *Doodle*, *Google*, *Strobl*, *puzzle*; semlegesek: *Ybl*, *Lidl*; elülsők: –. E felosztásért ld. Rebrus & Törkenczy 2019-et. Mindkét vizsgált toldalék háromalakú: *-On*, *-hOz*.) Ennek főbb eredményei a következők voltak:

- (1) az „eldöntendő kérdés” teszt (Siptár & Törkenczy 2000) alapján igazoltuk, hogy a szillabikus *l*-t tartalmazó kölcsönszavak tövét két szótagban ejtik a beszélők;
- (2) mind szillabikus *l*, mind betoldott redukált magánhangzó előfordult a tövek második szótagjának szótagmagjaként;
- (3) a semleges tövek szinte kizárólag az elülső alternánst (*Yblhöz*, *Lidlön*) választották, a hátsó töveknél ellenben megfigyelhető variáció (*Google-hoz* – *Google-höz*). Az elülső toldalékalternáns megjelenése mégsem függ a betoldott magánhangzó megjelenésétől, így kérdéses, mi váltja ki ezeket. Egy új harmonikus osztályt feltételezünk, ahol az /l/ képes kiváltani az elülső alternánst, ám átlátszó lehet hátsó tövek után.

Miért alkalmatlan hasonló eredmények megállapítására a korpusz? (1)–(3)-at beszélt nyelvi korpuszban mind vizsgálni is lehetne, viszont a vizsgált szavak közül a gyakoribbak előfordulása is marginális a BEA-ban. Az írott nyelvi korpuszok (MNSZ2, Szószablya) ellenben triviálisan nem képesek intonáció alapján megerősíteni a szótagszerkezetet (1), sem egy betoldott magánhangzó jelenlétét és az ő rá való kondicionáltságát vizsgálni (2–3), és ebből fonológiai következtetéseket levonni. Például egyik korpusz sem tartalmazza a *Strobl*nek nem-hátsó toldalékkal ellátott alakjait, noha kísérletünkben az alakok több mint fele ilyen volt; vagy az *Ybl*nek *-hOz*-zal és *-On*-nal ellátott alakjait.

Emiatt kénytelen a kutató olyan környezeteket előállítani, amelyben ki tud kényszeríteni a korpuszban ritkán vagy egyáltalán nem előforduló alakokat. Ehhez van szükség olyan célmondatokra, mint:

Strobl: Vajon a kopaszsága miatt volt mindig sapka ____?

Hivatkozások:

- Blaskovics, Á. & Ittész, A. (2022, megjelenés alatt). *Ybl* a Google-ön. A szillabikus *l* és a magyar magánhangzó-harmónia esete. In *Beszédtudomány 2022*.
- Gósy, M., Gyarmathy, D., Horváth, V., Grácsi, T., Beke, A., Neuberger, T., & Nikléczy, P. (2012). BEA. In *Beszéd, adatbázis, kutatások*, 9–24.
- Németh, L., Rung, A., Szakadát, I., Halácsy, P., Kornai, A., & Trón, V. (2003). A Szószablya projekt www.szoszablya.hu. In *Magyar Számítógépes Nyelvészeti Konferencia 1.*, 298.

- Oravecz, Cs., Váradi, T., & Sass, B. (2014). The Hungarian Gigaword Corpus. In *Proceedings of the ninth international conference on language resources and evaluation (LREC-2014)*, 1719–1723.
- Rebrus, P. & Törkenczy, M. (2019). Magyar harmónia: a dolgok állása. *Általános Nyelvészeti Tanulmányok*, XXXI, 233–333.
- Siptár, P. & Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford/New York: Oxford University Press.