

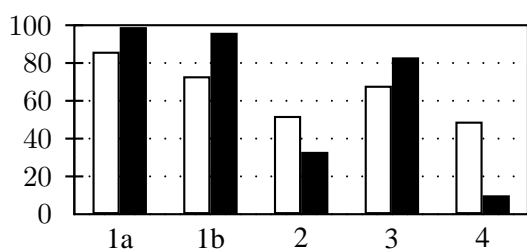
Van-e rendszer az izé-bizében? Labiális ikerítés a magyar nyelvben

Sóskuthy Márton, University of Edinburgh, m.soskuthy@sms.ed.ac.uk

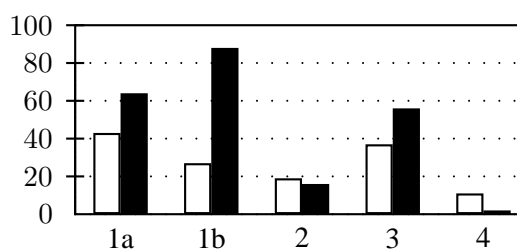
A labiális ikerítés kicsinyítő/becéző funkciójú morfológiai minta a magyarban, amely az alapszó megduplázásával áll elő; a jelenség érdekességét az ikerpár második tagjának kezdő mássalhangzója adja, amely mindig eltér az első tagétól, és jellemzően labiális (innen az elnevezés). Íme néhány példa: *cica-mica*, *csiga-biga*, *Julcsi-bulcsi*. Bár a minta meglehetősen produktív, a szakirodalomban igen kevés említés esik róla, és a létező leírások sem mondhatóak korszerűnek elméleti szempontból (ld. például Szikszainé Nagy 1993). Ennek részben az az oka, hogy a labiális ikerítés erősen variábilis és emiatt hagyományos nyelvészeti eszközökkel nehezen megragadható. Előadásomban éppen ezért modern számítógépes módszerekkel közelítem meg a jelenséget: a Magyar Webkorpuszból (Halácsy et al., 2004) automatikusan kigyűjtött adathalmazt gépi tanulóalgoritmusok és statisztikai módszerek segítségével elemzem. Az előadás végén a minta lehetséges eredetéről és a különböző nyelvelméletek számára való jelentőségéről is szót ejtek.

A vizsgált minta elkülönítése önmagában sem könnyű feladat: egyrészt a magyarban sok olyan hasonló funkciójú ikerpár létezik, amelyeknek a második tagja nem labiális kezdetű (pl. *angyal-kangyal*); másrészt számos olyan alakot találunk, amely látszólag beleillik a mintába, ám valójában nem ikerítés eredménye, hanem összetett szó (pl. *csillog-villog*). Mivel nem egyértelmű, hogy ezek az alakok *a priori* alapon kizárhatók a mintából, az adathalmazt egy igen általános keresőalgoritmus segítségével állítottam elő, amely az összes $O_1\{\dots\}_i-O_2\{\dots\}_i$ (O = szótagkezdet és $O_1 \neq O_2$) szerkezetű alakot megtalálta. Az így kapott szavak a szóalkotás módja szerint a következő csoportokba oszthatóak: 1) valódi ikerpárok (pl. *cica-mica*, *ici-pici*); 2) rímelő összetételek (pl. *csillog-villog*); 3) ikonikus párok, ahol egyik tag sem bír önálló jelentéssel (pl. *locs-pocs*); 4) kölcsönzavak (pl. *super-duper*). A 2-es, 3-as és 4-es csoport tagjai azért maradtak benne az adathalmazban, mert bár ezek az alakok más úton jönnek létre, feltételezhetjük, hogy az ikerítésnél érvényesülő tendenciák rájuk is befolyással vannak (pl. egy rímelő összetétel második tagja gyakrabban lesz labiális kezdetű).

Bár sok olyan ikerpár létezik, ahol a reduplikáns nem labiális kezdetű, a magyarban csak a labiális minta működik igazán produktívan, ezért az elemzésnél ezeket az alakokat vettem kiindulópontnak. Mivel az 1-es és a 3-as csoportban szereplő ikerpárok legalább egyik tagja önállóan nem létező alak, amely a szóalkotás eredményeként áll elő, a labiális minta itt produktívan működhet. Azt várjuk tehát, hogy ezekben a csoportokban magasabb lesz a labiális kezdetű alakok száma, mint a 2-es és a 4-es csoporton belül (az előzőnél ugyanis a szemantikai szempontok is fontos szerepet játszanak a második tag kiválasztásakor, az utóbbinál pedig nem kell érvényesülnie a magyar mintának az idegen nyelvi eredet miatt). Az 1. ábra igazolja ezt a várakozást. Érdekes ugyanakkor, hogy a 2-es csoporton belül még mindig lényegesen több a labiális kezdetű alak, mint azt a szókezdő labiálisok általános gyakorisága alapján várnánk – ez azt sugallja, hogy a labiális minta erre a csoportra is befolyással van. A labiális minta ezen felül bizonyos mértékű fonológiai kondicionáltságot is mutat: a második tag labiális kezdőhangjának képzési módja és zöngéssége bizonyos keretek között megjósolható az első tag alapján. Az adatok statisztikai elemzése (többdimenziós skálázással) alátámasztja ezt a megfigyelést: azt látjuk ugyanis, hogy 1) az első és a második tag kezdő mássalhangzója mindig eltér (ha az első [m] akkor a második [p] vagy [b], és fordítva) és 2) az első tag második szótagkezdete alapján részben jósolható a második tag szókezdő mássalhangzója (zöngés \rightarrow [b], zöngétlen \rightarrow [m]). Fonológiai kondicionáltságra utal az is, hogy azok a gépi tanuló algoritmusok, amelyek képesek kihasználni ezt a fajta információt (pl. TiMBL, Daelemans et al. 2007), lényegesen jobban teljesítenek a második tag szókezdő mássalhangzójának a megjóslásakor azokban a csoportokban, ahol a labiális minta erősebb (2. ábra).



1. ábra: a labiális reduplikánsok gyakorisága



2. ábra: TiMBL pontossági értékek

Hivatkozások

- Daelemans, W., J. Zavrel, K. van der Sloot & A. van den Bosch (2007). TiMBL: Tilburg Memory Based Learner, version 6.1, Reference Guide. ILK Research Group Technical Report Series no. 07-07.
- Halácsy, P., A. Kornai, L. Németh, A. Rung, I. Szakadát & V. Trón (2004). Creating open language resources for Hungarian. In: Proceedings of the 4th international conference on Language Resources and Evaluation (LREC2004).
- Sziksainé Nagy, I. (1993). *Az ikerítés helye, szerepe, szabályszerűségei a magyar nyelvben*. Magyar Nyelvtudományi Társaság Kiadványai, vol. 197.