

A melléknevek szerepe a nyelvi hatékonyságban

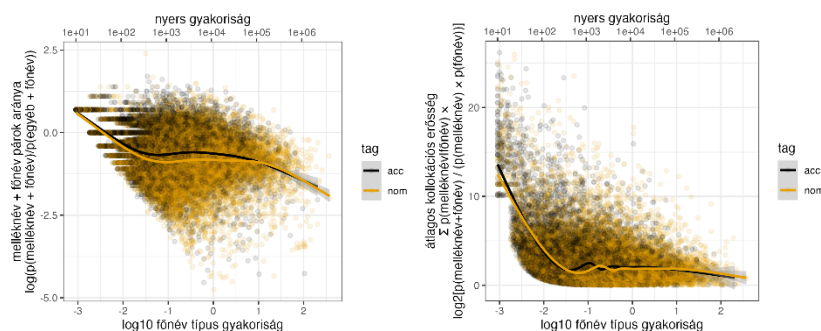
Az emberi kommunikáció hatékony: arra törekszünk, hogy minél több információt minél érthetőbben adjunk át. A nyelv a nehezen megjósolható információkat többszörösen kódolja, hogy megkönnyítse azok feldolgozását (Gibson 2019). A mondatalkotás szintjén ez azt jelenti, hogy a mondat bizonyos pontjain hányféle szó állhat, egy adott szót vagy szócsoporthoz milyen szó követhet. Vegyük ezt a három mondatot.

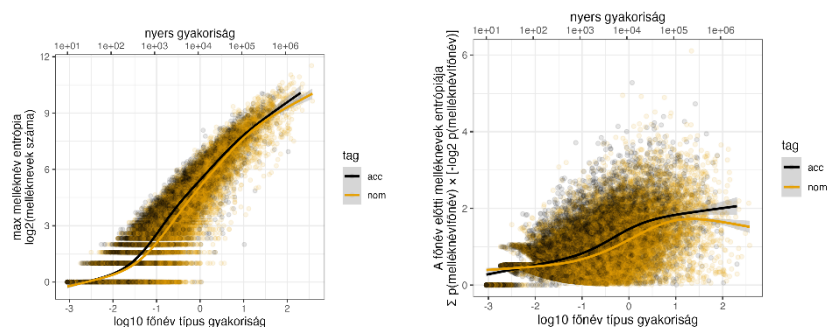
- (1) János kinyitott egy ... (ajtót / zárat / csomagot)
- (2) János megjavított egy ... (autót / csapot / mosógépet / párkapcsolatot/...)
- (3) János megjavított egy csöpögő ... (csapot / mosógépet / *lombfűrészt)

Az (1) mondatot kevesebb szóval lehet befejezni, mint a (2)-t. A (2)-es üres pozíciója többféle információt közölhet, ezért nehezebb kikövetkeztetni, milyen szó fog ott állni. Minél nagyobb a választható szavak halmazának a mérete, annál nehezebb lesz megjósolni, melyik szó fogja követni az elhangzottakat. Hogy egy szó mennyire jósolható meg, azt az *entrópiával* mérjük. Ez a fogalom Shannon (1948) információelméletéből származik, és azt fejezi ki, hogy mennyi bizonytalanság van egy esemény bekövetkezésével kapcsolatban. A (2)-es mondatban a lokális entrópia magas, ami nagyobb kognitív munkát jelent a hallgató számára. A (2) lokális entrópiáját csökkenti, ha a hiányzó pozíciót melléknévvel előzzük meg (3).

Dye és mtsai (2018) rámutattak arra, hogy nyelvi korpuszokban a nehezebben megjósolható főneveket olyan szavak előzik meg, amelyek megkönnyítik a hallgató számára, hogy azonosítsa őket. Az angolban a melléknevek látják el ezt a funkciót. Hogy a magyar nyelv entrópiacsökkentő mechanizmusait megfigyeljük, a Magyar Webkorpusz 2-ből (Nemeskey 2020) vettük ki a melléknév+főnév párokat. A korpusz teljességében 2,3 millió típust és 8,5 milliárd tokent tartalmaz. Az ebből vett minta minden tíznél gyakoribb melléknév + főnév párt tartalmaz, és összesen 1,2 millió típus és 128 millió token nagyságú. A mintánkba csak a nominatívuszi és akkuzatívuszi főnevek kerültek be.

Négy általánosítást tehetünk: i. a melléknevek gyakrabban módosítanak ritkább főneveket ($r = -0.16$, $p < 0.001$, Spearman korreláció alapján), ii. a gyakoribb főnevekhez többféle melléknév típus társulhat ($r = 0.9$, $p < 0.001$), iii. a gyakoribb főnevek előtt nehezebb megjósolni a mellékneveket: ha gyakoribb a főnév, a melléknevek összes entrópiája magasabb ($r = 0.51$, $p < 0.001$), iv. a melléknév + főnév párokban nehezebb megjósolni a főnevet a melléknévből (a két alak együttállása gyengébb), ha a főnév gyakoribb ($r = -0.14$, $p < 0.001$).





1. ábra: melléknév+főnév párok tendenciái

Dye és mtsai hasonló általánosításokat tettek az angol melléknév+főnév párok eredményei alapján. A hasonló tendencia azért érdekes, mert a magyar morfológiája nagyon különbözik az angolétól: a tematikus szerepeket nem a szórend, hanem toldalékok fejezik ki. A két különböző nyelv hasonló eredményei alátámasztják az információelméletet, mert csökkenteni igyekeznek a mentális terhet. Az ilyen típusú összehasonlító vizsgálatok segítenek feltárni azokat a mélyebb, univerzális nyelvi mechanizmusokat, amelyek a felszíni különbségek ellenére is működnek.

Hivatkozások:

- Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2018). Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. *Topics in cognitive science*, 10(1), 209-224.
- Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in cognitive sciences*, 23(5), 389-407.
- Nemeskey, D. M. (2020). Natural language processing methods for language modeling.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.