Laboratorija za eksperimentalnu psihologiju,
Odeljenje za psihologiju,
Filozofski fakultet Univerziteta u Beogradu

Da li modeli procesiranja sintakse zasnovani na veštačkim neuronskim mrežama *išta* objašnjavaju?

Vanja Subotić[1]

[1] Institut za filozofiju, Univerzitet u Beogradu

subotic_vanja@hotmail.com

Modeli procesiranja prirodnog jezika zasnovani na veštačkim neuronskim mrežama koriste se u računarskoj lingvistici i psiholingvistici još od uvođenja konekcionističke paradigme u u kognitivnu nauku. Različite kompjutacione arhitekture su tokom vremena bile implementirane u modelima — od rekurentnih neuronskih mreža devedesetih godina, do savremenih mreža sa dugoročnom i kratkoročnom memorijom i transformera. Stepen u kojem je način funkcionisanja veštačkih neuronskih mreža kompatibilan sa time kako ljudi procesiraju sintaksu je bio predmet kritike iz perspektive transformaciono-generativne gramatike, što je uglavnom dalje služilo za napad na konekcionističke ambicije da se ova paradigma ustanovi i kao hipoteza o ljudskoj kognitivnoj arhitekturi, a ne samo alternativni način modelovanja (za pregled, v. Buckner 2024).

Drugim rečima, napad se sastojao u argumentaciji da ovakvi modeli ne mogu dovoljno dobro da objasne ljudsko procesiranje sintakse usled strukturnih nedostataka, jer, bez unapred specifikovane hijerarhijske strukture, modeli ne mogu reprezentovati navodno urođeno gramatičko znanje zasnovano na hijerarhijskoj generalizaciji (za pregled, v. Chomsky 1957, 1965, 1980). Nasuprot tome, Lake & Murphy (2023: 25) ističu da je istraživačko ključno pitanje „da li su procesi modela na psihološki uverljiv način slični ljudskim procesima, te da li potencijalno pružaju uvid u ljudsku psihologiju". Suština problema se, stoga, svodi na sledeće: (A) da li ovi modeli na uverljiv način reprezentuju naše sintaksičke sposobnosti s obzirom na svoju kompjutacionu arhitekturu, i (B) da li imaju eksplanatornu vrednost u kontekstu računarske lingvistike. Tvrdim da je odgovor

na oba pitanja — (A) i (B) — potvrdan, i da ovi modeli pružaju kako-je-to moguće (how-possibly) i kako-je-to plauzibilno (how-plausibly) objašnjenja.

Svoj argument potkrepljujem: (1) usvajanjem tipa mehanističkog objašnjenja, za koje sam prethodno utvrdila da je primenljivo u slučajevima primene modela baziranih na veštačkim neuronskim mrežama u kognitivnoj neuronauci (Subotić 2024); (2) preciziranjem različitih faza otkrivanja mehanizama (Craver 2007); i (3) razlikovanjem tipova mehanističkih objašnjenja koja su trenutno izvodljiva u računarskoj lingvistici s obzirom na analizu konkretnih modela (Elman 1990, 1991; Linzen et al. 2016, McCoy, Frank i Linzen 2018).