

Absztrakt

A disszertáció az indonéz könnyűige-konstrukciókat (Light Verb Constructions, LVCs) elméleti nyelvészeti és korpuszalapú elemzés integrált megközelítésével vizsgálja. A kutatás elsődleges része empirikus alapot nyújt, négy különböző korpusz alapján értékeli az LVC-k gyakorisági és eloszlási mintázatait. A K-means klaszterezés három természetes csoportot azonosít (alacsony, közepes és magas gyakoriság), és az 1. hipotézist erős keresztkorpusz-rangsor-konzisztencia támasztja alá (Spearman $r_s = 0.891$, $p < .001$), ami megerősíti az LVC-k időbeli és műfaji stabilitását. A 2. hipotézis igazolja az eltérést a klasszikus Zipf-eloszlástól, és a Zipf–Mandelbrot-modellhez való közeledést. További modellezés alátámasztja a Menzerath–Altmann törvényt (morfémaalapú illeszkedés), míg a lexikai növekedést Heaps-törvény és Baayen-féle produktivitási mutatók részletezik. A lexikai eltérést időben Yule’s K, KL Divergencia és entrópiaindexek vizsgálják. Altmann (1967) szótáralapú modelljével való összevetés jelentős eltéréseket tár fel a morfémahossz, a típuseloszlás és a korrelációs viszonyok terén. A 3. hipotézist a klaszterek közötti morfoszemantikai-szintaktikai rétegzettség támasztja alá, amely egy strukturálisan aszimmetrikus, fokozatosan rétegzett LVC-rendszert mutat. A másodlagos rész a verbális elemek osztályozására összpontosít, megkülönböztetve a Valódi Könnyű Igéket (True Light Verbs) és a Homályos Cselekvő Igéket (Vague Action Verbs) aktionsart diagnosztikák alapján. A gépi tanulási modellek (Naïve Bayes és Random Forest) kiemelik, hogy a gyakorisági és szemantikai paraméterek erős prédikátorai az igei produktivitásnak. A záró rész a névszói összetevők elemzését végzi el, statikus és eseményszerű típusok szétválasztásával időbeli jellemzők alapján. Az eredmények azt mutatják, hogy a statikus értelmezések főként névszóvezéreltek, míg az eseményszerű olvasatok az ige–névszó kölcsönhatásából és az eloszlási mintákból erednek. A kutatás elméletileg megalapozott, adatvezérelt tipológiát nyújt az indonéz LVC-k számára, hozzájárulva a korpusznyelvészethez és a kevésbé leírt nyelvek LVC-rendszereinek modellezéséhez.