

PRAGMATIKAKO ERLAZIOZKO DISKURTSO-EGITURA: DESKRIBAPENA ETA BERE EBALUAZIOA HIZKUNTZALARITZA KONPUTAZIONALEAN

Tesiaren egilea: Mikel Iruskieta

Unibertsitatea: Euskal Herriko Unibertsitatea (UPV/EHU)

Saila: Euskal Hizkuntza eta Komunikazioa

Tesi-zuzendariak: Arantza Diaz de Ilarraza eta Mikel Lersundi

Tesiaren laburpena:

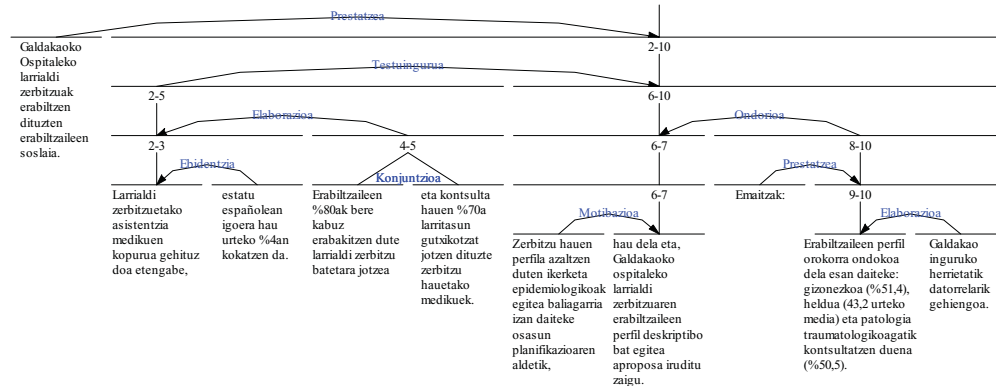
Testuek duten erlaziozko diskurtso-egitura deskribatzea da Diskurtsoaren Analisisiko zereginetako bat, Pragmatikaren barruan. Erlazio-egituran bi fenomeno nagusi nabarmendu daitezke: diskurtso-egituraren hierarkikotasuna eta testuzatien arteko koherentzia-erlazioak. Tesi honen helburua da euskarazko testuen erlazio-egitura deskribatzea erre-patroizko corpusa eraikiz, Hizkuntzalaritza Konputazionalean erabiltzeko. Ezaguna da erre-patroizko corpusak direla oinarri aplikazio aurreratuak egin ahal izateko: informazio-erazketan, laburpen automatikoan eta itzulpen automatikoan.

Horiek horrela, domeinu ezberdineko laburpen zientifikoak bildu ditugu eta erlazio-egitura deskribatzeko *Rhetorical Structure Theory* erabili dugu (W. C. Mann eta S. A. Thompson, 1987, «Rhetorical Structure Theory: A Theory of Text Organization», *Text* 8.3, 243-281), Hizkuntzalaritza Konputazionalean hainbat atazatarako baliatu izan dena. Marko teoriko horren pean etiketatze-metodologia eta ebaluatzeko metodo berritzailea proposatu dugu, etiketatzean dagoen arazorik handienari, analisten arteko subjektibotasunari, arretarik kendu gabe.

Bestalde, komunitate zientifikoaren eskuetan tesiko emaitzak jartzeko zabalkunde-lan handia egin dugu, corpus etiketatuaren gainean eraikitako web-tresnak (<http://ixa2.si.ehu.eus/diskurtsoa/>) ahalbidetzen baitu orain arte egin ezin ziren hainbat kontsulta egitea: testuko ideia nagusia eta horri lotutako erlazioak kontsultatu, analisi morfologikoan oinarritutako bilaketak egin, koherentzia-erlazioen maiztasunak eta adibideak kontsultatu, testu jakin baten erlazio-egitura ikusi eta erlazio-seinaleak kontsultatu.

Hona hemen corpus horretako testu baten erlazio-egitura eta horri dagozkion koherentzia-erlazio nagusien azalpenak: i) irakurlea tituluarekin artikulua irakurtzeko idazleak PRESTATZEN du (1-1 → 2-10); ii) aztertu beharreko gaia hobeto ulertzeko TESTUINGURUA zehazten du: larrialdi-zerbitzuetako erabiltzaileen soslaiaren zein den esanez (2-5 → 6-10); iii) lan hau interesgarria zergatik den MOTIBAZIOA erlazioarekin erakusten du (6-6 → 7-7), eta iv) laneko ONDORIOAK nabarmentzen ditu

(6-7 ← 8-10). Irudian, gezien noranzkoak erakusten du zein den unitaterik garrantzitsuena (hierarkia) eta erlazio-izenak zein den unitate horien arteko koherentzia-erlazioa. Hierarkian oinarritutako aplikazioetako bat laburpena da; izan ere, testua laburtu liteke bi gezi edo gehiago jasotzen dituzten unitateak (2, 7 eta 9) irakurriz gero.



Bukatzeke esan behar da web-tresna hori ez dela erlazio-egituraren kontsultak egiteko corpus bat gehiago, baizik eta etiketatzaileen arteko desadostasunak aztertzeko baliabidea, etorkizunean etiketatzaileen arteko adostasuna hobetzeko asmoz erabil daitekeena.