

EUSKAL LOKUZIOEN TRATAMENDU KONPUTAZIONALA

Tesiaren egilea: Ruben Urizar Enbeita

Unibertsitatea: Euskal Herriko Unibertsitatea (UPV/EHU)

Saila: Euskal Filologia saila

Tesi-zuzendaria: Iñaki Alegria Loinaz, Juan Carlos Odriozola Pereira

Tesiaren laburpena:

Hizkuntza konbinazio-sistema diskretua izanagatik, bere funtzionamendua ez dute bakarrik sistemaren arau libreek zehazten. Hiztunok, geure hizkuntz jardunean, askotariko hitz-konbinazio ihartuez baliatzen gara; konbinaziooi unitate fraseologikoak (UF) deritze. Bada, UFak oso ugariak dira hizkuntzetan eta berebiziko garrantzia dute hizkuntzaren jabetzan zein prozesaketan.

UFen artean hiru multzo nagusi bereizten ditugu. Batetik, enuntziatu osoa eratzen dutenak ditugu, hau da, enuntziatu fraseologikoak. Bestetik, sintagma osatzen dituztelarik, enuntziatuak eratu ahal izateko bestelako sintagma batzuekin konbinatu beharra dutenak. Horien artean, kolokazioak —hizkuntza-arauan finkatuak— eta lokuzioak —hizkuntza-sisteman finkatuak— bereizten dira.

Euskararen prozesamendu automatikoan, UFen tratamendua jorratu gabeko eremua zen. Hala bada, hizkuntzaren prozesamenduan hitz-konbinaziook unitate gisa ezagutzearen eta analizatzearen garrantziaz jabetuta, euskarazko lokuzioen prozesamendu automatikorako sistema bat garatzeko erronkari heldu diogu.

Helburu nagusi hori lortzeko, euskarazko lokuzioen prozesamendurako estrategia diseinatu eta oinarriak —linguistikoak zein konputazionalak— ezarri ditugu tesi-lan honetan. Lokuzioen azterketarako, aurrena kategoriaka sailkatu ditugu:

1. Izen-lokuzioak: *a bildua, euskaldun berri*
2. Aditz-lokuzioak: *adarra jo, begi bistatik galdu*
3. Adjektibo-lokuzioak: *batez besteko, estatubatuar*
4. Adberbio-lokuzioak: *behin eta berriro, hurrenez hurren*
5. Interjekzio-lokuzioak: *baita zera ere, hor konpon*
6. Zenbatzaile-lokuzioak: *apur bat, makina bat*
7. Izenordain-lokuzioak: *gutiz gehienak, zer edo zer*

8. Loturazko lokuzioak:

- Lokailu-lokuzioak: *azken finean, hain zuzen ere*
- Juntagailu-lokuzioak: *baita... ere, salbu eta*
- Menderagailu-lokuzioak: *-(e)la eta, -tu baino lehen(ago)*

9. Postposizio-lokuzioak: *-(r)en arabera, -(r)i begira*

Ondoren, lokuzio mota horien guztien deskripzio zabala egin dugu, bakoitzaren ‘osaera’, ‘aldagarritasuna’ eta ‘anbiguotasun semantikoa’ aztertuz.

Hala, euskarazko lokuzioen ezaugarriak aztertuta, eta eskuragarri ditugun baliabideak kontuan harturik, euskarazko lokuzioen prozesamendurako, egoera finituetako sisteman oinarritu gara. Horretarako, IXA taldeak garaturiko tresnak (batez ere MORFEUS analizatzaile morfosintaktikoa eta EDBL, Euskararen Datu-Base Lexikala) eta Murriztapen Gramatika (MG) formalismoa baliatu ditugu, besteak beste.

Halaber, diseinaturiko estrategiaren aplikazio errealak gauzatu ditugu. Horretarako, bi multzotan banatu ditugu lokuzioak, euren prozesamendurako bide ezberdinak eskatzen zituzten aldetik. Batetik, osagai guztiak hitz beregainak dituzten 2.200 bat lokuzio deskribatu ditugu EDBLn. Lokuzioak testuetan har ditzakeen azaleko gauzatze guztiak formalki kodetu ditugu, adierazpide orokor baten bidez (gauzatze-eskema). Bestetik, postposizio-lokuzioak (472 konbinazio) eta menderagailu-lokuzioak (131 konbinazio) detektatzeko, MG gramatika bana eraiki dugu.