

## Euskarazko testuen sinplifikazio automatikoa

Itziar Gonzalez-Dios

Ixa Taldea, UPV/EHU

Ikerketa-lan honetan euskararen tratamendu automatikorako baliagarria izango den idatzizko testuen sinplifikazio automatikoa aztertzen ari gara. Lan hau Ixa taldean<sup>1</sup> egiten da, hizkuntzaren tratamendu automatikoan lan egiten duen Euskal Herriko Unibertsitateko ikerkuntza-taldea.

Testuak sinplifikatzea edo erraztea, ingelesez *Text Simplification*, testu zail bat errazago edo ulergarriago bihurtzea da, jatorrizko testuaren esanahia mantenduz. Sinplifikazioa hainbat mailatan egin daiteke: lexikoa, sintaxia...; esaterako, sintaxian egitura sintaktiko konplexuak (perpaus koordinatuak, mendeko perpausak, aposizioak...) erraztea da helburu, lexikoan, berriz, hitz zailak edo maiztasun gutxikoak ezagunagoengatik edo maiztasun handiagokoengatik ordezkatzeta da, etab.

Testu errazek abantaila ugari eskaintzen dizkiete, bai gizakiei, baita makinei ere. Gizakien kasuan, hainbat azpitaldetara bidera daiteke sinplifikazioa: 1) Atzerriko hizkuntzak ikasten ari direnak, 2) Analfabeto edo gutxi alfabetatuak, 3) Haurrak, 4) Afasikoak, 5) Gorrak, 6) Gaixotasun kognitiboak dituztenak eta abar. Makinei dagokienez, sinplifikazioa aurreprozesu moduan erabil daiteke HPko zenbait sistematan: analizatzaile sintaktikoak (*parser*), itzulpen automatikoa, galde-erantzun sistemak, laburpen automatikoak egiteko sistemak; beste aparatuek (mugikorak, PDAk, tabletak...) dituzten pantaila txikietan informazioa modu erosoan agertzeko ere baliagarria izan daiteke.

Testuen sinplifikazio automatikoan lan gehienak ingeleserako egin diren arren, azken urteotan beste hizkuntzetarako ere sistemak sortzen hasi dira, hala nola Brasilgo portugueserako, arabierarako, suedierarako, gaztelaniarako, frantseserako, danierarako, euskararako eta italierarako. Gainera, azken urte hauetan ikerketa-lerro hau garrantzitsua bihurtu da eta uneoro testuak sinplifikatzeko metodo berriak argitaratzen ari dira.

Euskarari dagokionez, gure lana bi atazatan bana daiteke: alde batetik, euskarazko egitura sintaktiko konplexuen azterketa egiten ari gara eta, bestetik, testuen sinplifikazioa automatikoki egingo duten tresnen egokitzapenean edo garapenean dihardugu. Zeregin horretan perpausen mugak identifikatzen dituen gramatika eta aposizioak detektatu eta sailkatzen dituen gramatika osatu ditugu.

---

1 <http://ixa.si.ehu.es/Ixa>