

EUSKARAREN EZAGUTZA-BASE LEXIKALA: EUSKAL WORDNET

Tesiaren egilea: Elisabete Pociello Irigoien

Unibertsitatea: Euskal Herriko Unibertsitatea

Saila: Euskal Filologia

Tesi-zuzendaria: Eneko Agirre Bengoa eta Izaskun Aldezabal Roteta

Tesiaren laburpena:

Semantika konputazionalan aurrera egin ahal izateko, ezagutza-base lexikalak (EBLak) ezinbestekoak dira. Behar horri erantzutera dator hemen aurkezten dugun lana. Zehazki, *WordNet* deritzon EBLaren euskarako baliokidea (Euskal WordNet) eraikitzen emandako pausoak eta egindako ekarpenak azaltzen ditugu.

Lehenengo, nahiko genukeen EBL motaren ezaugarriak definitu ditugu, eta, ondoren, azterketa bibliografikoa egin dugu, ezaugarri horietara gehien egokitzen den formalismoaren bila. Hala, azterketa horren ondorioz, erabaki dugu *WordNet* eta horren ildotik garatu diren *EuroWordNet* eta *The Multilingual Central Repository* (MCR) eredueta oinarritzea euskarako EBLa.

EBLa eraikitzeke jarraitutako metodologiari dagokionez, WordNeteko ingeleseko kontzeptuak abiapuntutzat hartu ditugu eta horiei euskarako ordainak lotu dizkiegu. Horretarako, hasieran, estaldurari eman genion garrantzia —ingeleseko kontzeptu guztien euskal ordainak automatikoki sartuz—, eta gero ordain horien kalitateari erreparatu genion, kontzeptuz kontzeptuko eskuzko orrazketaren bitartez lehenengo, hitzez hitzeko eskuzko orrazketaren bitartez ondoren, eta eskuz etiketatzen ari garen corpus bateko (*EuSemcor* [Agirre *et al.*, 2006]) informazioa erabiliz gaur egun.

Euskal WordNeten egin diren orrazketa horietan —kontuan izanda ingeleseko WordNetaren gainean ari garela garatzen euskarakoa—, ingelesaren eta euskararen arteko hainbat bereizgarri linguistiko azaleratu dira: lexikalizazioari dagozkionak eta hierarkiari dagozkionak, hain zuzen ere. Bi bereizgarri horien azterketa egin dugu, eta EBLaren eraikuntzan horrelakoak topatzean zer irizpideri jarraitu behar zaion ere definitu dugu.

Azkenik, Euskal WordNet informazio sintaktiko-semantiko gehiagorekin aberasteko asmoarekin, ingeleseko kirol-aditz batzuentzat automatikoki eskuratutako objektu/subjektuen hautapen-murriztapenen azterketa egin dugu, bi helbururekin: batetik, hainbat eskuratze-teknika erabiliz ingeleseko eta euskarako corpus ezberdinetatik eskuratutako hautapen-murriztapenak aztertzea, konparatzea eta ebaluatzea; eta bestetik, hautapen-murriztapenak eleanitzak diren ala ez egiaztatzea.