



Rosa Miren Pagolak zuzendutako ikerlanaren aurkezpena

## Bonaparte corpusari teknologia berriak aplikatzea: web orriaren aurkezpena

2011ko Hizkuntzen Astearen barruan (martxoaren 8tik 10era), euskal testuen analisirako ikerlan hau aurkeztu zen:

Hizkuntzaren teknologia Bonaparteren corpusari aplikatuak: web orriaren aurkezpena. Aurkeztu den web orriari ageri den informazioa *TesiTek* izeneko diziplinarteko taldeak azken urteotan egin duen ikerlanari dagokio. Tresna informatiko bat garatu da euskal testuen

azterketa linguistikoa egiteko, berrerrabiltzeko *corpus* digital baten ediziotik abiatuta. Emaitzak <http://bonaparte.deusto.es> webean daude, aplikazioak zabalik dituela eta erabiltzaile guztien eskura. Corpora Bonaparte ondareko dokumentuek osatzen dute, euskaraz dagoen dokumentu bildumarik garrantzizkoena euskalkien eta diakroniaren ikuspegitik. Edizioari atxikita, testuen oharpenez gain, halako edizioetan izan ohi diren elementuak ageri dira: deskripzio bibliografikoa, dokumentuei buruzko informazioa, iturrien aipamena, etab. Testuen oharpenari dagokionez, XML markazio hizkuntza erabili da, TEI (*Text Encoding Initiative*) etiketatzearen bidez. Nazioartean ezaguna den markazio sistema hori testu digitalizatuak aurkezteko erabiltzen da. Gure kasuan, ikerketaren alderik berritzaileenera bideratuta dago: euskalki guztien hainbat garaitako testuen analisi (morfologiko eta sintaktiko) erdi automatikoa egiteko modua emango duen tresna informatiko bat garatzera.