



Arantza Diaz de Ilarraza
EHUko Informatika Fakultateko irakasle eta ikerlaria

Ikerketa azpiegitura digitalen garrantzia

Azkeneko urteetan berrikuntza sozial eta kulturalako ikerketa azpiegiturretan asko inbertitu da European, duten potentziala dela eta. Iker-

keta azpiegitura digitalek ikerkuntza bultzatzen dute, lankidetzaz, diziplinartekotasuna, datu zientifikoak erabiltzeko eta haietaz baliatzeko ezagutzak, trebetasunak eta gaitasunak ematen baitituzte. Azpiegitura horiek ezagutza aurreratzeko sortuak dira, esparru zientifiko anitzeko komunitate zabal bati begira eta atzipen askeko eta lehia askeko FAIR printzipioei jarraituz. ES-FRIren bide orrian badaude gizarte eta giza zientzien arloetako ikerketa azpiegiturak eraikitzen dituzten proiektuak; CLARIN da horietako bat.

CLARINen xedea ahozko zein idatzizko hizkuntzaren azterketari laguntzeko azpiegitura ireki bat eskaintzea da: fenomeno sozialak eta dinamika kulturalak islatuko dituen datu multzo gisa, batetik, eta datu horiek modu errazean eta jakintza tekniko gutxiarekin ustiatzeko tresnak lantzen dituen, bestetik, ikerkuntza diziplinarteko testuinguru digitalean egin ahal izateko. CLARINek sarbide erraza eta iraunkorra eskaintzen du datu linguistiko digital elebakar zein eleaniztunetara (izan idatzizko, ahozko, bideo edo multimodal); bestalde, baita tresna aurreratuak ere datuak aurkitzeko, esploratzeko, ustiatzeko, etiketatzeko eta aztertzeko, datuok gordetzen diren zentroen kokapen fisikoa edozein dela ere. Asmoa da hizkuntzarekin lotuta dagoen teknologiarako sarbidea modu bateratuan ematea, muga linguistiko, nazional, kultural eta denborazkoetatik haragoko ikerketa konparatiboa bultzatzeko.

CLARIN mota askotariko datuak integratzeko sortua da; besteak beste, honako hauek: Europako Parlamentuko saioetako erregistroak (ParlaMint), ahozko corpusak, ikasleen corpusak,

literatura testuak, sare sozialetako datuak, dokumentu historikoak, ahoz transmititutako historia datuak nahiz erakundeen artxiboetakoak, irrati artxiboak, eta abar. Adibidez, Parlamentuko grabazioen eta testuen azterketak oso garrantzizkoak dira diskurtso politikoan, jarrera historikoak, erlijio-soak edo kulturalak islatzeko. Ikertzaileek Talk of Europe datu multzoa erabili izan dute denboran zehar termino politikoak aztertzeko: Europako Parlamentuan finantza krisiaren eztabaidentan, adibidez; edo hizkuntza aztertze-ko emozioen garraiatzaile gisa, eta hizkuntzak dinamika kultural eta sozialekin erlazionatutako beste fenomeno batzuekin duen korrelazioa neurtzeko.

Datuak intentsiboak diren ekosistema batean lan egiteko aukerari esker, gizarte zientzia eta humanitateetako ikerketa iturriek eta metodoek izugarrizko eraldaketa izan dute. Kultura ondarearen material bolumen handiak digitalizatu dira, eta, sorreraz digitala den materiala handituz doan heinean, aurrekaririk gabeko modu batean handitzen ari da metodologia berrien eta tresnen eskaera. Ikerketa azpiegitura ondo hornituz eta beharrak modu egokian kudeatuz erantzun dakioko eskae-

ra horri, ikerketa paradigma berriek gizarte eta kultura berrikuntzarako duten potentziala handitzera orientatuz. Esaterako, azpiegitura honetan, lehen aldi luze batean egiten zen ikerketa askoz denbora gutxiagoan egin daiteke egun, testu bolumen handiagoa era gardenagoan.

CLARINek eskaintzen duen inguru-nean ikerketa ugari egiten ari dira: migrazio ereduak, historia intelektuala, denbora eta eskualdeen arteko hizkuntza aldaerak, eta baita parlamentuko diskurtsoa eta generoa aztertzen dituztenak ere, gutxi batzuk aipatzearen.

CLARINen azpiegitura eraiki dute bertako kideek (eta behatzaileek), eta kideen artean finantzatzen dute azpiegitura hori. Estatuak edo gobernuarteko erakundeak (unibertsitate, ikerketa erakunde, liburutegi eta artxibo publikoak) izan ohi dira CLARINeko kide edo behatzaile. Estatu bakoitzeko kideek “partzuergo nazionala” osatzen dute, eta beraien arteko bat da koordinatzailea. CLARINeko jakintza zentroek (K-Centre) sarbidea eman behar dute bildutako ezagutza horretara guztira ikertzaileek datu eta tresnokin modu eraginkorrean lan egin dezaten.

Europako herrialde gehienek urteak daramatzate proiektu honetan, baina Espainiako eta Frantziako estatuak ez dira oraindik CLARINeko kide (Frantzia behatzaile da); beraz, estatu horietako hizkuntzak zeharka baino ez daude bertan, beste hizkuntzetarako sortu diren tresna eta produktuen artean. Euskara ez dago presente CLARINen, eta euskararako garatutako hainbat tresna ez daude bertako azpiegituran txertatuta. Tresna horiek azpiegitura horretan jarri ezean eta azpiegitura horretan garaturiko hainbat hizkuntza teknologia baliatu ezean, aukera bat galtzen ari gara hizkuntza komunitate gisa sozialki zein kulturalki aurrera egiteko eta berritzeko. •



CLARIN proiektuaren logoa.