

Euskal Hiztegia prozesamendurako baliabide gisa

Jose Mari Arriola

UPV/EHU

Laburpena

Euskal Hiztegia *prozesamendurako baliabide lexikal gisa hartzeak izan dituen zenbait emaitza eta, hiztegi gintzari begira ere, hiztegiak egiteko beste modu batzuen premiaz arituko gara, hiztegi gintzari informatikak eta bereziki corpus hizkuntzalaritzak eman dion astinduz, alegia. Azken urteetan, areagotu egin dira hiztegien eta corpusen arteko loturei buruzko azterketak. Izan ere, corpusak ezinbesteko bilakatu dira hiztegi gintzan. Hala ere, hiztegien eta corpusen arteko interakzioaz dagoen aniztasunaz gain, aniztasun horren atzean hurbilpen teoriko eta kontzeptualizazio desberdinak ezkutatu ohi direla ikus dezakegu.*

0. Sarrera

Ibon Sarasolak *Euskal Hiztegia* (EH, 1996) lantzerakoan, seguru asko, nekez pentsa zezakeen Hizkuntzalaritza Konputazionalaren aztergai edota lehengai izan zitekeenik bere hiztegia. Izan ere, giza erabiltzaileari zuzenduriko hiztegi orokorra dugu, ez zen berariaz prozesamendurako sortu. Baina, bereziki 80. hamarkadan piztu zen interesa edota beharra, alegia hiztegiak prozesamendurako baliatzearena, EHri ere egokitu zitzaion, IXA¹ taldeak baliabide lexikalen aldetik zuen beharrari erantzuteko, batik bat. Hain zuzen ere, hizkuntzaren prozesamenduari begira egindako horietako zenbait lan bilduko ditugu. Hala ere, biltze-lan horri heldu aurretik eta, laburki bada ere, hiztegi gintzari begiratu azaleko bat eman nahi nioke eta, aski ezaguna den arren, Sarasolak hiztegi gintzan egin duen lana ere gogoratuko dut gain-gaienetik.

Hizkuntza-teknologiak lagun, hiztegi gintzan aldaketa sakonak gertatu dira azken hamarkadan, bereziki corpus hizkuntzalaritzak eta corpusak usiatzeko teknologiak lexikografia irauli egin dutela esan dezakegu. Ildo horretatik, Hanks-ek (2012) corpus hizkuntzalaritzaren iraultza nabarmentzen

¹ <http://ixa.si.ehu.es/Ixa>

du. Espekulazioari corpusetako ebidentziek hartu diote aurrea eta, horrez gain, hiztegegintzaren parte handiena automatizatu egin da, formatu elektronikoa landuko dira eta hortik eratorriko dira merkaturatuko diren hiztegiak ere, gehienetan formatu digitalean. Euskal Hiztegegintzan ari diren UZEI² eta Elhuyar³ kontuan hartzen baditugu, garbi ikus dezakegu horiek ere iraultza horretan murgilduta daudela. Elhuyarrek hiztegien modernizazioan hartu duen bidea erakusten digu Leturiak (2013)⁴ eta, etorkizunari begira esaten diuenak ekarriko ditut hona:

Orain arte landu ditugun mota horretako teknologia gehienek hiztegirako hitz eta terminoak eta haien ordainak erazten zituzten corpusetatik; baina, horiez gain, hiztegi batek definizioak, adierak eta adibideak behar ditu. Bada, orain hasi gara horiek modu automatikoan lortzeko modua ere lantzen, hau da, hitz baten definizioak, adierak eta horietarako adibide egokiak testuetatik edota webetik automatikoki erazten.

Prozesu horretan, hiztegiak egiteko moduari erreparatzen baldin badiogu, jauzi ikaragarria gertatu zen garai batean zapata-kutxetan gordetzen ziren fitxetatik, informazio hori ordenagailuen bidez lantzen hastearekin; nahiz eta hasiera batean, testu-prozesadore soilak baino ez erabili, ondoren, etorriko baitziren datu-baseak. Halaber, jauzi nabarmena gertatu zen corpora hiztegien lehengai bihurtzen hasten denetik.

Laburbilduz, bi alde horiek lirake azpimarragarri: batetik, hiztegiak egiteko moduak automatizatzea eta, bestetik, corpora lehengai izatea hiztegi-tako artikuluen informazioa hornitzeko, hala nola sarrerak, adibideak, definizioak eta abar. Ildo horretatik, uste dut Ibon Sarasolak hiztegegintzari ekiterakoan jorratu zituen zenbait gai ere izan dutela eraginik euskal hiztegegintzaren modernizazio-prozesuan, besteak beste, honako hauek: hiztegiaren tamaina, hiztegi-sarreraren hautaketa, corpusak ustiatzea adibideak hornitzeko eta definizioak lantzeko ereduak.

Hiztegegintzako lanetatik, *Orotariko Euskal Hiztegia* kontuan hartzen badugu, nolabait, Sarasolak⁵ berak «filologia industrial» deitu izan duen aroaren sustatzaile dugu. Seguru asko, ingeniariaren buruak izan zuen eraginik hiztegegintzarekin lotutako hainbat gai antolatzen: hiztegiaren tamaina, hiztegi-sarreraren hautaketa eta hiztegia aurrera eramateko beharrezkoak diren azpiegiturari dagozkion beste hainbat gai (zenbat orrialde inprimatu eguneko, horretarako behar zen tresneria, eta abar).

² <http://www.uzei.com/igb-2/lexicon-and-dictionaries/>

³ <http://www.elhuyar.org/hizkuntza-zerbitzuak/es/Hiztegegintza-zerbitzuak>

⁴ Igor Leturia, «Hiztegegintza Teknologiaz modernizatzen», *ELHUYAR aldizkaria*, [<http://zientzia.net/artikuluak/hiztegegintza-teknologiaz-modernizatzen/>], 2013-10-10.

⁵ Ibon Sarasola. 2012. «*Orotariko Euskal Hiztegia* eta euskal hiztegegintza». *Koldo Mitxelena Katedraren III Biltzarra*. [<https://www.youtube.com/watch?v=zx5GjrlOufY&noredirect=1>]

Halaber, Azkaratek (2008), Sarasolak hiztegegintzari egin dizkion ekarpenetatik lexikoa sistematizatzeke ahalegina nabarmentzen du:

Hau da, euskalkietako hitz-ondaretik abiatuta, batasunak nahitaezko duen sistematizazio eta mailakatze-lana da tradizioaren altxorrean *Hiztegi Batuak* dakarrena. Eta kontu honetan, hiztegegintza lan osoan bezala, ezinbestekoa da Ibon Sarasolaren lana aitortzea, bera izan zen eta Arantzazuko Biltzarren ondotik, *Euskal Hiztegi Arauemailea* lantzean lehenik eta *Euskal Hiztegiaren* bidez ondoren, euskal lexikoaren itsaso zabala sistematizatu eta mailak ezartzen behin eta berriz saiatu dena.

Definizioez den bezainbatean ere, (Alberdi et al 2008) lanean frogatzen dute Sarasolak atal horretan ere euskal lexikografiarako oinarri sendoak jarri zituela:

Apenas existe una tradición lexicográfica monolingüe vasca. Se puede decir que dicha tradición es muy reciente y que se asienta sobre las bases establecidas por Ibon Sarasola en su *Euskal Hiztegia*. Pues bien, en dicho diccionario se fijan unas estrategias sintáctico-discursivas para la definición que se han probado adecuadas, y que, más allá de la lexicografía, entroncan perfectamente con la tradición escrita: una de las claves de dichas estrategias es el uso mesurado de estructuras apositivas para añadir especificaciones al núcleo informativo (hiperónimo).

Sarrera gisa, Ibon Sarasolak hiztegegintzan jokatu duen rolaz zertzelada batzuk eman ondoren, hurrengo atalak honela egituratu diot: lehenengoan, EHTik hizkuntzaren prozesamendurako informazio lexikala eskuratzeko zenbait lanen aipamena laburbilduta; bigarrean, berriz, prozesamenduari begira egindako lan horiek hiztegegintzari egin dizkioten zenbait ekarpen eta, ildo horretatik, hirugarren atala, corpusen eta hiztegien arteko interakzioaz dagoen eztabaidari eskaintzen diot; azkenik zenbait ondorio.

1. *Euskal Hiztegia* euskararen prozesamendurako oinarri

Literaturan hainbat aldiz aipaturiko «lengoaia naturalaren arloko zintzur-gune estua» (Walker et al 1995), alegia lexikoiak ezagutza lexikalaz hornitzea izan da arazo nagusietariko bat Lengoaia Naturalaren Prozesamenduan (LNP). Izan ere, aplikazioek neurri errealeko eta erabilera desberdinetarako prestaturiko lexikoi konputazional erraldoiak behar dituzte. Arazo horri aurre egiteko, bi irtenbide nagusi jorratu ziren: 1) dauden baliabideak berrerabilteza eta 2) baliabide berriak eraikitzea. Lehenengo bideari ekin zitzaion nagusiki 80. hamarkadan. Izan ere, hizkuntzalaritza konputazionalaren arloan, hedatuta zegoen hiztegiak helburu konputazionalarekin erabiltzea, bereziki, euskarri magnetikoan zeudenak (*Machine Readable Dictionaries*, MRD).

Izan ere, hiztegi horiek baliabide lexikal oparotzat hartzen ziren, lexikografoen hainbat urtetako lanaren ondorioz informazio lexikal aberatsa zutela pentsatzen zen. Behar horri erantzuteko, IXA taldean askotariko baliabide lexikalak erabili izan dira (Artola 2004) eta, horien artean, hiztegien atalean, EH zertarako baliatu den ikusiko dugu.

1.1. *Euskararen Datu-Base Lexikala (EDBL) hiztegi-sarrera gehiagoz hornitzeko*

Euskal hiztegi gintzan, elebakarrean behintzat, iturri garrantzitsua izan da EH. IXAk euskara estandarerako aplikazioak jomugan izanik eta, kontuan harturik euskal hiztegi gehienek EHTik hartzen zituztela hiztegi-sarreak behintzat; kontsentsu hori aprobetxatzea izan zen gure asmoa. Ondoren, Hiztegi Batuaren bidez (Euskaltzaindia 2000), Euskaltzaindiak ofizialtasuna eman dio batasun horri. Estandarizatutako hitz horiek ere EDBLn izango dute sarrerarik, noski. Halaber, beste hiztegi batzuk ere baliatu izan dira, hala nola Elhuyarren Hiztegia (Elhuyar 1998), *Basque-Spanish/Spanish-Basque dictionary* (Morris 1998), sinonimoen hiztegia (UZEI 1999). Bestelako hiztegiez baliatzea beharrezkoa izan dugu, EHk literatur tradizioa oinarri duenez gero, askotan erabilera berriagoetan ageri diren hitzak jasotzen ez dituelako.

Bestalde, sarrera-hornitze horretan zuzentzaile-egiaztaizailearen edota analizatzaile morfologikoaren erabiltzaileek ere parte hartuko dute, bereziki, hutsuneak edota akatsak antzematen.

1.2. *Euskal WordNet osatzeko*

Euskal WordNet eta EHren arteko mapaketa edota bateratze bat egitea izan zen helburu, hau da, Euskal WordNet⁶ EHTik eratortzen diren erlazio lexiko-semantikoekin aberastea. Halaber, interpretazio semantikorako beharrezkoak diren hainbat eta hainbat erlazio lexiko-semantiko txertatu ziren (adibidez garagardoa garagarrez egina dagoela, okinaren lanbidea ogia egitea dela, arrantza pertsonen —eta astoek— egiten dutela, eta abar.).

Horretarako, Euskal Wordneten adierak eta EHrenak lotuko dira. Oro har, lotura horiek bat batekin edota hainbat batekin izan dira; hainbat hainbat batekin, aldiz, ez da aurkitu. Izan ere, WordNeten egiten den adiera banaketa xeheagoa da. Edozein modutan, ez da harritzekoa bat ez etortze hori, izan ere, adiera-banaketa oso irristakorra da.

Bestalde, sarreren kasuan bezala, literatur tradizioan oinarritzeak dakarren gabezia dugu, zenbait adiera ez agertzearena alegia.

⁶ <http://ixa2.si.ehu.es/mcr/>

1.3. *Aditzen patroiak erauzteko adibideetatik*

EHko aditzen adibideak aztertu ziren (Arriola et al 1999; Arriola 2000), aditzen oinarritzko azpikategorizazio-patroiak erauzteko xedearekin. Aditzen adibide guztiak analizatu ziren *Constraint Grammar* formalismoa (Karlsson et al 1995) euskararako murriztapen-gramatikaren bidez aplikatuta, baina gramatika horren bidez lortutako analisi sintaktikoak ez ziren izan behar bezain fidagarriak. Beraz, emaitza erdi-automatiko horiek ezin izan ziren baliatu EDBL aberasteko edota analizatzailea hobetzeko. Hala ere, azaleko syntaxian egindako lan horrek erakutsi zuen zer atal sendotu behar ziren gramatika konputazionalan.

1.4. *Erlazio lexiko-semantikoak erauzteko definizioetatik*

EHTik eratorritako hiztegi ezagutza-basea eraikitzeke, hiztegiko definizioak erdi-automatikoki analizatu ziren erlazio lexiko-semantikoak erauzteko (Agirre et al 2001). Analisi horien emaitzak EHren Ezagutza-basearen kontzeptuak sailkatzeko atalean gorde ziren. EDBLren eta EHren Hiztegi Ezagutza-basearen arteko erlazioak lexikoari buruzko informazio gramatikala eta semantikoa erabiltzeko aukera eskaintzen digu.

Laburbilduz, giza erabiltzaileari zuzenduriko hiztegia hainbat zereginetarako baliatu izan da Hizkuntza Naturalaren Prozesamenduaren arloan. Emaitza horiek ikusita, esan dezakegu dauden baliabide lexikalak berrerabiltzea eta, hutsetik hasi ordez, existitzen dena aintzat hartzea hurbilpen egokia dela.

2. *Euskal Hiztegia LNPrek ikuspetigik aztertzearen zenbait ondorio*

Aurreko atalean laburbildu ditugun LNPko zenbait lanetan, ikusi dugu hainbat informazio-mota eskuratu ahal izan zirela EHTik, beraz berrerabilgarria izan zen. Hala ere, beharbada, hutsune nabarmenenak aplikazioek ezagutu beharreko hainbat hitz edota adiera ez agertzearena izan zen. Gabezia horiek berriro ere, pil-pilean dagoen corpusaren gaira eramango gaituzte 3. atalean.

Artikuluen adibideez esate baterako, Arriola (2000) lanean lexikografoak hautaturiko adibideak egokitzen jotzen dira aditzen patroio sintaktikoak aztertzeko. Izan ere, corpusarekin alderatuta aztergai genuen aditza bertan aurkituko genuen behintzat. Dena den, adibideen «egokitasunaz» hitz egiterakoan autore guztiak ez datoz bat, eta Arriola (2000) lanean bertan ere, nabaritzen da nolabaiteko kezka hori. Haatik, adibideen alde onak eta ez hain onak jasotzen dira eta, EHko adibideez osaturikoa, corpus-eredu «bitxi» gisa aurkezten du. Batetik, corpusetik atereak direlako eta, bestetik, lexikografoak aukatzen dituelako. Bataren eta bestearen aldeak direla eta, maiztasuna ekarriko

dugu gogora. Izan ere, corpusek hiztegiatako adibideek ez bezala, estatistika baliatzeko aukera ematen dutela aipatzen zen. Hala ere, lan hartan bezala, esango nuke kontua ez dela soilik estatistikoki esanguratsuak diren esaldien azterketa egitea. Esaldi guztiak lirateke interesgarri hizkuntzari buruzko eza-gutza osatzeko, ez bakarrik maizen gertatzen direnak. Baina, aldiz, erabilera «jatorren» argudioak alde batera utzi behar direla iruditzen zait.

EHko zenbait atal errebisatu eta berraztertu ondoren, hiztegiak eta corpusak konbinatzeko beharrari eutsiko nioke.

Edozein modutan, berrerabili ahal izateko ezinbesteko lehenbiziko urratsa ekarriko dugu gogora lehendabizi, hiztegiaren prestatze-lana hain zuzen ere. Bereziki, EH erdi-automatikoki prozesatu ahal izateko eta informazioa eskuratu ahal izateko ezinbesteko urratsa izan zen. Hala ere, aldi berean, hiztegi-gintzari edota lexikografoaren lanarekin zerikusia duten zenbait gai ere ezin aipatu gabe utzi. Izan ere, lexikografoak buruan zuen egitura azaleratze-rakoan eta egitura horren adierazpidea formalizatzerakoan⁷ Arriolak (2000) garbi ikusten zuen hiztegi-gileari lana modu sistematikoagoan egiteko laguntzak eskaini behar zitzaizkiola. Ildo horretatik, hiztegiak editatzeko inguru-neak oso lagungarriak izango dira lexikografoak hiztegia modu errazean eta malguan eguneratzerako.

Bestalde, egitura-formalizatze horrek EHren bertsio elektronikoa (Arregi et al 2003, 2007) ahalbideratu zuen. Bertsio elektronikoa horretan *Hiztegi Baturko* informazioa ere gehitu zen modu esplizituan. Baina, hiztegi-gileari begira ez ezik, erabiltzaileari begira ere onurak ekarriko zituen bertsio elektronikoa garatu zen. EHren bertsio elektronikoa egituratu hori CD-ROMean, interneten edo testu-prozesadorean integra zitekeen. Halaber, hiztegi-kontsulta aurreratuagoak jarri nahi ziren erabiltzailearen eskura, hiztegi-sarrera bateko eremu jakin bati buruz galdetu ahal izango zen, esate baterako. Tamalez, bertsio elektronikoa hori prototipoan baino ez zen geratu.

3. Corpusak eta hiztegiak

Ezin uka daiteke, gaur egun, corpora lexikografoaren ezinbesteko lanabesa bilakatu denik. SEGen ere atarian,⁸ esate baterako, corpusen garrantzia azpimarratzen da, bai hiztegi-gintzari, bai hizkuntzalaritzari begira, oro har:

Hiztegiek informazio handia behar dute, hizkuntzaren erabileraren inguruko informazioa: hitzak, bai, sarrera egokiak aukeratzeko, baina hitzekin bakarrik ez da aski, hitzak beren testuinguruetan ikusi behar dira, ongi zehazteko zer adierazi nahi duen hitz bakoitzak, nola interpretatu behar den sarrera

⁷ EHren bertsio egituratua lortzeko, TEIko (*Text Encoding Initiative*) gida-lerroei jarraiki; SGML (*Standard Generalized Markup Language*) lengoiaia estandarra erabiliko da.

⁸ Pello Salaburu (2011). [<http://www.ehu.es/seg/hizt/6>]

bakoitza. Corpusek egiazko hizkuntza adibideak biltzen dituzte, eta ezinbesteko bihurtu dira hizkuntzaren zinezko erabilera aztertzeko.

COBUILD (Collins Birmingham University International Language Database) taldeko lexikografoak izan ziren aitzindari corpus informatizatuak erabiltzen hiztegegintzan. Izan ere, *Collins Cobuild Dictionary of English Language* (1987) hiztegia egiteko, ingelesaren corpus (20 milioi) adierazgarria zorrotzasunez aztertu zuten. COBUILD hiztegiaren lehen edizioaren sarrean, Sinclair-ek maisuki adierazten du urrats horren garrantzia:

For the first time, a dictionary has been compiled by the thorough examination of a representative group of English texts, spoken and written, running to many millions of words. This means that in addition to all the tools of the conventional dictionary makers —wide reading and experience of English, other dictionaries and of course eyes and ears— this dictionary is based on hard, measurable evidence. No major uses are missed, and the number of times a use occurs has a strong influence on the way the entries are organized. Equally, the large group of texts, called the corpus, gives us reasonable grounds for omitting many uses and word-forms that do not occur in it. It is difficult for a conventional dictionary, in the absence of evidence, to decide what to leave out, and a lot of quite misleading information is thus preserved in the tradition of lexicography (Sinclair et al 1987: xv).

Aipamen horretatik, bi ezaugarri azpimarra nahi nituzke: batetik, corpus adierazgarria osatzearena eta, bestetik, neur daitezkeen ebidentzietan oinarritzeko aukera. Bi gai horiek eztabaida iturri dira eta, helburuen arabera, irizpideak moldatu egin izan dira. Hain zuzen ere, corpusen adierazgarritasunari dagokionez, Igone Zabalak (2013b)⁹ ere Sinclair maisuaren hitzak jasotzen ditu eta, horrez gain, corpus orekatuek bete beharreko irizpideak jorratzen ditu. Edozein modutan, Igone Zabalak dioen bezala, oso kontziente izan behar gara esku artean dugun corpusaren ezaugarriez datuak aztertzeko eta interpretatzeko orduan. Hala ere, nik uste dut corpusen inguruan sortzen diren eztabaidagai edota aztergai horiek guztiek ez luketela ezkutatu behar corpusaren analisiak hizkuntzari buruz dugun ezagutza handitzen lagun diezagukeela.

Corpus hizkuntzalaritza diziplina beregaina den aldetik, alde askotariko gaiak biltzen ditu. Baina, horiek alde batera utzi eta, corpusen eta hiztegien arteko harremanari dagokionez, Igone Zabalak (2013a)¹⁰ berak esandakoak oso egokiak iruditzen zaizkit:

Hiztegien eta corpusen arteko harremana gurpil zoro baten modukoa da, zeren eta corpusak beharrezkoak baitira lan lexikografikoetarako baina, bestalde, hiztegiek eskaintzen dituzten informazioek eragin handia baitute

⁹ <http://31eskutik.com/2013/04/01/corpusen-adierazgarritasuna/>

¹⁰ <http://31eskutik.com/2013/03/01/corpusak-eta-hiztegiak/>

kontsultak egiten dituzten hiztunek egingo dituzten erabileretan eta, azken finean, etorkizunean corpusetan aurkituko ditugun hitz, adiera eta erabileretan.

Gurpil-zoro hori dela eta, corpusaren eta hiztegiaren arteko interakzioari helduko diogu. Esate baterako, zenbait egileren arabera (Agirre et al 2006), hiztegiak hitzen adierak emateaz arduratzen dira, hitz hauen testuinguruez gehiegi arduratu gabe. Aldiz, corpusek hitzaren testuinguruari eta erabilerari buruzko informazioa eskaintzen digute. Horrela bada, bi baliabide hauen abantaileri probetxu ateratzearen, hiztegiak eta corpusak elkarrekin erabiltzea proposatzen dute. Izan ere, hain garrantzitsua izanik corpusa, badira zenbait egile proposatzen dutenak corpusa bera ez ote den izango etorkizuneko hiztegi (Alonso 2009). Ildo horretatik, corpusaren eta hiztegiaren arteko interakzioari gagozkio, egokia iruditzen zait bereizketa hau: Kilgarriff (2005) lanean bi bide bereizten dira:

1) «poner el corpus en el diccionario» (PCED), que representa la actual Lexicografía basada en corpus; 2) «poner el diccionario en el corpus» (PDEC), que sería la defendida por los que codifican la información en el corpus, tal y como se hace en algunos trabajos de desambiguación semántica automática o en cualquier tarea de anotación de corpus que necesite información léxica.

Gaur egun, PCED estrategia nagusi bada ere, oraindik badira proiektuak non corpusaren papera hizkuntzalariaren introspektzioaz murrizten duten:

Por esta razón, este diccionario se reclama como perteneciente a un tipo de lingüística con corpus (frente a lingüística de corpus): «una lingüística en la que el corpus está al servicio del investigador, de forma que los datos encontrados se filtran por su introspección, se evalúan y se completan con otros que el corpus no proporciona, pero que la introspección considera naturales» (Bosque 2004: CLIX).

Baina, eztabaida hori gaintitze aldera, zenbait autorek proposatzen dute corpusa bera dela hiztegi edo etorkizuneko hiztegi, behintzat (Abaitua 2006). Alegia, corpusa ez litzateke besterik gabe hiztegia egiten laguntzeko tresna, baizik eta corpusak berak hiztegi gisa jokatu luke. Corpusen eta hiztegien arteko lotura hori indartu zuenariko aintzindari Sinclair dugu, horrela, *The Dictionary of the Future* delako hitzaldian, Sinclair-ek hiztegia corpuserako sarbide-interfaz gisa hartzen du: *a device through which the user will observe the living language* (Sinclair 1987).

Corpusak hiztegi gisa erabiltzeak hitzak erabilera bizian kontsultatzeko aukerak emango dizkio erabiltzaileari. Jakina, horretarako, ezinbestekoa izango da corpusak informazio linguistikorekin etiketatzea eta corpus horiek kontsultatzeko tresnak garatzea. Corpusak linguistikoki etiketatzeak

izugarritzko garrantzia du. Etiketatzeko-lanek aurrera egiten dute laguntza informatikoei esker, baina hala eta guztiz ere, eskuzko lan-zama oraindik ikaragarria da. Beraz, corpusak eguneratu eta kontsultagai jartzeak kostu handiak izango ditu eta, erabat eskuz egitea ezinezkotzat jotzen dute (Voutilainen et al 2012).

Bestalde, corpusekin batera, corpus horiek ustiatzeko eta kontsultatzeko tresnek berebiziko garrantzia dute. Lexikografoen lanari dagokionez, lan horretan LNP arloaren ekarpena bereziki, esan daiteke lehen lexikografoek eskuz egin beharreko zenbait lan konputagailuari ematea izan dela. Hala ere, hiztegiak egiteko moduak erabat irauli diren une honetan, badira zenbait galdera lexikografoen etorkizunaz ere:

Will mechanization lead to lexicographers becoming deskilled? is there still a role for informed human judgement? have I been colluding in a process that will lead to my own redundancy? and so on. (Rundell 2009)

Zer rol jokatuko duten lexikografoek ez dago argi, baina hori hurrengo lan baterako utziko dugu.

4. Zenbait ondorio

Lan honetan, *Euskal Hiztegiaren* gainean hizkuntzaren prozesamenduan ikuspegitik egindako zenbait lan laburbildu ditugu. Lan horiek hizkuntzaren prozesamenduan ez ezik, hiztegi gintzan ere izan zuten eraginik. Egindakoa laburbildu ondoren, gaitetik bada ere, corpusen eta hiztegien arteko interakzioari heldu diogu. Interakzio horiek hainbat ikuspuntutatik azter daitzeko, baina uste dut biak ala biak beharrezko izango ditugula etorkizunean ere. Beraz, kontua ez litzateke zein gailendu, baizik eta bata bestea elikatzea eta, elikatze horretatik biak aberastea.

Bukatzeko, esan beharra daukat aurkezturiko hurbilpenak modu sakonagoan beharko liratekeela aztertu. Azken eginkizun hau etorkizunerako lanen artean sartuko litzateke.

Aipamenak

- Abaitua, J. 2006. «Taxonomías y ontologías para la gestión de recursos lexicográficos». *Atti del Convegno Internazionale Glossari, dizionari, corpora: Lessicologia e lessicografia delle lingue europee*. Gargnano del Garda, Italia (2006ko maiatzaren 25-27), Università degli Studi di Milano.
- Agirre E., I. Aldezabal & E. Pociello 2006. «Euskararako ezagutza-base lexiko-semantikoaren eredu-hautaketa eta garapena: EuskalWordNet». *Gogoia*. 237-266.
- Agirre E., O. Ansa, X. Arregi, X. Artola, A. Díaz de Ilarraza, M. Lersundi, D. Martínez, K. Sarasola & R. Urizar. 2000. «Extraction of semantic relations from a

- Basque monolingual dictionary using Constraint Grammar». *Proceedings of Euralex*. Stuttgart.
- Agirre, E. & M. Lersundi. 2001. «Extracción de relaciones léxico-semánticas a partir de palabras derivadas usando patrones de definición». *Proceedings of SEPLN 2001*. Jaén.
- Alberdi, X. & J. Garcia de los Salmenes. 2008. «Definición lexicográfica y orden de la información de las palabras: el caso del euskera». In Elisenda Bernal & Janet DeCesaris (arg.), *Proceedings of the XIII EURALEX International Congress* (Barcelona, 2008ko utzailaren 15-19an)
- Alonso Ramos, M. 2009. «Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionarios?». In P. Cantos Gómez & A. Sánchez Pérez (arg.), *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus*. Murcia: AELINCO. 1191-1207.
- Arregi, X., J.M. Arriola, X. Artola, A. Díaz de Ilaraza, E. Garcia, V. Lascurain, A. Soroa & L. Uria. 2007. «Semiautomatic Construction of the Electronic Euskal Hiztegia Basque Dictionary (eEHBD)». *The XVIth biennial conference of the Dictionary Society of North America*. Chicago.
- Arregi, X., J. Arriola, X. Artola, A. Díaz de Ilaraza, E. García, B. Laskurain, K. Sarasola, A. Soroa & L. Uria. 2003. «Semiautomatic conversion of the *Euskal Hiztegia* Basque Dictionary to a queryable electronic form». *T.A.L. journal* 44-2. 107-124.
- Arriola, J.M., X. Artola, A. Maritxalar & A. Soroa. 1999. «A Methodology for the Analysis of Verb Usage Examples in a Context of Lexical Knowledge Acquisition from Dictionary Entries». *Proceedings of EACL'99*. Bergen.
- Arriola, J.M. 2000. *Euskal Hiztegiaren azterketa eta egituratzea ezagutza lexikalaren eskuratzeko automatikoari begira. Aditz-adibideen analisisa Murriztapen Gramatika baliatuz, azpikategorizazioaren bidean*. doktore-tesia, UPV/EHU.
- Artola, X. 2004. «Laying Lexical Foundations for NLP: the Case of Basque at the Ixa Research Group». *First Steps for Language Documentation of Minority Languages: Computational Linguistic Tools for Morphology, Lexicon and Corpus Compilation. 4th International SALT MIL (ISCA SIG) LREC*. 2004ko maiatzaren 24an. Lisboa.
- Azkarate, M. 2008. «Hiztegi gintza eta euskararen normalakuntza». *Euskalgintza XXI. mendeari buruz. XV. Biltzarra*. Iker-19. Euskaltzaindia. 171-182.
- Bosque, I. (dir.) 2004. *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- Euskaltzaindia. 2000. *Hiztegi Batua*. Bilbo: Euskera 5. [<http://www.euskaltzaindia.net/hiztegiabatua/>]
- Elhuyar. 1998. *Elhuyar Hiztegi Txikia*. Usurbil: Elhuyar Kultur Elkarte.
- Hanks, P. 2012. «The Corpus Revolution in Lexicography». *International Journal of Lexicography* 25 (4). 398-436.
- Hanks, P. 2009. «The Impact of Corpora on Dictionaries», In P. Baker (arg.), *Contemporary Corpus Linguistics* (Series: *Contemporary Studies in Linguistics*). London: Continuum. 214-236.
- Karlsson, F., A. Voutilainen, J. Heikkilä & A. Anttila. 1995. *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Berlin: Mouton de Gruyter.
- Kilgarriff, A. 2005. «Putting the Corpus into the Dictionary», *Proceedings MEANING Workshop*, Trento.

- Leturia, Igor. 2013. «Hiztegegintza Teknologiaz modernizatzen». *ELHUYAR al-dizkaria*. 2013/10/10 [http://zientzia.net/artikuluak/hiztegegintza-teknologiaz-modernizatzen/]
- Mitxelena, L. & I. Sarasola. 1989. *Orotariko Euskal Hiztegia* (II-IV). Bilbo: Euskaltzaindia.
- Morris, M. 1998. *Morris Hiztegia*. Klaudio Harluxet Fundazioa: Donostia. [http://www1.euskadi.net/morris/indice_e.htm.]
- Rundell, M. 2009. «The road to automated lexicography: First banish the drudgery... then the drudges?». *Proceedings of eLexicography in the 21st Century Conference*. Louvain-la-Neuve, Louvain: Université Catholique de Louvain.
- Rundell, M. 2002. «Good Old-fashioned Lexicography: Human Judgment and the Limits of Automation». In M.-H. Corréard (arg.), *Lexicography and Natural Language Processing. A Festschrift in Honour of BTS Atkins*. Euralex. 138-155.
- Salaburu, Pello. 2011. «Corpusak eta hiztegiak». *Sareko Euskal Gramatika (SEG)*. [http://www.ehu.es/seg/hizt/6]
- Sarasola, I. 1996. *Euskal Hiztegia*. Donostia: Kutxa Fundazioa.
- Sarasola, I. 2012. «Orotariko Euskal Hiztegia eta euskal hiztegegintza». *Koldo Mitxelena Katedraren III. Biltzarra*. Vitoria-Gasteiz, 2012/X/8-11. [https://www.youtube.com/watch?v=zx5GjrlOufY&noredirect=1]
- Sinclair J. 1987. *Collins COBUILD English Language Dictionary*. John Sinclair.
- Voutilainen, A., T.K. Purtonen & K. Muhonen. 2012. «Outsourcing Parsebanking: The FinnTreeBank Project». In Diana Sousa, Krister Lindén & Wanjiku Nganga (arg.), *Shall we Play the Festschrift Game?: Essays on the Occasion of Lauri Carlson's 60th Birthday*. Springer Verlag. 117-131.
- Walker D.E., A. Zampolli & N. Calzolari (arg.). 1995. *Automating the Lexicon: Research and Practice in a Multilingual Environment*. Oxford: Oxford University Press.
- Zabala, Igone. 2013a. «Corpusak eta hiztegiak». *31 eskutik*. 2013/03/01. [http://31eskutik.com/2013/03/01/corpusak-eta-hiztegiak/]
- Zabala, Igone. 2013b. «Corpusen adierazgarritasuna». *31 eskutik*. 2013/04/01. [http://31eskutik.com/2013/04/01/corpusen-adierazgarritasuna/]