

Hauta-Lanerako Euskal Hiztegiaren analisi erdiautomatikoa

J.M. ARRIOLA,* X. ARTOLA, A. SOROA
(E.H.U. / U.P.V.)

Abstract

The aim of this paper is to present the basic prerequisites which must be met to be able to carry out a computer-based analysis of the Basque dictionary Hauta-Lanerako Euskal Hiztegia (HLEH). Bearing in mind that HLEH is a rich source of information designed for practical use, the first step, previous to the automatic processing, has been the structuring of the text (of the dictionary). The main steps of this structuring process have been labeling, writing the grammar which is necessary for the analysis, and the analysis itself. We have adapted the results of our analysis to the directives of the TEI, so that it will be useful for both those who analyze Basque and those who simply use the language.

1. Sarrera gisa**

Lengoaia naturalaren prozesamenduko (LNP) aplikazioen problemarik larriena lexikoi konputazional "handia"¹ eraikitzean datza, hots, sistema errealetarako jakintza lexikala eskuratzea ezinbestekotzat jotzen da arlo honetan aurrera egin nahi bada. Horrela, LNPrako lexiko-sistemen eraikuntzarako laguntza automatikoak garatzea eta dauden baliabide lexikalez baliatzea dira harturiko irtenbide nagusienak. Era berean, lexikografia konputazionala, lexiko-sistemen azterketa, errepresentazioa eta erabilpena, gero eta garrantzi handiagoa hartzen ari dira.

Gure lana hainbat alderdi desberdin bildu ohi dituen lexikografia konputazionalaren alorrean koka daiteke. Euskarri magnetikoan dauden hiztegiak (Machine Readable Dictionaries edo MRDak) erabili izan dira oinarritzko informazio-iturri gisa LNPrako lexikoiak hornitzeko zeregin horretan (Boguraev & Briscoe 1989; Castellón 1993) eta bide horretatik jo nahi dugu guk ere, hiztegiaren edukia aztertuz, maila lexikaleko informazio semantikoa eta sintaktikoa erauzte aldera. *Hauta-*

*Eusko Jaurlaritzako Bekaduna

** Proiektu honek Gipuzkoako Foru Aldundiaren laguntza du.

(1) Literaturan hainbat aldiz aipaturiko "lengoaia naturalaren arloko zintzurgune estua", hots, neurri errealeko lexikoi konputazional "erraldioiak" erabilera desberdinetarako prestatzea.

Lanerako Euskal Hiztegia euskarri magnetikoan izateak atek zabaldu dizkigu lexicografia konputazionalaren eremuan oparo jorratu den bide honi ekiteko.

Informatikaren aroak aldaketa sakona ekarri du lexicografiaren mundura. Informatika dela medio, astindua ere ezagutu du hiztegegintzak azken urteotan: ederki ahantzia dago paperezko fitxekin lan egiten zeneko sasoi. Gaur, ordenadorea da lexicografiaren lanabes arruntena. Eta, hori horrela, ordenadorez irakurgarriak diren hiztegiak ("Machine Readable Dictionary") sortu eta lantzen dira gehienbat.

Ordenadorez irakurgarriak izateak ez du esan nahi ordea, hiztegi horiek guztiak, informazioaren egituraketari dagokionean, maila berekoak direnik. Ez da inola ere gauza bera, esate baterako, testu-prozesadore batez editatutako hiztegia, testu huts dena, edo datu-base baten gisara osatutakoa. Eta hiztegi datu-baseen artean ere bada, noski, alderik.

Garbi dago, hori bai, datu-baseak testua baino oinarri askoz ere egokiagoa direna langintza informatikorako eta, hala berean, hiztegien bertsio elektronikoak sortzeko. Hori kontuan, azken urteotan euskararen lexikoari buruzko informazioa Euskararen Datu-Base Lexikalean (Agirre et al. 1994) bildu dugu, egun 60.000 sarreratik gora dituen datu-biltegi handian. Datu-base horretan euskararen ezaugarri morfologiko eta sintaktikoak jaso dira, baina ezaugarri semantikoak ez ditugu sartu gaurdaino. Kontuan izanik Euskararen Datu-Base Lexikala² dela eskuartean ditugun aplikazio desberdinetarako sostengu nagusia, lan honen helburuetariko bat EDBLren aberasketa dugu. Jakintza lexikal ahalik eta osoena behar da, informazio anitzez osatu nahi baita EDBLko sarrera oro (informazio sintaktiko, semantikoa, etab.). Honekin batera, helburu nagusienak, hauexek lirateke:

- ondoren egingo diren azterketa linguistikoak egin ahal izateko "iturri lexikal" honen oinarritzko prestaketa.
- analisia bera, hots, analisiaren emaitza.

2. HLEHren ezaugarri nagusiak

HLEH hiztegia euskara batuaren biltegi gaurkotua da, hastapenetan arau-emaile izateko bokazioa zuena baina gerora hauta-lanetan erabiltzeko moldatu dena. Euskarazko hiztegi elebakarra da, orokorra, taxuz egindakoa, tradizio literarioan oinarritua, 30.000 sarrera inguru dituena eta gaur egun lexikoa zertan den jakiteko erreferentzia nagusia.

Hiztegia artikuluz osaturiko bilduma dugu, unitate nagusi hau sarrera-buruak eta honi dagokion informazioak osatuko dute.

Hona hemen segidan hiztegiko artikuluen osakera:

- Sarrera: letra lodiz bereizten dira. Eratorriek ere badute beren sarrera.
- Kategoría: eremu honetan kategoría gramatikalak nahiz ezaugarri morfologikoak ager daitezke. Sarreraren ondoren dator normalean kategoría.

(2) EDBLk, lexikoari buruz informazio-biltegi erraldoia den neurrian, euskararen prozesamendu automatikoan zeregin desberdin askotarako oinarri lexikala izan behar du, hau da, arloko lan desberdinetan beharrezko diren lexikoen euskarri eta iturri.

- Data: sarreraren agerpen-data hiztegi edo testu literario batean.
- Definizioa: adiera bakarreko sarreretan definizio-testuak osatzen du eremu hau. Adiera bat baino gehiago dagoenean bereizgarri desberdinak erabili dira definizioak banatzeko, edota ñabardurak markatzeko.
- Erlazioak: eremu honetan sinonimia edo antonimia bezalako adiera-erlazioak adierazteko kode lexikografiko ohikoak erabiltzen dira: Ik., Ant., ...
- Adibidea: testu etzanez adierazten da adibidea, normalean definizioaren ondoren.
- Laburdurak: erabiltzerakoak, tematikoak, hizkuntz mailak, euskalki-argibideak, etab.
- Azpisarrerak: maiuskulaz ageri dira eta sarreratzat har daitezke, baina beste sarrera baten baitakoak.
- Azalpen gramatikalak: eremu honetan sarrerari buruzko informazio gramatikalak aurki daitezke.

HLEH irakurgarria da ordenadoren (badago MRD bertsioa): testu fitxategi baten gisara, hiztegi artikuluak sekuentzialki editatu dira eta informazio motak banatzeko bereizgarri tipografikoak erabili dira. Hona hemen MRDko sarrera bat:

abalketu, abalke, abalketzen. da-du *ad.* (1571). *Ipar. edo jas. A. da ad.* Lotsatu. *Abalke zaitez, zerbitzari laxoa. Erori orduko abalketu ziren, eta estali zituzten beren gorputzak piko-hostoz.* □ *Zerbaitz edo norbaitz abalketu. Abalketzen baita erokeria haiek ikusteaz eta entzuteaz.*

B. du ad. Ik. abalkerazi. 1. (1571). Lotsatu. *Neure bihotzeko hasperenek abalketzen ninduten. 2.* (XIX b.). Beldur erazi. *Mamu horrek txoria abalketzen ditu.*

1. Irudia. *Abalketu* sarrera MRDn azaltzen den bezala.

Horrelaxe bada, HLEH informazio-iturri joria da, baina informazio hori era gordinan dago eta, giza erabiltzaileari zuzendua dagoenez, inprimatutako bertsioen akatsak ditu, bereziki, informazioaren egituraketa desegokia LNPren ikuspegitik. Beraz, premiazkoa da edozein prozesamendu automatikori heldu aurretik hiztegi-testua egituratzea. Ondoko puntuetan zehazten dira egituratze-lanetan jarraituriko urratsak.

3. Etiketatzea: unitateetan segmentatzea, unitate bakoitzari etiketa bat esleituz (Tokenizazioa)³

Iturburu-testua —hiztegia— segmentatzea da ataza honen helburua. Etiketatzean artikuluak identifikatzeko behar diren item-ak eta eremuak identifikatzen dira. Badira osagai horien artean informazio lexikografikoari buruzko kodeak eta estiloari edo letra motari dagozkion kode tipografikoak. Gainontzeko testu-atalak TX, hots,

(3) Tokenizazioa PERL programeriaz baliatuz gauzatu da, prozesu honen emaitza testu soila (ASCII kodez) eta guk definitu ditugun etiketekin geratuko delarik.

testu bezala kodeturik azalduko dira. MRDko sarreraren ikus daitekeenez (ikus 1. Irudia), nahiz eta sarreraren zatiak esplizituki adierazita ez egon, lagungarri izan daitezkeen kode tipografiko eta lexikografikoak egon badaude, eta hauetaz baliatuko gara gure lanari aurre egiterakoan. Kode horiek guztiak (ikus 2. Irudia) lagungarriak dira giza-erabiltzaileak hiztegia kontsultatzerakoan. Baina ordenadoreak uler ditzan azalarazi egin behar dira eta horretarako 3. irudian azalduko ditugun tokenak sortu ditugu.

	Kode tipografikoak
	Lodia kodea
	Etzana kodea
	Azpimarra kodea
	Kode lexikografikoak
Adiera Multzoa	"A.", "B.", "C."
Adiera Zenbakiak	1.", "2." ...
Adiera Xeheak	"a)", "b)", "#1", "#2"
Erreferentziak	"Ik."
Antonimoa	"Ant."

2. Irudia. Kode tipografiko eta lexikografikoak.

Tokena	Deskripzioa	Oharrak
[SH] / [SB]	Hiztegiko sarrera baten hasiera /bukaera	
[LH] / [LB]	Lodi hasiera/bukaera	Kode tipografikoa. Word fitxategian lodiz markatutako zati guztiak.
[EH] / [EB]	Etzana hasiera/bukaera	Kode tipografikoa. Word fitxategian etzanaz markatutako zati guztiak.
[ZH] / [ZB]	Zenbaki hasiera/bukaera	Bakarrik lodiz dauden zenbakiak markatu dira.
[AMH] / [AMB]	Adiera multzo hasiera/bukaera	Kode lexikografikoa.
[AZH] / [AZB]	Adiera zenbaki hasiera/bukaera	Kode lexikografikoa.
[HZH] / [HZB]	Homografo zenbaki hasiera/bukaera	Kode lexikografikoa.
[IKH] / [IKB]	Erlazioaren hasiera/bukaera	Kode lexikografikoa.
[ANH] / [ANB]	Antonimoaren hasiera/bukaera	Kode lexikografikoa.
[AXH] / [AXB]	Adiera xehearen hasiera/bukaera	Kode lexikografikoa.

[DH] / [DB]	Dataren hasiera/bukaera
[LAH] / [LAB]	Laburduraren hasiera/bukaera
[KAH] / [KAB]	Kategoria hasiera/bukaera
[MH] / [MB]	Maiuskula hasiera/bukaera

3. Irudia. Erabilitako *tokenak*

Ikus daitekeen bezala, kode tipografiko eta lexikografikoez gain baditugu bestelako batzuk ere, hala nola: *data hasiera / bukaera*, *laburduraren hasiera / bukaera*, *kategoria-ren hasiera / bukaera* eta *maiuskula hasiera / bukaera*. Etiketa hauek sortzearen arrazoia analisi xehea egin nahi izatea da. Jakina, zenbat eta etiketatze finagoa orduan eta informazio xeheagoa jaso ahal izango dugu. Hona hemen MRDko *abalketu* sarrera etiketaturik:

[SH][LH]ahalketu, ahalk, ahalketzen. [LB]da-du [KAH]ad. [KAB]-[DH](1571)[DB][LAH]. Ipar. edo jas. [LAB][LH][AMH]A.[AMB][LB] da [KAH]ad. [KAB]Lotsatu. [EH]Ahalk zaitetz, zerbitzari laxoa. Erori orduko ahalketu ziren, eta estali zituzten beren gorputzak piko-hostoz. [EB][AZH]#1[AZB][EH] Zerbaitez edo norbaitez ahalketu. Ahalketzen baita erokeria haiek ikusteaz eta entzuteaz. [EB][LH][AMH]B.-[AMB][LB] du [KAH]ad. [KAB][IKH]Ik. ahalkerazi. [IKB][LH]1. [LB][DH](1571).[DB] Lotsatu. [EH]Neure bihotzeko hasperenek ahalketzen ninduten. [EB][LH]2. [LB][DH](XIX b.).[DB] Beldur erazi. [EH]Mamu horrek txoriak ahalketzen ditu.[EB][SB]

4. Irudia. *Abalketu* sarrera etiketatzea burutu ondorik.

4. Analisirako gramatikaren idazketa

Gramatikak⁴ artikuluen egitura orokorra jasoko du eta horretarako DCGa (Definite Clause Grammar) garatu izan da. Gramatika honen oinarrizko zatiak, noski, tokenizazio fasetik ateratakoak dira, hau da, *tokenak* eta *tokenen* arteko testuak. Parte txiki hauetatik abiatuta eta posizioaren arabera, analizatzailea sarrerren egitura asmatzen saiatzen da poliki-poliki.

Metallengoia batez baliatuko gara sarrerren egitura orokorra deskribatzeko:

- <A>/ A edo B, aukerako sinboloak.
- <A> A eta B, Aren jarraian B.
- [<A>] aukerako sinboloa.
- <*NULL> sinbolo hutsa.
- <*A> bukaerako sinboloa.

Erregelen sintaxia:

<ELEMENTUA> = <EL1> <EL2>... <ELN>.

(4) Gramatika Prolog programazio lengoiaiaz idatzia izan da, Definite Clause Grammar (DCG) gramatika-mota erabili izan delarik.

Hona hemen sarreren egitura orokorra jasotzen duen gramatika:

```

<SARRERA> = <LEMA> [<ERLAZIOAK>] <KATEGORIA>
[< *DATA>] [<DEFINIZIO_ADIBIDEAK>].
<LEMA> = [<HOMOGRAFO_ZENBAKIA>]
[<LEMA_HILA> / <LEMA_ESTANDARRA>].
<HOMOGRAFO_ZENBAKIA> = <*HH> <*ZENBAKIA> <*HB>.
<LEMA_HILA> = <*GUR> <*LH> <*LEMA> <*LB>.
<LEMA_ESTANDARRA> = <*LH> <*LEMA> <*LB>.
<KATEGORIAK> = [<*AZPIKATEGORIA>] <KATEGORIA>.
<KATEGORIA> = <*EH> <*KAT> <*EB>.
<DEFINIZIO_ADIBIDEAK> = <DEFINIZIOA> [<ADIBIDEAK>]
<DEFINIZIO_ADIBIDEAK> / <*NULL>.
<DEFINIZIOA> = [<ADIERA_ZENBAKIA>][<ADIERA_MULTZOA>] <*DEFINI-
ZIOA> [<ERLAZIOAK>].
<ADIERA_ZENBAKIA> = <*LH> <*ZENBAKIA> <*LB>.
<ADIERA_MULTZOA> = <*AMH> <*ADIERA_MULTZOA> <*AMB>.
<ERLAZIOAK> = [<SINONIMOA> / <ANTONIMOA>]
<ERLAZIOAK> [<ADIBIDEAK>]. / <*NULL>.
<SINONIMOA> = <*IKH> <*SINONIMOA> <*IKB>.
<ANTONIMOA> = <*ANH> <*ANTONIMOA> <*ANB>.
<ADIBIDEAK> = <*EH> <*ADIBIDEA> <*EB>.

```

5. Analisia

Analisi-prozesuaren emaitza hiztegiko testua izango da, baina aurredefinitutako egitura baten arabera antolatuta. Emaitza hiztegi-informazio egituratua izango dugu, artikulua, hauetako atalak eta atal bakoitzeko eremu zein azpi-eremu bereizten direlarik: definitutako hitza, kategoria, definizioa, adibidea, etab. Egitura honetan kode tipografikoak ez zaizkigu azaltzen, hemendik aurrera eman behar diren urratsetarako ez baitute interesik. Hala eta guztiz ere, artikulua jatorrizko formatoa berreraiki nahi balitz, ez litzateke inongo arazorik izango kode tipografiko horien berreskurapenerako.⁵ Eskuarlean dugun HLEHren laginak 22.767 artikulua ditu, 19.290 artikulua analizatu ditugularik, eta horietarik %90 ondo analizatua dago. Emaitzak epaitzerakoan bi puntu izan beharko liriateke kontuan: laginaren ezaugarriak⁶ eta ale xeheko analisia egin dugula.

Badugu, orain arte egindako lanaren ondorioz, *Hauta-Lanerako Euskal Hiztegia* euskarri magnetikoan, non bere sarreren sakoneko egitura ezagututa dagoen. Konputagailuz trata daitekeenez, hiztegi honek euskararako baliabide lexikal paregabea eskaintzen digu. Hala ere, baliabide lexikal anitzi gertatzen zaion bezala, gureak ere arazo nagusi bat dauka, hots, baliabide honetan oinarritzen den edozein tratamenduk gure hiztegiko errepresentazioari lotuta egon behar baitu. Arazo hau ez da berria eta azken hamarkadan komunitate zientifikoak hainbat saio egin du sarrera

(5) Jatorrizko bertsioa errespetatzea izan baita beti ere gidatu gaituen printzipio nagusia.

(6) Hamar urteetan egindako lan lexikografikoa izateak izan du eraginik, ezen aletik alera aldaketa franko aurkitu izan baititugu, analisiaren prozesua zailagoa eginez.

lexikalen errepresentazio orokor bat lortzeko. Hau dela eta, HLEHren azterketaren emaitza (Text Encoding Initiative, TEI) delakoak testu elektronikoak kodetzeko eta trukatzeko bere gidalerroak —TEI P3—, (Sperberg-McQueen & Burnard 1994)n proposatzen diren etiketen bidez adieraztea erabaki dugu.

Egituratzeko lanaren emaitza gisa hiztegiaren analisisa erdietsi dugu, lorturiko emaitza egituratu honetatik formatu estandarera pasatzea erabaki da, hots, TEIn proposatzen diren etiketen bidez adieraziko dugu analisiaren irteera. Adibidez, 5. irudian *ahalketu* sarrera ikus daiteke TEI formatuan jarririk.

```

<entry>
  <form><orth>ahalketu, ahalke, ahalketzen.</orth></form>
  <GramGrp><subc>da-du </subc><pos>ad. </pos></GramGrp>
  <usg type=time>.(1571)</usg>
  <usg type=geo>. Ipar. edo jas. </usg>
  <sense n = 'A.'>
    <GramGrp><subc> da </subc><pos>ad.</pos></GramGrp>
    <def>Lotsatu. </def>
    <eg>Ahalke zaitetz, zerbitzari laxoa. Erori orduko ahalketu ziren, eta estali zituzten beren gorputzak piko-hostoz. </eg>
    <sense n = '#1'>
      <eg> Zerbaitez edo norbaitez ahalketu. Ahalketzen baita erokeria haiek ikusteaz eta entzuteaz.</eg>
    </sense>
  </sense>
  <sense n = 'B.'>
    <GramGrp><subc> du </subc><pos>ad.</pos></GramGrp>
    <xr type = syn>Ik. ahalkerazi.</xr>
    <sense n = '1. '>
      <usg type=time>.(1571) </usg>
      <def> Lotsatu. </def>
      <eg>Neure bihotzeko hasperenek ahalketzen ninduten.</eg>
    </sense>
    <sense n = '2. '>
      <usg type=time>(XIX b) </usg>
      <def> Beldur erazi. </def>
      <eg>Mamu horrek txoriak ahalketzen ditu.</eg>
    </sense>
  </sense>
</entry>

```

5. Irudia. *Ahalketu* sarrera TEI formatuan.

TEI aukeratu dugu HLEHren MRDtik abiatuta lortutako baliabide lexikala adierazteko erredutzat, besteak beste arrazoi hauengatik:

- Lexikografoen lanerako onuragarria delako, informazioaren egituraketari dagoikienean kodaketa-sistema finkoa eskaintzen baitu. Ez da inola ere gauza bera,

esate baterako, testu-prozesadore batez editatuko hiztegia, testu huts dena, edo honelako adierazpide aberatsago bat erabiliz osatutakoa.

- Honen inguruan garatzen ari diren tresnak kontuan harturik, informazioaren atzipena, kontsulta-aukeren ugaltasuna eta aberastasuna, etab. ahalbideratu daitezkeelako.
- Estandarra denez, informazioaren berrerabilgarritasuna eta elkartrukatzea gauzatu baitaitezke.

6. Akatsen tratamendua

Automatikoki eginiko analisiaren ondorik, ezagutu gabe geratu diren artikuluen azterketa egin da ikusteko zein ziren ez ezagutzearen arrazoiak. Horrela bada, akatsak sailkatu ditugu aurrera begira hauen tratamendu erdiautomatikoa bideratzeko:

- 1) Errore tipografikoak edota sakatze akatsak, hau da, Word testu-prozesadorean aritzerakoan lexikografoak eginikoak, askotan hiztegiaren zenbait atal ez dagokienez letra-motaz azaltzean ezin izango ditugu analizatu. Adibidez, sarrearen adibidea dena definizio eremuan txertatzea.
- 2) Espero ez ziren elementuen agerpena. Esate baterako, hiztegiaren sarreran ematen diren laburduren zerrendatik at dauden laburdurak ager daitezke.
- 3) Egileek zehazten dituzten zenbait kode lexikografiko, beraiek adieraziriko eginkizunez gain beste erabilera batzuetan ere aurki daitezke.

Bestalde, zenbait sarrerak duen egituraren konplexutasunagatik gure gramatikak ezin izan ditu ezagutu. Arrakasta kopurua %100 edukitzeko modu bakarria eskuzko prozesu baten menpe dago: sarrera akastunak eskuz analizatu eta gaizki analizatu diren sarrerak zuzendu.

HLEH oso osorik datu-base lexikal batean gorde nahi bada, akats hauei aurre egiteko ingurune egoki bat prestatu behar da lexikografoaren lana arintzeko. Beraz, analisi automatikoaren bidez ezin izan dena ezagutu, erdiautomatikoki landu beharko da.

7. Ondorioak

Azterketa honek ematen du aukera hiztegia, modu honetara egiteak dakartzan gabeziak eta akatsak nabarmenarazteko. Informatikaren aroak aldaketa sakona ekarri du lexikografiaren mundura. Lexikografoen lanerako euskarri ezinbestekoa da, egungo datu-base lexikala: artikuluen idazketarako, kontsistentzia bermatzeko, bertsio desberdinak lantzeko, eta, normalizazio bidean dagoen hizkuntza baten kasuan, baita gertatzen diren aldaketa edo hartzen diren erabakien berri eguneratua gordetzeko ere.

Bestalde, testu huts izatetik, formatu eta adierazpide aberatsago batera pasatzeko oinarriko urratsa ere bada. Adierazpide berrituak prozesamendu informatikorako atea zabaltzen ditu, bai informazioa bildu eta gordetzeko —datu-base erlazionalen moldean adibidez—, bai informazioa erabiltzeko, hitz eta adieren arteko erlazioak ustiatuz. Hiztegi elektronikoei eskatzen zaizkien ezaugarriak, informazioa atzitzeko arintasuna, etab., lortu ezinezkoak dira ez bada alde aurretik hiztegien azterketa sakonik egiten. Paperezko formatuan argitaratu ondoren, hiztegi horren bertsio elektronikoa sortzea jomuga interesgarria iruditzen zaigu.

Iritzi hori zenbait arrazoi desberdinek justifikatzen du:

- Helburu lorgarria: gaur egun maneiatzen ditugun teknika eta metodoek ego-kiera ematen dute HLEH bezalako hiztegia paperezko formatutik bertsio elektronikora pasatzeko.
- Erabilgarria: zinez uste dugu euskarazko hiztegien erabiltzaileek eskertuko dutela, eta ohitura hartu ahala gehiago eskertu ere, hiztegiak bertsio elektronikoetan ere argitaratzea. Gure iritziz, badago nahikoa arrazoi hiztegi elektronikoen erabilera paperezkoena baino aberatsago gertatuko dela uste izateko. Onura horiez ohartu eta hiztegi elektronikoak erabiltzen ikastea denbora eta ohitura da batez ere. Esan behar da, gainera, badirela dagoeneko euskarazko hiztegiak bertsio elektronikoan merkaturatu direnak.
- Euskararen modernizatorako aurrerapausoa: teknologia berrien aroan euskarak ez luke atzera geratu behar eta, alde horretatik begiratuta ere, oso mesedegarri litzateke lan hau.
- Aurreko puntuarekin lotuz, hiztegi elektronikoak oinarrizko baliabidearen rola jokatu luke beste hainbat aplikazio automatikorako.

Bibliografia

- Agirre E., Arregi X., Arriola J. M., Artola X., Díaz de Ilarraza A., Insausti J. M., Sarasola K., 1995, "Different issues in the design of a general-purpose Lexical Database for Basque", in *Proceedings of the First Workshop on Applications of Natural Language to Databases. (NLDB 95)*. Versailles, 299-313.
- Agirre, E., Arregi X., Arriola J. M., Artola X., Insausti J., *Euskararen Datu-Base Lexikala (EDBL)*. UPV/EHU / LSI / TR8-94.
- Amstler, R., 1981, "A Taxonomy for english Nouns and Verbs", in *Proceedings of the 19th annual Meeting of the Association for Computational Linguistics, (ACL'81)*, Stanford, California, 133-138.
- Boguraev, B. and Briscoe, T. (arg.), 1989, *Computational Lexicography for Natural Language Processing*. New York: Logman.
- Boguraev, B. et al., 1991, *Database models for Computational lexicography*. Research Report RC 17120, IBM Research Center, Yorktown Heights, NY.
- Byrd, R. J. et al., 1987, "Tools and Methods for Computational Lexicography", *Computational Linguistics* 13, ns. 3-4.
- Calzolari, N. and Picchi, E., 1986, "A project for Bilingual Lexical Database System", in *Proceedings of the Second annual Conf. of the Centre for the New OED*, University of Waterloo, Waterloo, Ontario, 79-82.
- Castellón, Irene, 1993, *Lexicografía Computacional: Adquisición automática de conocimiento léxico*. Tesis Doctoral, Universidad de Barcelona.
- Martí, M. A. and Castellón, I., 1990, "Gramática para el análisis del diccionario Vox", *Boletín SEPLN* 10, Donostia, 123-143.
- Sarasola, I., 1984-95, *Hauta-Lanerako Euskal Hiztegia*. G. K. Donostia.
- Sperberg-McQueen, C. M., Burnard, L., (arg.), 1994, *Guidelines for Electronic Text Encoding and Interchange*. Chicago & Oxford.