

Hiztegietan oinarritutako hitzen adiera-desanbiguazioa euskaraz

E. Agirre, O. Ansa, M. Aranzabe, X. Arregi, J.M. Arriola, X. Artola, A. Díaz de
Ilarraza, M. Lersundi, D. Martínez, K. Sarasola, A. Soroa

UPV/EHU/LSI/TR 10-2000
2000ko Uztaila

Abstract

In this paper we focus on the Word Sense Disambiguation (WSD) task, using a Monolingual Dictionary of Basque in MRD (Machine Readable Dictionary) format as main linguistic resource (*Euskal Hiztegia*, Sarasola 1996). As a result of this work, we have automatically disambiguated the senses of the semantic relations extracted from the definitions of the MRD in a previous work (Agirre et al., 2000). The WSD method is based in (Rigau et al., 1997) and has obtained promising accuracy: 80% with polisemous nouns, 66% with polisemous verbs and 84% with polisemous adjectives. We also have created a semantic network of concepts from the disambiguated relations. This network represents one first approach to create an ontology of Basque concepts, which is a very important linguistic resource for Natural Language Processing (NLP) development.

Aurkibidea

1. Sarrera	4
2. Desanbiguazio algoritmoak	5
2.1. Hiztegiko Lehen Adiera (HLA)	5
2.2. Teilakapena	5
2.3. Agerkide-tza-sinplea	5
2.4. Agerkide-tza-bektorea	5
2.5. Yarowsky (1992).....	5
2.6. Dentsitate/Distantzia kontzeptuala.....	6
2.7. Karov and Edelman (1998)	6
3. Algoritmoen parametrizazioa	7
4. Erabilitako baliabide lexikoak eta tresna linguistikoak.....	8
5. Ebaluazioa	8
6. Hierarkiaren eraikuntza automatikoa	11
7. Aurreko lanekiko konparaketa	12
8. Ondorioak eta etorkizunerako lana.....	12
9. Bibliografia.....	14
A. Eranskina: Laginetan aztertu ez diren kasuak	16
B. Eranskina: Algoritmoen parametroen doiketa	22
C. Eranskina: Izenen sinonimo multzoak.....	24

1. Sarrera

Txosten honetan hitzen adiera-desanbiguaziorako metodo bat (Rigau et al., 1997) aplikatzen zaio *Euskal Hiztegiari* (Sarasola, 1996). Hitzen adiera-desanbiguazioa arreta handia jasotzen ari den ikerkuntza arlo bat da gaur egun. Arazo hau lantzea beharrezkoa da LNPko (Lengoaia Naturalaren Prozesamenduko) zenbait aplikazio garatzeko: itzulpen automatikoa, lengoaiaren ulermena, ontologiaren erauzketa hiztegietatik abiatuz... (Ide and Veronis, 1998)

Erabili den metodoak ezagutza-iturri desberdinak (hiztegiak, ontologiak eta corpusak) erabiltzen dituzten zenbait algoritmo konbinatzen ditu. Teknikaren aplikazio-eremua *Euskal Hiztegia* (Sarasola, 1996) izan da. Hiztegi horren definizioetatik erauzitako erlazio semantikoak (Agirre et al., 2000a) hartu ditugu abiapuntutzat, eta hauen gainean desanbiguazio semantikoa aplikatu dugu. Ondoren desanbiguaturiko erlazio horietan oinarrituriko hierarkia baten lehen hurbilpena eraiki da automatikoki.

Metodo honen bidez aztertu diren erlazio semantikoak sinonimia eta hiperonimia izan dira, eta hauek erabili dira hierarkia eraikitzeko. Hiztegietakoa definizioetan askotan "genus et differentia" egitura erabiltzen da; hau da, hitzaren hiperonimoa ematen da eta hiperonimo bera duten hitzetatik desberdintzeko ezaugarriak. Hiztegietatik erauz daitekeen erlazio garrantzitsuenetako bat hitzen adieren arteko hiponimo/hiperonimo erlazioa da (Amsler, 1981; Vossen and Serail, 1990). Ontologiaren bizkarrezurra osatzen du erlazio honek, adibidez *WordNets* (Fellbaum, 1998).

Hitzen adiera-desanbiguazioa burutzeko, desanbiguatu nahi den hitza duen definizioa hartzen da lehenik. Ondoren, hitz horrek *Euskal Hiztegian* dituen adieretatik testuinguru horretan egokiena zein den asmatzen saiatzen da algoritmoa. Horretarako euskararen analisirako tresna linguistikoak (lematizatzailea, kategoria etiketatzailea; Morfeus, (Urkia, 1997)) eta baliabide lexiko lagungarriak (*WordNet*, (Fellbaum, 1998); *Morris* hiztegi elebiduna, (Morris, 1999)) erabili dira. Prozesu honen helburua hitz-adieren arteko erlazio semantikoak identifikatzea da eta izen, adjektibo eta aditzen arteko erlazioei aplikatzen zaie.

Lan honetan ingelesarekin emaitza onak dituzten desanbiguazio-algoritmoak aztertu eta euskarara egokitu dira. Gaur egun metodo "gainbegiratuak" dira hitzen adiera-desanbiguazioan emaitza onenak lortzen ari direnak. Algoritmo hauek adiera bakoitzerako etiketaturiko adibideak erabiltzen dituzte adibide berriak desanbiguatzeko. Hauen arazo nabari bat etiketaturiko adibideak lortzearena da. Ingeleserako eskuz etiketaturiko hainbat corpus erabilgarri dago (*Semcor*, (Miller et al., 1993); *DSO*, (Ng and Lee, 1996)), eta oso erabilia izan da algoritmo gainbegiratuak lantzeko (nahiz eta bere garraia garritasuna eztabaidatua izan den; (Ng, 1999, Agirre et al. 2000b)). Adibide etiketatuak lortzeko metodo automatikoak ere erabiltzen dira. Hauek batez ere Internet (Mihalcea and Moldovan, 1999) eta ontologiak (Leacock et al., 1998) erabiltzen dituzte iturri bezala. Euskararako ez dago eskuz etiketaturiko corpora, eta adibideak automatikoki lortzeko metodoak ez dira aztertu. Ez dago argi gaur egun Interneteko informazioa euskararako adibideak lortzeko erabil daitekeen, ingelesarekin alderatuz oso gutxi agertzen baita euskara sarean. Gainera, ez dago LNPrako prestatua dagoen euskarazko Ontologia zabalik (lan hau bide horretan joango litzateke).

Nahiz eta, normala denez, ingelesarekiko desabantailan egon, hitzen adiera-desanbiguazioa LNPko aplikazio askotarako beharrezkoa den ataza bat da, eta nahitaez landu beharrezkoa. Euskararako dauden baliabide linguistikoek, gainera, algoritmo mota desberdinak lantzeko aukera ematen digute emaitza onekin, lan honek frogatzen duen bezala.

Aipatu diren murriztapenak direla eta, erabili diren algoritmo guztiak "ez-gainbegiratuak" dira; hau da, ez dituzte erabiltzen aurretik eskuz desanbiguaturiko adibide zerrendak. Aztertu diren algoritmoak bi multzotan sailka daitezke: hiztegiak eta corpusak erabiltzen dituztenak alde batetik, eta ontologiak erabiltzen dituztenak bestetik. Euskarazko kontzeptuen ontologia zabal erabilgarririk ez dagoenez, ontologietan oinarrituriko metodoetan ingelese-euskara itzulpena egin da hiztegi elebidunen bidez, eta ondoren ingelesezko ontologiak erabili dira.

Guztira desanbiguaziorako zazpi metodo aztertu dira. Lehenik bakarka probatu dira eta ondoren konbinaketa burutu da. 2. atalean metodo hauen ezaugarriak azaltzen dira. Metodo batzuentzat parametro desberdinak aztertu dira, eta hauek 3. atalean azaltzen dira. 4. atala erabilitako baliabide lexiko eta tresna linguistikoei eskainia dago. 5. atalean desanbiguazio metodoaren ebaluazioa egiten da. 6. atalean hierarkiaren eraikuntza aztertzen da eta ondorengo bi puntuetan aurreko lanekin konparaketa egin eta ondorioak ematen dira.

2. Desanbiguazio algoritmoak

Aurretik aipatu den bezala, literaturan agertzen diren zenbait algoritmoren emaitzak konbinatu ditugu. Konbinatzeko, algoritmo bakoitzak pisu normalizatu bat esleitzen dio adiera posible bakoitzari, eta ondoren guztien batura egiten da. Aztertu diren metodoak sinpleenetatik (1, 2, 3 eta 4) konplexuenera (5, 6, eta 7) doaz. Informazio iturriei dagokienez, alde batetik MRDa (*Machine Readable Dictionary*-a) soilik erabiltzen dutenak daude (hiztegi moduan edota corpus moduan: 1, 2, 3, 4 eta 7 algoritmoak) eta bestetik ontologiak erabiltzen dituztenak, *WordNet* gure kasuan (5 eta 6 algoritmoak).

2.1. Hiztegiko Lehen Adiera (HLA)

Algoritmo sinpleena da hau. Adierak hiztegian duten ordenaren arabera sailkatzen dira. Heuristikoak suposatzen du adiera erabilienak agertzen direla lehenik, eta hauei ematen die pisu maximoa.

2.2. Teilakapena

Testuingurua eta adiera bakoitzaren hiztegi-sarrerako definizioaren artean dauden hitz berdinak kontatzen dira. Pisu maximoa ematen zaio testuinguruarekin hitz berdin gehien duen adierari eta gainontzekoak honen arabera normalizatzen dira.

2.3. Agerkidetza-sinplea

Hau aplikatzeko lehenik *Euskal Hiztegi*ko definizio eta adibideak erabiliz hitz bikoteentzat agerkidetza pisuak lortzen dira, *association ratio* eta *mutual information* neurriak erabiliz.

Desanbiguatzean, adiera bakoitzeko, bere hiztegi sarreraren definizioko hitzek testuinguruko hitzekin dituzten agerkidetza pisuak batzen dira. Pisu gehiena duen adierari balio maximoa ematen zaio eta gainontzekoak honen arabera normalizatzen dira.

2.4. Agerkidetza-bektorea

Metodo hau (Wilks et al., 1993) artikuluan aurkezturikoan oinarritzen da eta honek ere hiztegi osotik erauzitako agerkidetza-informazioa erabiltzen du. Kasu honetan testuinguruarentzat eta adiera desberdinentzat bektoreak eraikitzen dira. Bektoreetan jatorrizko perpausetan dauden hitzak (1 pisuarekin) eta hauen hitz agerkideak (dagozkien agerkidetza pisuekin) sartzen dira.

Desanbiguatzeko testuinguru eta adiera bakoitzeko bektoreen arteko distantzia kalkulatu da (kosinu funtzioa erabiliz) eta hurbilenari pisu maximoa ematen zaio.

2.5. Yarowsky (1992)

(Yarowsky, 1992) metodoan oinarritzen da algoritmo hau. Lan honetan Yarowsky-k *Roget* thesaurusa erabili zuen eta guk *WordNet* erabiliko dugu. Euskarazko hitzekin lan egin ahal izateko

beharrezkoa izango da ingelesa-euskara hiztegi elebidun bat; *Morris* hiztegia (Morris, 1999) izan da guk aukeratu duguna.

Metodo honetan lehenik entrenamendu fase bat burutuko da. Honela *WordNet*en dauden 26 eremu semantiko bakoitzeko hitz adierazgarrienak lortuko dira, *Euskal Hiztegia* corpus moduan erabiliz. Horretarako eremu semantiko bakoitzeko hitzen inguruan (perpauis beretan) dauden hitzak hartuko dira eta "mutual information" neurriaren bidez dagozkien pisuak lortuko dira.

Desanbiguatzean, testuinguruarekin eta adiera desberdinekin bektoreak eraikitzen dira. Kasu honetan, sarrera bakoitzeko hitzek *WordNet* kategoria bakoitzari zenbat pisu ematen dioten hartzen da bektore bakoitzean. Bektoreen arteko distantzia kalkulatu da aurreko kasuan bezala, kosinu funtzioa erabiliz eta adiera hurbilenari pisu maximoa lotzen zaio.

2.6. Dentsitate/Distantzia kontzeptuala

Algoritmo honen oinarrian *WordNet* ontologia (ik. 4. atala) erabiliz hitzen arteko distantzia kontzeptuala ebaluatzen duten metodoak daude. Bereziki, aztertu direnak Sussna-ren distantzia kontzeptuala (Sussna, 1993) eta Agirrerren dentsitate kontzeptuala (Agirre, 1999) izan dira. Metodo hauek izenekin soilik aplikatu dira.

Algoritmoak egiten duena zera da: testuinguru eta adiera bakoitza osatzen duten definizioetatik hitz nagusiak (sinonimo edo genus bezala markatuta daudenak) erauzten dira lehenbizi. Hitz hauen ingelesezko ordainak lortzen dira ondoren (*Morris* hiztegia erabiliz). Adiera bakoitzeko, loturiko ingelesezko itzulpen eta testuinguruko itzulpenen artean *WordNet* ontologian dagoen distantzia kontzeptuala ebaluatzen da (horretarako bi metodo erabili dira, Sussnaren distantzia kontzeptuala eta Agirrerren dentsitate kontzeptuala oinarritzat dituztenak). Azkenik, antzekotasun handiena ematen duen adierari ematen zaio balio maximoa.

2.7. Karov and Edelman (1998)

Algoritmo honen funtsa definizio honetan dago: antzeko perpauisak antzeko hitzak dituztenak dira, eta antzeko hitzak antzeko perpauisetan agertzen direnak dira. Definizioaren zirkulartasuna ebazteko prozesu iteratibo bat burutzen da.

Lehenik entrenamendu fase bat burutu behar da. Honetan, desanbiguatuko den hitzaren adiera desberdinei hiztegiko perpauisak lotuko zaizkio, adiera horren definizioan dauden hitzak dituzten perpauisak hain zuzen. Perpauis hauek adiera jakin baten adierazgarri izango dira.

Desanbiguatzean, perpauis berria (testuinguru) eta adiera desberdinei loturiko hitzen antzekotasuna ebaluatuko da. Horretarako perpauis guztiekin eta hauek dituzten hitzekin matrizeak eraikiko dira. Bi matrizeek (perpauis-matrizea eta hitz-matrizea) iteratiboki beren balioak elkar eguneratuko dituzte antzekotasun formulen bidez. Azkenik, iterazioa amaitzean perpauis berriari bere perpauis antzekoenari loturiko adiera emango zaio.

3. Algoritmoen parametrizazioa

Algoritmoak aztertzean parametro desberdinak probatu dira, emaitza onenak lortzearren. Hauek izan dira landu direnak:

- Adibideak: adiera bakoitzeko ezaugarriak erauztean, hiztegi-sarrerako definizioak bere horretan edo adibideekin batera erabil daitezke. Emaitza hobeak lortu dira adibideak ere hartzean. Hala ere, desanbiguatu nahi den hitzaren testuinguru bezala definizioa soilik hartuz emaitzak hobeak dira, aurretik pentsa zitekeen bezala.
- Lemak/Hitz-formak: adierako eta testuinguruko hitzak konparatzean lemak edo hitz-formak erabili daitezke. Hiru erara probatu da: lemak, hitz-formak eta bien konbinaketa. Enpirikoki ikusi da lema erabiltzean emaitza hobeak lortzen direla algoritmo gehienetan, baina agerkidetzeta-bektoreekin, lema eta hitz-formen konbinaketa da onena.
- *Baztertze-hitz* zerrenda:

Perpau osoa erabili beharrean, maiztasun altua duten eta informazio gutxi ematen duten hitzak (*baztertze-hitzak*) baztertuz probak egin dira. Hitz hauek lortzeko *Euskal Hiztegiako* lema-zerrenda hartu da eta maiztasun handieneko 50ak erabili dira. Hiru aukera probatu dira:

- Ez-baztertu: hitz guztiak hartzen dira.
- Baztertu: kategoria nagusietakoak (izen, aditz, adjektibo eta adberbioak) ez diren hitzak baztertzen dira.
- Baztertu2: kategoria nagusietakoak ez diren hitzak eta maiztasun handieneko hitzak baztertzen dira.

Erabilitako algoritmoaren arabera, lagungarria izan daiteke *baztertze-hitzak* erabiltzea edo ez.

- Agerkidetzak:

Algoritmo batzuek hitzen agerkidetzei buruzko informazioa erabiltzen dute. Gure kasuan, bi hitz agerkideak direla esango dugu definizio berean agertzen badira. Agerpen kopurua zenbatuz eta normalizatuz ("*association ratio*" edo "*mutual information*" erabiliz; (Rigau et al., 1997)) hitz bikote agerkide bakoitzari pisu bat esleitzen zaio.

Agerkidetzeta datuak *Euskal Hiztegiako* definizio eta adibideetatik erauzi dira. Datu hauek erauztean hiru faktore eduki dira kontuan:

- lemak edo hitz-formak erabili
- sarrera bakoitzeko definizioak bakarrik edo adibideekin
- "*association ratio*" edo "*mutual information*" erabili hitzen erlazioa ebaluatzerakoan

Beraz, aukera hauen arabera 8 (2x2x2) agerkidetzeta fitxategi sortu dira eta guztiak probatu dira.

- Inplementazio murriztapenak:

(Karov and Edelman, 1998) algoritmoa aplikatzean, konputazionalki oso garestia denez, murriztapen batzuk ezarri behar izan dira enpirikoki: iterazio-kopurua (3), adiera bakoitzeko perpaus kopurua (100) eta desanbiguatzen diren perpausen kopurua (50).

4. Erabilitako baliabide lexikoak eta tresna linguistikoak

- *Euskal Hiztegia* (Sarasola, 1996): euskarazko hiztegi elebakarra. Euskara estandarra jasotzen du eta nagusiki literatura tradizioan oinarrituta dago. Guztira 30.715 sarrera eta 41.699 adiera ditu. Hitzen adieren definizioekin batera adibideak ematen ditu sarrera batzuentzat
- *WordNet 1.5: WordNet Datu-Base Lexikala* (Fellbaum, 1998) Princeton Unibertsitatean garatua izan da. Hiztegi estandar baten moduan, *WordNetek* hitzen definizioak ditu. Hiztegi estandar batekiko duen aldean, hitzen bidez antolatua egon beharrean, kontzeptuen bidezko antolamendua izatea da. *WordNeteko* oinarritzko unitatea sinonimo multzoa edo "synset" izenekoa da, eta kontzeptu lexikalizatu bat adierazten du. Adibidez, *WordNet 1.5-en shot* hitzaren bi adierak bereizteko {shot, snapshot} eta {shot, injection} synset-ak daude. Kategoría bakoitzak antolamendu berezia du *WordNeten*: izenak eta aditzak hierarkikoki antolatzen dira, eta adjektiboak antonimiaren bidez. Kontzeptuen artean erlazio desberdinak daude: hiponimia/hiperonimia, meronimia/holonimia, entailment (aditzetan)...
- *Morris* hiztegi elebiduna: (Morris, 1999) aurretik esan bezala, beharrezkoa da hiztegi elebidun bat *WordNet* ingelesezko ontologia erabiltzeko. Formatu elektronikoa eskura dagoen 80.000 sarrerako euskara-ingelesa/ingelesa-euskara hiztegi hau aukeratu genuen
- Agerkidetza fitxategiak: algoritmoen parametrizazioa azaltzean aipatu den bezala, hitzen agerkidetzari buruzko datuak erauzi dira *Euskal Hiztegiatik*. Agerkidetza fitxategi hauetako informazioa lengoaiaren tratamenduaren beste lanetarako erabil daiteke
- Euskararako analizatzaile morfosintaktikoa (MORFEUS): euskara eratorpen handiko lengoaiaren izanik, beharrezkoa da analisi morfosintaktiko sendoa burutzea. Horretarako erabili da MORFEUS tresna (Alegria et al., 1996; Urkia, 1997)

5. Ebaluazioa

Sistema ebaluatzeke eta parametroak doitzeko bi linguistak eskuz desanbiguatu dituzten ausazko laginak erabili dira. 400 definizio hartu dira kategoría bakoitzerako (izen, adjektibo, eta aditzak); linguistek hitz baten definizioa eta definizio horretan agertzen den hitz bat eman zaie (sinonimo edo hiperonimo bezala identifikatu dena). Azken hitz hau *Euskal Hiztegiako* zein adieratan agertzen den adierazi dute linguistek eta ondoren sistema automatikoki gauza bera egiten saiatu da.

Kasu batzuetan adituek hitzaren adiera egokia (edo hitz-sarrera bera) *Euskal Hiztegian* ez zegoela adierazi dute. Beste zenbaitetan ikusi da sinonimo edo hiperonimo bezala aukeraturiko hitzak benetan ez zuela funtzio hori betetzen definizioan. Kasu hauek ebaluaziotik baztertu egin dira, eta baita sinonimo/hiperonimo monosemikoak zituzten kasuak ere. Beraz guztira 358 kasu geratu dira izenentzat, 394 adjektiboentzat eta 359 aditzentzat.

Aipatu behar da aditzetan baztertu diren kasu askotan *collocation* baten osagai bat identifikatu dela genus bezala, eta ez "kolokazio" osoa. Hau gertatu da batez ere "egin" eta "izan"

aditzekin; adib.: “hegan egin”, “uste izan”... kasu hauek detektatzea beharrezkoa litzateke adieradesanbiguazioa egin aurretik. Bestalde, erabilitako MRDa erdi-automatikoki erauzi denez, bertan ematen diren erroreak eragina dute; adibidez “tresna” hitza sarritan identifikatzen da genus bezala izenentzat (ontologietan goiko mailetan agertu ohi den hitza delako), eta normalean adiera jakin batekin erabiltzen da hiztegiatiko definizioetan. Gure kasuan, jatorrizko hiztegian dagoen sarrera anbigua dela eta, adiera egoki hori galdu egin da tratamendu automatikorako, eta, beraz, baztertu behar izan ditugu kasu horiek. Arrazoi hauengatik bazterturiko kasuen zerrenda A eranskinean ematen da.

Adituek adiera bat baino gehiago eman dute ontzat kasu batzuetan. Gure sistema metodo heterogeneo askoren konbinaketa denez, bi adieraren arteko berdinketa gertatzeko aukera eskasa da, eta, normalean, emaitza bakarra ematen du. Gure sistemak adiera egokietako bat asmatzen badu ontzat eman dugu.

Hiztegiaren eta laginen polisemiari buruzko datuak 1. taulan aurkezten dira. Datu hauen arabera, ikus daiteke ebaluatzerakoan emaitzak asko aldatuko direla kategoria batetik bestera. Aditzetan ematen da polisemia handiena; gure helburua sinonimo eta genus polisemikoen desanbiguazioa da, eta hiztegian mota honetako aditzek 4,75 adiera dituzte batez beste, adjektibo eta izenek 2,41 eta 2,82 dituzten bitartean. Gainera, adjektiboen kasuan, linguistek batez beste 1,42 adiera eman dituzte ontzat laginean. Beraz, errazagoa izango da adjektiboekin asmatzea.

Kategoria	Batez besteko adierak <i>Euskal Hiztegian</i>			Batez besteko adierak laginetan	
	Sarrerak	Sin./Genus	Sin./Genus Polisemikoak	Adierak	Adiera "onak"
Izenak	1,19	1,85	2,82	2,86	1,08
Aditzak	1,34	3,86	4,75	4,53	1,10
Adjektiboak	1,12	1,41	2,41	2,38	1,42

1. taula: *Euskal Hiztegiaren* eta laginen adiera banaketa

2. taulan agertzen dira algoritmo desberdinekin lorturiko doitasuna eta estaldura. Algoritmo bakoitzari 2. ataleko zenbakia eman zaio (‘6’ Distantzia Kontzeptuala izango da eta ‘6b’ Dentsitatea). Erreferentzia bezala, adiera ausaz aukeratuz asmatzeko itxaropen matematikoa ematen da. Lehen algoritmo sinplea ere (beti Hiztegiko Lehen Adiera ematea, HLA hemendik aurrera) erreferentzia bezala erabili izan da sarritan literaturan.

Esan bezala, algoritmo bakoitzaren parametroak enpirikoki doitu dira. Doiketa hau izenentzat soilik egin da eta parametrizazio bera erabili da beste kategorientzat. Agian emaitzak hobetu zitezkeen algoritmoak kategoria bakoitzerako egokituz, baina hau ez da aztertu. Parametrizazio desberdinentzat lorturiko emaitzak B Eranskinean agertzen dira.

Izenei dagokienez, aipatu behar da probatu diren algoritmo guztiek ez dituztela emaitza onak eman. Bereziki, ontologietan oinarrituriko metodoak (dentsitate eta distantzia kontzeptualak) ez dute HLA gaintitu eta besteekin konbinatzean ez dituzte hobetu sistemaren emaitzak. Aurretik esan bezala, metodo hauek aplikatu ahal izateko euskarazko ontologia bat ez zegoenez, ingelesezko *WordNet* ontologia erabili da eta honek eragin handia izan du emaitzetan. Gainera, metodo hauek izenentzat soilik aplika daitezke. Beste algoritmo guztiek gaintitu dituzte ausazkoa eta HLA. Bestalde, (Karov and Edelman, 1998) metodoa (7) izenekin probatu da eta konbinatzean emaitzak igo ditu (%80,1etik %80,4ra), baina konputazionalki oso garestia denez ez da aplikatu beste kategoriekin. Orokorrean, izenekin lortu diren emaitzek %15.6an gaintitu dute HLA eta %37.7n ausazko metodoa. Izenekin eman da hobekuntza onena oinarritzko metodoekiko.

Adjektiboekin lortu dira emaitza orokor onenak (%84). Aurretik esan den bezala, hau espero zitezkeen, adjektiboetan adiera asko ematen baitira ontzat batez beste (1,42). Kasu honetan %10,9n gaintitzen da HLA eta %13n ausazkoa.

Aditzekin lortzen dira emaitza kaxkarrenak, oso polisemikoak baitira. Ausazkoa metodo guztiek gainditzen badute ere, HLA "Teilakapen" algoritmoak soilik gainditzen du (gainera estaldura baxuarentzat). Hala ere, metodoak konbinatzean HLA %4,2n gainditzen da eta ausazkoa %34n, %66ko doitasuna lortzen delarik.

Orokorrean, doitasun onena lortzen duen algoritmoa "Teilakapena" (2) delakoa da; baina bere estaldura baxua da. Agerkidetzetan oinarriturikoak (3,4) eta Yarowskyren metodoak (5) estaldura altua dute baina doitasuna nabarmenki jaisten da, batez ere aditzetan. Nahiz eta algoritmo batzuen emaitzak kaxkarrak izan, eginiko esperimenduetan ikusi da guztiek hobetzen dituztela metodoaren emaitzak (dentsitate eta distantzia kontzeptualak izan ezik, aipaturiko arrazoiengatik). Beraz, doitasun eta estaldura eskaseko heuristikoek ere lagun dezakete beste heuristikoek ez duten ezagutzarekin (Rigau et al., 97).

	Ausaz	(1)	(2)	(3)	(4)	(5)	(6a)	(6b)	(7)	Batura (1-5,7)
Izenak										
• Doitasuna	%42,7	%64,8	%85,3	%74,2	%72	%70,6	%57,9	%43,3	%65	%80,4
• Estaldura	%100	%100	%32,4	%98,6	%100	%98,8	%24,5	%25,1	%80,7	%100
Aditzak										
• Doitasuna	%32	%61,8	%80	%59,2	%61	%53,9	-	-	-	%66
• Estaldura	%100	%100	%19,5	%99,1	%100	%99,1	-	-	-	%100
Adjektiboak										
• Doitasuna	%61	%73,1	%96	%78,5	%83,2	%77,3	-	-	-	%84
• Estaldura	%100	%100	%38	%94,6	%100	%98,7	-	-	-	%100

2. taula: algoritmoen doitasuna eta estaldura kategoriak

3. taulan algoritmo konbinaketa desberdinek izenen laginarekin eman dituzten emaitzak eta behar izan duten CPU denbora daude "Sun Sparc 10" makina batean (384 Mb Ram eta 360 Mhz.). Oinarritzko konbinaketa 1-5 algoritmoek osatzen dute. Gainontzeko zutabeetan, oinarritzko konbinaketa horri beste algoritmoak gehitzean gertatzen dena ikusten da. (Karov and Edelman, 1998) algoritmoa (7) eta distantzia kontzeptuala (6a) oso garestiak dira konputazionalki eta emaitzen igoera nabarmenik ekartzen ez dutenez, *Euskal Hiztegia* osoaren analisitik baztertzea erabaki da. Dentsitate kontzeptuala (6b) ere ez da erabili, bere emaitza baxuak direla eta.

	(oinarritzkoa, 1-5)	(+7)	(+6a +7)	(+6b + 7)
Doitasuna	%80,16	%80,4	%76	%77,3
Estaldura	%100	%100	%100	%100
Exekuzio denbora	< 10'	11 ordu	18 ordu	11 ordu 30'

3. taula: algoritmo konbinaketan emaitzak eta exekuzio denborak izenen laginekin

Metodoa laginekin probatu ondoren, *Euskal Hiztegiatik* erauzitako sinonimo eta genus guztiei aplikatu zaie. Aipatu behar da hitzen arteko erlazio semantikoaren doitasuna %97 baino altuagoa dela kategoria guztientzat (Agirre et al., 2000a). Horretarako azterturiko lehen bost algoritmoen konbinaketa erabili da. Datuak 4. taulan agertzen dira. Hitz hauetako asko monosemikoak izango dira, eta, beraz, amaieran lortzen ditugun adieren arteko erlazio semantikoek %90 baino doitasun altuagoa izango dute izenentzat, %70 aditzentzat eta %95 adjektiboentzat. Estaldurari dagokionez, emaitzarik gabe geratzen diren kasu bakarrak *Euskal Hiztegian* sarrera bezala agertzen ez direnak dira, gainontzeko kasuetan emaitzaren bat itzultzen da.

	<i>Guztira</i>	<i>Monosemikoak</i>	<i>Polisemikoak</i>	<i>Hiztegia ez daude</i>	<i>Estaldura</i>	<i>Doitasuna</i>
Izenak	26.461	12.328 (%46,6)	10.909 (%41,2)	3.224 (%12,1)	%87,8	%90,6
Aditzak	7.700	1.715 (%22,3)	5.530 (%71,8)	455 (%5,9)	%94,1	%70,2
Adjektiboak	2.389	1.547 (%64,7)	652 (%27,3)	190 (%7,9)	%92,1	%95,2

4. taula: *Euskal Hiztegiatik* erauzitako erlazio semantikoen desanbiguazioa

6. Hierarkiaren eraikuntza automatikoa

Euskal Hiztegiatik erauzitako erlazio semantiko desanbiguatuetatik abiatuz, sare semantiko baten eraikuntza automatikoa aztertu zen. Helburua, informazioa *WordNet* moduan errepresentatzea zen (ik. 4. atala), kontzeptu bakoitzerako sinonimo multzoak erabiliz eta kontzeptuen arteko loturak ezarriz. Horretarako, *WordNetek* eskaintzen duen ‘Grinder’ konpilatzailea erabili genuen. Tresna honek, sinonimo multzo eta genus erakusleetatik abiatuz sare semantikoa eraikitzen du. Lan hau izen eta aditzekin egingo da, adjektiboek ez baitira antolamendu hierarkikoarekin errepresentatzen.

Lehen urratsa sinonimo multzoak eraikitzea izan zen. Hau egiteko kontuan eduki behar da hiztegiko sinonimoek duten egitura. Hiztegiaren egilearen arabera (Ibon Sarasola), 4 definizio mota bereizten dira *Euskal Hiztegia* ('sin' bidez sinonimoa adierazten da, eta 'def' bidez definizioa):

1. sin, sin, ...
2. sin, def
3. def, sin
4. def

Lehen mailako sarrerak, garrantzitsuenak, 3. eta 4. motakoak dira. Bigarren mailako sarrerak, 1. eta 2. motakoak dira. 'sin' horien definizioetan ez da hiztegiko sarrera azaltzen. Gainera "ik." eremua dago. Garbi dago denak ez direla maila bereko sinonimoak. Informazio hau, momentuz, ez da erabili sinonimo multzoen sorkuntzan eta aurrerago egiteko lana bezala geratzen da.

Hiru hurbilpen hartu dira kontuan *Euskal Hiztegiako* sinonimo desanbiguatuak multzokatzeko, lehen bi hurbilpenei buruzko datu gehigarriak C Eranskinean ematen dira:

- lehen hurbilpena: sinonimiaren trantsitibotasuna erabili zen eta honen arabera sortu ziren multzoak. Adibidez, "jator#1"¹ eta "jatorri#2" sinonimotzat ematen badira eta baita "jatorri#2" eta "ama#3", orduan "jator#1" eta "ama#3" ere sinonimotzat hartuko dira. Baina bide hau jarraituz, izenentzat 3.245 hitzez osaturiko "sinonimo" multzoa sortzen da. Hau gertatzen da hiztegiatiko definizioetan benetan sinonimoak ez diren hitzak, hots, ia-sinonimoak (near-synonyms) sarritan ematen direlako. Honetaz gain, desanbiguazio automatikoa ematen diren akatsek (ia %10 izenentzat) ere eragina dute. Jarraian, sinonimiaren trantsitibotasuna nola hedatzen den ikus daiteke:

jator.1 jatorri.2
jatorri.2 ama.3
ama.3 iturburu.1
iturburu.1 iturri.3
iturri.3 etorki.3
etorki.3 sorburu.1
sorburu.1 iturburu.1
iturburu.1 sustrai.2
sustrai.2 zimendu.1
zimendu.1 oinarri.1
oinarri.1 oin.6

¹ *jator* hitzaren 1 adiera.

oin.6 pata.1
 oin.6 oinkari.2
 oinkari.2 oinezko.1
 oin.6 zango.2
 oinarri.1 sustrai.2
 sustrai.2 jatorri.2
 jatorri.2 sorburu.1
 jatorri.2 etorburu.1
 etorburu.1 etorki.3
 etorki.3 iturburu.2
 iturburu.2 iturbegi.1
 iturburu.2 hasiera.2
 hasiera.2 hasialdi.1
 ...

- bigarren hurbilpena: trantsitibotasuna mugatzea erabaki zen eta soilik benetako sinonimo izateko posibilitate gehienak zituztenak aukeratzea heuristiko baten bitartez. Honela, sinonimo bezala agertzen ziren bikoteen artean, genus berak zituztenak bakarrik onartzen ziren (edo bikoteko batek genusa zuenean eta besteak ez). Modu honetan sorturiko sarean *synseten* banaketa askoz orekatuagoa zen eta gehienez 37 sinonimoko multzo bat sortu zen. Multzo hauetatik abiatuz bi sare semantiko banatu sortu ziren izen eta aditzetarako. Sarearen datu batzuk 5. taulan ematen dira:²

	Izenak	Aditzak
Synset kopurua	12.219	3.485
Synset-en batez besteko hitz kopurua	1,53	1,49
Isolaturiko synset kopurua	1.210	126
Erro-synset kopurua	1.492	180
Hosto-synset kopurua	10.490	3.067

5. taula: sare semantikoaren ezaugarriak

- hirugarren hurbilpena: beste aukera bat sorturiko sare semantikoaren unitate bezala adiera uztea zen eta beraz sinonimoak ez multzokatzea. Aukera honen azterketa ere aurreragoko geratu da (Ansa eta Arregi, 2000).

7. Aurreko lanekiko konparaketa

Gure lanaren oinarrian (Rigau et al., 1997) metodoa dago. Artikulu horretan %82 eta %83ko doitasuna ematen zen hurrenez hurren frantses eta gaztelaniazko hiztegietatik erauzitako erlazioentzat. Gure lanean kategorien arabera emaitzak oso desberdinak izan dira eta %95 lortu dugu adjektiboentzat, %90 izenentzat, eta %70 aditzentzat (honen zergatia 8. atalean aipatzen da). (Rigau et al., 1997)-n azaltzen ziren metodoez gain, beste bat probatu dugu (Karov and Edelman, 1998), arrakasta gutxiarekin emaitza aldetik.

Beste hitzen adiera-desanbiguazio lanekin konparaketa egitea zaila da. Kontuan eduki behar da hiztegi sinonimo eta genusen desanbiguazioak hitzen adiera-desanbiguazio orokorrak ez dituen zailtasun batzuk dituela: hitzek polisemia altuagoa izaten dute, testuinguru murrizta, "one sense per discourse" ezin aplikatu... (Rigau et al., 1997).

8. Ondorioak eta etorkizunerako lana

Lan honetan hitzen adiera-desanbiguaziorako metodo ez-gainbegiratuak aztertu eta aplikatu dira euskararako. Aplikazio eremua hiztegietatik informazio-erazketa eta Datu-Base Lexikalen eraikuntza automatikoa izan da. Aurretik esan bezala, algoritmo gainbegiratuak emaitza hobek

² Hurbilpen hau jarraituz sorturiko sare semantikoa ikusgai dago IXA taldeko partaideentzat. Ik.: http://ixa.si.ehu.es/taldea/azpataldeak/lexikosemantika/materiala/EH_interfaz.html

lortzen dituztela dirudi gaurko literaturan, baina lan honetan emaitza aipagarriak lortu dira, izenekin %90, eta adjektiboekin %95eko doitasuna (monosemikoak barne). Hala ere, aditzekin emaitzak ez dira onak izan (%70). Hau arrazoi desberdinengatik gertatzen da:

- aditzek beste kategoriek baino polisemia altuagoa izaten dute (ik. 1. taula)
- genus bezala identifikaturiko aditzak gehienetan oso orokorrak dira eta ontologiaren goi mailetan agertzen dira (adib.: "egin", "izan"...). Aditz hauen adiera banaketa oso lausoa izaten da eta sarritan gizakientzat ere zaila gertatzen da adiera egokia identifikatzea. Hala ere, normalean adiera jakin batekin agertzen dira hiztegiko definizioetan eta hori kontuan hartuz posiblea litzateke aditz horiek modu berezian lantzea

Bestalde, genus bezala identifikaturiko aditz "orokor" horiek zenbaitetan "kolokazio" baten osagai moduan agertzen dira (adib.: "hegan egin", "uste izan"...). Kasu hauek gure analititik baztertzen dira, baina etorkizunean aurretik identifikatu beharko lirateke prozesatu ahal izateko. Aurrerago landu beharreko beste arazo bat MRDaren erroreen zuzenketa da. IXA taldea lanean ari da dagoeneko *Euskal Hiztegiatik* erauzitako informazioaren eskuzko orrazketan.

Aplikatu diren algoritmoen dagokienez, aipagarria da (Karov and Edelman, 1998) metodoak ez dituela emaitza onak lortzen. Beren artikuluan 4 hitz soilik aztertzen dituzte eta arazoa oso desberdinak diren bi adieraren artean bereiztera mugatzen dute. Beren emaitzak %90tik gorakoak badira ere, lan honetan, hiztegiko izenetara eta dituzten adiera finetara hedatzean, %65eko doitasuna besterik ez da lortu. Horretaz gain, 3. taulan ikusten den bezala, kostu konputazionala izugarria da.

*WordNet*eko distantzian eta dentsitatean oinarrituriko metodoek ere emaitza kaxkarrak lortu dituzte doitasun eta estaldurari dagokionez. Nahiz eta (Rigau et al., 1997) artikuluan metodo hauek gaztelaniara eta frantsesera arrakastaz hedatu hiztegi elebidunen bidez, ez da gauza bera gertatu euskara eta gure baliabideekin eginiko saiakerekin. Hala ere, (Yarowsky, 1992) metodoak emaitza onak lortzen ditu, eta kasu honetan ere beharrezkoak dira ingelesa-euskara ordainak.

"Teilakapena" metodoak oso doitasun altua lortu du (%80 aditzentzat, %85 izenentzat eta %96 adjektiboentzat), bere estaldura oso baxua bada ere. Badirudi *Euskal Hiztegian* sarritan erabiltzen direla hitz berdinak semantikoki antzekoak diren adierak definitzeko. Emaitza hauek kontuan izanik, aukera bat izango litzateke algoritmo honi pisu gehiago ematea besteekin konbinatzean, baina hau ez da aztertu. Agerkideztetan oinarrituriko metodoek ere nahiko doitasun altua ematen dute (aditzetan izan ezik) eta ia erabateko estaldura.

Orokorrean, lan honetan ikusi da algoritmo ez-gainbegiratu heterogeneoen konbinaketek emaitza onak ematen dituztela (Rigau et al., 1997) eta metodo berrien gehikuntza aberasgarria izaten dela. Emaitzak hobetzeko modu bat algoritmo sinple gehiago sartzea izan daiteke, metodoa muturreraino eramanez. Konbinaketa egiteko modu desberdinak ere azter daitezke, algoritmoen eraginkortasuna kontuan hartuz.

Euskarazko ezagutza-base aberats bat hitzen adiera-desanbiguaziorako oso lagungarria litzateke. Lan honetan ezagutza basearen eraikuntza automatikoa aztertu da, baina saiakera moduan. Ez dira sakonean landu prozesu honek dituen zenbait arazo:

- o ezagutza-basearen goiko mailaren azterketa
- o hiztegiko erlazioen benetako izaera ezagutza-basean: benetako sinonimo eta genusak zeintzuk diren
- o sinonimia eta hiperonimia ez diren erlazioen integrazioa ezagutza-basean
- o automatikoki sorturiko ezagutza basearen "kalitatea" neurtu: ezagutza basearen errore motak identifikatu eta zenbakitu

IXA taldeak badu euskararen ezagutza-base aberats bat eraikitzeke asmoa. Lan hau bide horretan joango litzateke. Hala ere, helburu hau lortzeko eskuzko orrazketa ezinbestekoa da. Baliabide honen bitartez, desanbiguaziora ontologietan oinarrituriko metodoak aplikatu ahal izango

lirateke. Honetaz gain, adibide etiketatuak automatikoki lortzeko lagungarria izan daiteke (Internet edo euskarazko corpusetan bilaketak eginez).

9. Bibliografia

- Agirre E. *Kontzeptuen arteko erlazio-izaeraren formalizazioa ontologiak erabiliaz: Dentsitate Kontzeptuala*. Tesi-txostena. Informatika Fakultatea, UPV-EHU. 1999.
- Agirre E., Ansa O., Arregi X., Artola X., Díaz de Ilarraza A., Lersundi M., Martinez D., Sarasola K. and Urizar R. *Extraction of semantic relations from a Basque monolingual dictionary using Constraint Grammar*. EURALEX 2000. Stuttgart, Germany.
- Agirre E. and Martinez D. *Decision lists and automatic word sense disambiguation*. COLING 2000, Workshop on Semantic Annotation and Intelligent Content. Luxembourg.
- Alegria I., Artola X., Sarasola K. and Urkia M., *Automatic morphological analysis of Basque*. In *Literary & Linguistic Computing* vol. 11, nº4 (1996) 193-203, Oxford, Oxford University Press. 1996.
- Amsler, Robert. *A Taxonomy for English Nouns and Verbs*. In *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, pages 133-138. Standford, California. 1981.
- Ansa, O. eta Arregi X. *EHEBren diseinua*. Argitaratzeko zain dagoen barne-txostena. 2000.
- Fellbaum, C. *WordNet, An Electronic Lexical Database*. Cambridge, MA: MIT Press. 1998.
- Ide N. and Véronis J. *Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art*. *Computational Linguistics*, vol. 24, n. 1. 1998.
- Karov Y. and Edelman S., *Similarity-based Word Sense Disambiguation*. *Computational Linguistics*, vol. 24, n. 1. 1998.
- Leacock C., Chodorow M., Miller G. A., *Using Corpus Statistics and WordNet Relations for Sense Identification*. *Computational Linguistics*, vol. 24, n. 1. 1998.
- Mihalcea R. and Moldovan D. I., *An Automatic Method for Generating Sense Tagged Corpora*. *Proceedings of AAAI'99*. Orlando (Florida). July 1999.
- Miller G.A., Leacock C., Teng R., and Bunker R.T., *A Semantic Concordance*, in *Proceedings of the ARPA Workshop on Human Language Technology*, (1993).
- Morris, M. *Morris Hiztegia (Basque-English/English-Basque)*. Klaudio Harlouxet Fundazioa. 1999.
- Ng H.T. and Lee H. B., *Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-based Approach*, in *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*. ACL, (1996).
- Ng H.T., Lim C.Y. and Foo S.K. *A Case Study on Inter-Annotator Agreement for Word Sense Disambiguation* in *Siglex 1999 - ACL 1999*. Workshop on Standardizing Lexical Resources.
- Rigau G., Atserias J., Agirre E. *Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation*. *Proceedings of ACL'97/EACL'97*, 48-56. Madrid, Spain. July 1997.
- Sarasola, I., *Euskal Hiztegia*, Donostia, Gipuzkoako Kutxa. 1996.
- Sussna, M. *Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network*. In *Proceedings of the Second International Conference on Information and Knowledge Management*. Arlington, Virginia. 1993.
- Urkia M., 1997, *Euskal Morfologiaren Tratamendu Informatikorantz*, Gazteiz, UPV/EHU, PhD Thesis.
- Vossen P. and Serail I. *Word-Devil, a Taxonomy-Browser for Lexical Decomposition via the Lexicon*. *Esprit BRA-3030 Acquilex Working Paper* n. 009. 1992.
- Wilks Y., Fass D., Guo C., McDonald J., Plate T. and Slator B. *Providing Machine Tractable Dictionary Tools*. In *Pustejovsky J. ed. Semantics and the Lexicon*, pages 341--401. 1993.

Yarowsky D., *Word-Sense Disambiguation Using Statistical Models of Rogets Categories Trained on Large Corpora*. In Proceedings of the 14th International Conference on Computational Linguistics (Coling'92), pages 454-460. Nantes, France. 1992.

A. Eranskina: Laginetan aztertu ez diren kasuak

Jarraian kategoriaka aurkezten dira analisitik bazterturiko kasuak eta honen zergatia azaltzen da. "X" karakterearen bidez genusa ondo identifikatu ez dela adieraziko dugu, eta "?" bidez adiera egokia ez dagoela aukeren artean.

- **Izenak:** baztertu diren 42 adibideetatik, 24 kasutan genusa gaizki identifikatua zegoen, 10etan adiera egokia ez zegoen hiztegian, eta gainontzeko 8etan arrazoi desberdinak eman dira (polisemiarik ez, errepikaturiko kasuak...). Aipatu behar da "tresna" 6 aldiz ematen dela genus bezala, eta bere adiera zuzena MRDan ez dagoenez, kasu guztietan baztertu egin da.

Jarraian genus egokia ez izategaitik edo adiera zuzena MRDan ez zutelako bazterturiko kasuak:

34: bijilia GENUS Eliz besten bezpera, barau eta haragi uzte egin behar zena.

 behar 1 () Errua.
 behar 2 (H2.A1.)Zerbaierako nahitaezko den zera; zerbait beharrezko gertatzea.
 behar 3 (H2.A2.)Ilan egin.

OK: 2X

83: haize_orratz GENUS Kanpandorreetan-eta kokaturik, haizearen norabidea adierazten duen tresna, askotan oilar itxura duen metalezko atal birakor batek osatua.

 tresna 1 (A1.) Soinekoa, jantzia.
 tresna 2 (A2.) Zamariei eta, karga eraman ahal izateko, jartzen zaizkien gauzakiak.
 tresna 3 (A3.) Izenondo gisa.

OK: ?

94: igorgailu GENUS Uhin elektromagnetikoen bidez ahotsa edo irudia igortzen duen tresna.

 tresna 1 (A1.) Soinekoa, jantzia.
 tresna 2 (A2.) Zamariei eta, karga eraman ahal izateko, jartzen zaizkien gauzakiak.
 tresna 3 (A3.) Izenondo gisa.

OK: ?

95: ilargi GENUS Lurraren satelitea, Eguzkiak argitzen duena, gaueko argizagi nagusia; haren itxura Lurretik ikusten den bezala.

 nagusi 1 (MA.A1.) Etxe bateko jauna zerbitzariet buruz; zenbait erakunde edo talderen burua mendekoei buruz; lantegi edo fabrika bateko jabea.
 nagusi 2 (MA.A2.) Maisua.

OK: X

110: kazetagintza GENUS Kazetak egiteko jarduna.

 egiteko 1 (A1.) Norbaitek egiteko duena, norbait ari dena edo norbaiti dagokiona; arazoa.
 egiteko 2 (A2.) Egintza askotan gaitesgarritzat hartua.

OK: X

111: kilowatt GENUS Elektrizitate ahalmen banakoa.

 ahalmen 1 (A1.) Zerbait egiteko ahala, eskua, gaitasuna.
 ahalmen 2 (A2.) (ADIB.) Makina, motor baten ahalmena: denbora jakin batean egiten duen lanaren bidez neurtzen dena. Ahalmena watt-etan neurtu.

OK: X

191: diplomazia GENUS Politikaren adarra, estatuen arteko harremani dagokiona; delako harremanak zertzen dituzten pertsona eta erakundeen multzoa.

 pertsona 1 (A1.) Giza motako banakoa.
 pertsona 2 (A2.) Literatura lan baten ekintzan esku hartzen duten gizakietako bakoitza.
 pertsona 3 (A3.) Izenordainak eta adizkiak aldatzen dituen gramatika kategoria, hitz egiten duenari.

OK: 1X

195: eresi GENUS Nahi bizia, gutizia.

 bizi 1 (H1.A1.)Landareen eta animalien oinarritzko nolakotasuna, hazteko, gaiberritzeko, ugaritzeko... ahalmenean datzana.
 bizi 2 (H1.A2.)Bizitza.

OK: X

197: esne_bide GENUS Santio bidea.

 bide 1 (H1.A1.)Bertatik ibiltzeko prestatua den lur zerrenda.
 bide 2 (H1.A2.)Helburu bat iristeko balio duen zera; ezer lortzeko era edo moldea.
 bide 3 (H1.A3.)Aukera, abagunea.
 bide 4 (H1.A4.)Zergatikoa.
 bide 5 (H1.A5.)Bide.

OK: ?			
218: kautxu	GENUS		Gai malgua, hedakorra eta zaila, tropikoko hainbat landareren esne edo latex-az, edo sintesi bidez hidrokarburo asegabenez, egiten dena.
		esne	1 (A1.) Ugaztun emeen ugaztetan eratzen den isurkari zuria, hazkurri handikoa, ume jaio berrien berezko elikagaia.
		esne	2 (A2.) Zenbait landareri darien edo zenbait hazi sakatuz ateratzen den isurkaria, esnearen antzekoa.
OK: X			
221: kongresu	GENUS		Biltzarra; berezk., hainbat pertsonak beren ideiak edo ikerketak elkarri adierazteko aldian-aldian egiten dena.
		ideia	1 (A1.) Gauza baten irudikatze abstraktu eta orokorra, gogoak edo adimenak eratu.
		ideia	2 (A2.) Burutapena; ustea; asmoa.
OK: X			
229: mokoilo_edo_mokoloi	GENUS		Tretzako soka eta amuen nahastea edo korapikoa.
		soka	1 (A1.) Elkarrekin bihurtzen diren ezpartzu edo kalamuzko hari multzoa.
		soka	2 (A2.) Erabilera bereziak.
OK: X			
235: ontzigintza	GENUS		Ontziak batez ere lehen adieran egiteko jarduna.
		egiteko	1 (A1.) Norbait egiteko duena, norbait ari dena edo norbaiti dagokiona; arazoa.
		egiteko	2 (A2.) Egintza askotan gaitesgarritzat hartua.
OK: 1X			
243: pottoko	GENUS		Behorraren kume arra.
		arra	1 (A1.) Zabalduriko eskuko hatz lodienaren eta txikiaren arteko bitartea, luzera neurri gisa erabilia izan dena.
		arra	2 (A2.) Ahurra.
OK: X			
250: txorimalo	GENUS		Hazia ereindako soro edo alorretatik txoriak uxatzeko ezartzen den giza irudia, arropa zaharrez jantzia.
		soro	1 (A1.) Zelaia, belardia, belar saila.
		soro	2 (A2.) Lantzen den lur saila.
		soro	3 (A3.) Bigarren ebakialdiko belarra.
OK: X			
276: epaiki	GENUS		Ebakitzeko tresna, berezk. zerra.
		tresna	1 (A1.) Soinekoa, jantzia.
		tresna	2 (A2.) Zamariei eta, karga eraman ahal izateko, jartzen zaizkien gauzakiak.
		tresna	3 (A3.) Izenondo gisa.
OK: ?			
289: ipurtats	GENUS		Ugaztun haragijalea, arre iluna, txikia eta gorputz-luzea, Europa, Asia eta Ipar Afrikako basoetan bizi dena eta usain nardagarria botatzen duena.
		haragijale	1 (A1.) Haragia jaten duena.
		haragijale	2 (A2.) Animaliez mintzatuz, batez ere haragiz elikatzen dena.
OK: X			
299: kontrabaxu	GENUS		Biولين familiako musika tresnetan handiena eta soinu apalekoena.
		soinu	1 (A1.) Hotsa.
		soinu	2 (A2.) Musika tresna baten hotsa, belarriari atsegin zaion hotsa.
		soinu	3 (A3.) Musika, batez ere herri dantza saioetako.
OK: X			
316: transformadore	GENUS		Elektr. Korrante elektrikoaren era edo tenka aldatzeko balio duen tresna.
		era	1 (A1.) Zerbait agertzen, gertatzen edo gauzatzen den itxura berezi edo jakina.
		era	2 (A2.) Aukera, abagunea.
		era	3 (A3.) Erosotasuna, apeta.
OK: X			
322: arnegu	GENUS		Nork bere fede edo erlijioari uko egitea; uko egin.
		fede	1 (A1.) Erlijioa egia eta dogmei buruzko sinestea; bertute teologala, Elizak erakusten duena egia ziurtat sinestean datzana.
		fede	2 (A2.) Erlijio baten sinesgaien multzoa.
		fede	3 (A3.) Asmoa, borondatea.
OK: X			
326: baraila	GENUS		Ahoaren gain eta azpialdeetan dauden eta hortz-haginak ezartzen diren bi hezur arkuetako bakoitza; hainbat intsekturen ahoko kideko atalak.
		gain	1 (A1.) Edozein gauzatan, lurretik urrunen dagoen aldea.
		gain	2 (A2.) Lekuzko kasu atzizkiez, singularrean.
		gain	3 (A2.Xa.)Dagokion izen sintagmak -ren kasu atzizkia hartzen duela.
		gain	4 (A2.Xb.)Delako sintagmako izenak bizigabeko zerbait adierazten badu eta eskuinean adjektiborik ez badu, -ren atzizkia eta mugatzailea gal ditzake; orobat delako izenak hitz elkarketako era har dezake.
		gain	5 (A2.Xc.)Dagokion izen sintagmarik gabe.
OK: X			
329: buruxka	GENUS		Gari, arto eta antzeko landareetan, ardatz antzeko batera biltzen diren lore edo aleen multzoa.
		lore	1 () Lorearen korola osatzen duten hosto moduko bakoitza.

	lore	2 (A1.)	Hainbat landareren zati koloretsu eta askotan usaintsua.
	lore	3 (A2.)	Zerbaiten alde edo zati hoberena, ederrena.
OK: X			
341: mintzo	GENUS		Aditu zuela zerutik mintzo bat zioela:... Mintzo ozena eta sarkorra. Aitaren hitza mintzo bihurtua. Aingeru horien mintzo isila begiz ez ikusi arren. Bere mutuan, ez du mintzoaren beharrik oraingo bizitzaren alegia azaltzeko. Kontzientziako mintzoa. Ekin zion kilkerak bere kantu mintzo-bakarrari.
	mintzo	1 (A1.)	Aditu zuela zerutik mintzo bat zioela:... Mintzo ozena eta sarkorra. Aitaren hitza mintzo bihurtua. Aingeru horien mintzo isila begiz ez ikusi arren. Bere mutuan, ez du mintzoaren beharrik oraingo bizitzaren alegia azaltzeko. Kontzientziako mintzoa. Ekin zion kilkerak bere kantu mintzo-bakarrari.
	mintzo	2 (A2.)	Elea, hizkuntza.
OK: X			
352: akademia	GENUS		Jakintsuen, erti edo letra gizonen elkarte.
	elkarte	1 (A1.)	Elkartasuna.
	elkarte	2 (A2.)	erd.: akzio elkarte, kapitala akziotan banatua dagoena.[*Euskal Herriko Autonomia Elkarte. *Europako Ekonomia
Elkarte.			
OK: ?			
353: aldezkari	SIN		Aldekoa; defenditzailea.
	alde	1 (MA.A1.)	Gauza, toki, bitarte bati buruz, hartan bereiz daitezkeen zatietako bakoitza. Gauzaki edo puntu bati buruz, haren ingurua; hura inguratzen duen bitartetik nolabait bereizten den zatia.
	alde	2 (MA.A2.)	Izaki edo gauzaki abstraktuez.
	alde	3 (MA.A3.)	Poligono bat ixten duten lerroetako bakoitza; angelu bat osatzen duten bi lerroetatik bakoitza.
	alde	4 (MA.A4.)	Bi zer edo gauza elkarrengandik bereizten dituen nolakotasun desberdintza.
	alde	5 (MA.A5.)	Taldea.
	alde	6 (MA.A6.)	(ADIB.) Alde! Alde hemendik!
OK: ?			
357: elektrodo	SIN		Bateria, gas hodi etab.-etako buruetako bakoitza.
	bateria	1 (A1.)	Artilleria pezen multzoa; peza horiek kokatzeko babeslekua.
	bateria	2 (A2.)	Elektrizitatea metatzen duen tresna.
	bateria	3 (A3.)	Orkestra batean, perkursio tresnek osatzen duten multzoa.
OK: X			
371: ontzigintza	GENUS		Ontziak batez ere lehen adieran egiteko jarduna.
	egiteko	1 (A1.)	Norbaitek egiteko duena, norbait ari dena edo norbaiti dagokiona; arazoa.
	egiteko	2 (A2.)	Egintza askotan gaitesgarritzat hartua.
OK: X			
373: pala	GENUS		Pilotan jokatzeko tresna, heldulekuaren muturrean mehea eta bestean aski zabala den zurezko atal batez osatua.
	tresna	1 (A1.)	Soinekoa, jantzia.
	tresna	2 (A2.)	Zamariei eta, karga eraman ahal izateko, jartzen zaizkien gauzakiak.
	tresna	3 (A3.)	Izenondo gisa.
OK: ?			
374: piano	GENUS		Musika tresna handia, klavierduna, harpa gisako egitura batean kokatzen diren eta mailu moduko batzuek jotzen dituzten hariak dituena.
	tresna	1 (A1.)	Soinekoa, jantzia.
	tresna	2 (A2.)	Zamariei eta, karga eraman ahal izateko, jartzen zaizkien gauzakiak.
	tresna	3 (A3.)	Izenondo gisa.
OK: ?			
377: zizaia	GENUS		Guraize edo artazi handien moduko tresna, metalak eta ebakitzeko erabiltzen dena.
	tresna	1 (A1.)	Soinekoa, jantzia.
	tresna	2 (A2.)	Zamariei eta, karga eraman ahal izateko, jartzen zaizkien gauzakiak.
	tresna	3 (A3.)	Izenondo gisa.
OK: ?			
391: taxidermia	GENUS		Hildako animaliak, disezio bidez, bizi itxuran gordetzea.
	animalia	1 (A1.)	Sentitzeko eta berez mugitzeko gauza den izaki bizia.
	animalia	2 (A2.)	Adimenik gabeko animalia.
OK: X			
394: estremo	SIN		Burua, muturra.
	buru	1 (A1.)	Gizonaren edo animalien gorputzaren goiko edo aurreko zatia, ahoa eta zentzumenak kokaturik daudena.
	buru	2 (A2.)	Burua adimenaren iturburutzat hartua.
	buru	3 (A3.)	Gauzaki baten mutur zabala edo biribila punta-ri aurrez aurre.
	buru	4 (A4.)	Buruzagia.
	buru	5 (A5.)	Buruxka.
	buru	6 (A6.)	b-, d-, g-, / p-, t-, k- aldaera.
	buru	7 (A7.)	Jendeaz, aziendaz mintzatuz, banakoa.
OK: ?			

➤ **Aditzak:** baztertu diren 41 kasuetatik, 22 aldiz genusa gaizki identifikatu da, eta gainontzeko 19etan adiera egokia ez zegoen hiztegian. Kasu gehienak ondorengo aditzekin osaturiko “kolokazio”-ak dira: *izan, eman, utzi, egin, ekarri*. Baztertu diren beste kasu batzuk hauek dira: *goratu, hartu, jarraitu, barreiatu* eta *jokatu*.

Jarraian aditzen hiztegi sarrerak aurkezten dira, alfabetikoki ordenatuak, eta ondoren bazterturiko kasuak:

- **barreiatu:**

barreiatu 1 (A1.) Bildua dagoena han-hemenka jarritz bereizi.
 barreiatu 2 (A2.) Desegin, suntsitu.
 barreiatu 3 (A3.) Ohartasuna edo buru bildutasuna galdu.

- **egin:**

egin 1 () [Laguntzailerik gabe ere jokatzeko da: Dagit, dagi, dagite; dagitzat, dagitza; negien, zegen, zegiten; zegitzan; banegi, balegi; balegitza; legike; legizke; begi, egizu, begite; begitza, egitzazu; degidan, degizkidan, begit, begizkit; degion, degizuen, degien, degieten; degit, degizkit, zegidan, zegizkidan; degio, degizu, degie, degiete; balegio, balegizkio, legioke, legizkioke; balegizu, balegie.
 egin 2 (A1.) Izaki bat gauzatu.
 egin 3 (A2.) Gauzatu.
 egin 4 (A3.) Apustu egin.
 egin 5 (A4.) Hitz egin.
 egin 6 (A5.) (ADIB.) Egin dezagun kontu: demagun, eman dezagun. Egin dezagun hil dutela mutil prestu bat. Egin dezagun kontu badakizula zuri dagokizuna.
 egin 7 (A6.) Ezkerrean ezartzen den aditzaren indargarri gisa, edo hura esaldiaren galdegaia dela adierazteko; delako aditza era burutuan ezartzen da, eta aspektuaren marka egin-ek hartzen du.

- **ekarri:**

ekarri 1 () [Laguntzailerik gabe ere jokatzeko da: dakart, dakar; dakartza; nakar; dakarkio; nekarren, zekarren; zekartzan; nindekarren; zekarkion; bekar; bekartza; bekarzio.
 ekarri 2 (A1.) Mintzatzen ari dena aurkitzen den lekura zerbait edo norbait eraman.
 ekarri 3 (A2.) Zerbait, adierazten den zerbaiten ondorio izan.
 ekarri 4 (A3.) Fruitu eman.
 ekarri 5 (A4.) Gorde, begiratu.

- **eman:**

eman 1 () [Laguntzailerik gabe ere jokatzeko da: demadan, emak/-n, bema, demagun, emazu, emazue, bemate; bemakio, emaiou; bematza; bemazkio.
 eman 2 (A1.) Norbait zerbaiten edukitzaile egin; zerbait norbaiten esku ezarri.
 eman 3 (A2.) Aditz izenkiaz.
 eman 4 (A3.) Izenondo gisa, objektua -(ri eta -ra deklinabide atzizkiaz). Zalea, atxikia.
 eman 5 (A4.) konpletiba batez: argudio edo gogoeta baten abiapuntutzat erabiltzen diren esapideak.

- **goratu:**

goratu 1 (A1.) Goretsi, goraiatu.
 goratu 2 (A2.) Gorago joan edo eraman; mailaz igo.
 goratu 3 (A3.) Harrapatu.

- **hartu:**

hartu 1 (A1.) Nork bere eskuan zerbait jarri, erabiltzeko, edukitzeko, hartaz jabetzeko.
 hartu 2 (A2.) Nork zerbait bereganatu edo eskuratu.
 hartu 3 (A3.) Nork bere gorputzean barneratu.
 hartu 4 (A4.) Zerbaiten jabe edo edukitzaile bihurtu, berezk. bortxaz.
 hartu 5 (A5.) Zerbaiti lotu erabiltzeko.

- **izan:**

izan 1 () da eta du.
 izan 4 (MA.) [Laguntzailerik gabe, edo beste aditzen laguntzaile gisa ere jokatzeko da: naiz, haiz, da, gara, zara, zarete, dira; nintzen, zen, ginen, ziren; banintz, balitz, balira; naizateke, dateke; nintzatekeen, zatekeen; nintzateke, litzateke; banadi, badadi, bagaitz; nadin, dadin, gaitezen, ditezen; nendin, zedin, gintezen, zitezen; banendi, baledi; naiteke, daiteke, daitezke; nintekeen, zitekeen, gintezekeen, zitezkeen;

ninteko, liteke; zait, zaio, zaie; zaizkit; natzaio; gatzazkio; zitzaidan, zitzaion; nintzaion; gintzaizkion; balitzaio; litzaioke; dakion; nakion; gakizkion; zekion; nenkion; balekio; lekioko.

izan	5 (MA.A1.)	Izatasunaren jabe gertatu.
izan	6 (MA.A2.)	Lotura aditza, subjektua eta predikatua elkartzuz.
izan	7 (MA.A3.)	Adizlagun batez edo deklinabide kasu bat hartzen duen izen sintagma batez.
izan	8 (MA.A4.)	Aditz iragangaitzen laguntzaile gisa.
izan	9 (MB.)	X ea.. [Laguntzailerik gabe, edo beste aditzen laguntzaile gisa ere jokatzeko da: dut, duk/n, du, dugu, duzu, duzue, dute; ditu; nau; gaitu; nuen, zuen, genuen zuten; zituen; ninduen; gintuen; banindu; banu, balu; duket, duke; nauke; gaituzke; nindukeen; dezadan, dezan; ditzan; nazan; gaitzan; nukkeen, zukeen; luke, lituzke; nezan, zezan; nintzan; gintzan; baneza, baleza; dezaket, dezake; nezakeen, zezakeen; nezake, lezake; diot, dio; dit; digu; nion, zion; zidan; zigun; banio, balio; balit, baligu; dizkio; zizkion; dioke; ziokeen; lioke; dizkioke; diezaion; ziezaion; baliezaio; diezaioko.
izan	10 (MB.A1.)	Adierazten denaren jabetasuna edo gozamina izan.
izan	11 (MB.A2.)	Nork beregan izan.
izan	12 (MB.A3.)	Aditz iragankorren laguntzaile gisa.

- jarraitu:

jarraitu	1 ()	Dena dela, adizkiok, darraio eta darraie izan ezik ez dira gaur egun ia erabiltzen.
jarraitu	2 (A1.)	Norbaiten edo zerbaiten ondotik joan edo ibili.
jarraitu	3 (A2.)	Beste zerbaiten ondotik etorri edo gertatu.
jarraitu	4 (A3.)	Egoera edo jarrera jakin batean irau.
jarraitu	5 (A4.)	Norbait edo zerbait gidarizat, arautzat, norbaiten iritziak edo erakutsiak onartu.
jarraitu	6 (A5.)	Era burutua izenlagun gisa. Etenik edo etenaldirik ez duena.

- jokatu:

jokatu	1 ()	du.
jokatu	2 (A1.)	Joko jakin batean aritu; jokoa jardun.
jokatu	3 (A2.)	da batez ere ipar. edo du.
jokatu	4 (A3.)	Izen baten deklinabide erak, eta batez ere aditz baten adizkiak, hurrenez hurren esan edo idatzi.
jokatu	5 (A4.)	Abere arrak emea estali.
jokatu	6 (A5.)	Antzerkiaz eta mintzatuz, teatro idazlanean kontaktzen dena egiaz gertatuko balitz bezala eman, antzezlarien bidez irudikatu.
jokatu	7 (A6.)	: hartaz irri egin.

- utzi:

utzi	1 (A1.)	Nolabait loturik gauden edo dagokigun norbait edo zerbaitengandik bereizi.
utzi	2 (A2.)	(ADIB.) Obra mardula utzi digu Orixek, eta bizitz eredu beteagoa. Bere azken finean aztore haietarik bata utzi zuela, zor bat baitzuen, hura pagatzeko.
utzi	3 (A3.)	Aurrean dena, eskaintzen dena ez hartu.
utzi	4 (A4.)	Irauten duen seinale bat egin.
utzi	5 (A5.)	Ez eragotzi, debekurik ez jarri.

2: abusatatu	GENUS	Gehiegi egin edo erabili.	OK:?
4: airatu	GENUS	Hegaldatu, hegan egin .	OK:X
19: asistitu	SIN	Etorri, joan, izan .	OK:?
34: behartu	GENUS	"Zerbaiten" (ipar., "zerbaitez"). Behar izan .	OK:?
69: entzueratu	izan	GENUS Aditzea izan , entzutea izan.	OK:?
74: erausi	GENUS	Zaunka, erausi egin .	OK:X
91: eskuratu_edo_eskura	GENUS	Eskura ekarri , menderatu.	OK:?
115: gogor_egin	GENUS	Aldapa egin , zail gertatu.	OK:X
130: harritu	GENUS	Harri eta zur eginik, zer esan edo egin jakin gabe gelditu edo utzi .	OK:?
132: haurdundu	GENUS	Haurdun gelditu edo utzi . lk.	OK:?
143: hizketatu	GENUS	Norbaitekin hizketaldi bat izan .	OK:?
148: hustu	GENUS	Ontzi bati edo zerbait daukan zerbaiti barruan duena atereaz, hutsik utzi edo gelditu.	OK:?
158: iruzurtu	GENUS	Iruzur egin .	OK:X
166: jaiki	SIN	Altxatu; goratu .	OK:?
168: jarri	GENUS	Paperean eman , idatzi.	OK:?
172: jo	GENUS	Kopuruei buruz, zenbatekoa, adierazten dena edo antzekoa dela uste izan .	OK:X
187: larrapastatu	GENUS	Larrapast egin , labaindu.	OK:X
190: larritu	GENUS	Larritasunak hartu , egoera larrian jarri.	OK:?
220: neurritu	GENUS	Neurrira etorri, neurrira ekarri .	OK:?
224: oinazetu	GENUS	Oinaze, estira eman edo hartu.	OK:?
227: ondoeztu	GENUS	Ondoeza egin , ondoezik jarri, alditxartu.	OK:X
244: publikatu	GENUS	Zabaldu, hedatu barreiatu .	OK:X
256: tiranizatu	GENUS	Tiraniaz agindu; tiraniaz jokatu .	OK:2?
259: tokikotu	GENUS	Tokiko egin .	OK:X
266: tuztatu	GENUS	Tu egin , tuz zikindu.	OK:X
284: zil_izan	GENUS	Zilegi izan .	OK:X
300: azkendu	GENUS	Azkena eman , bukatu, amaitu.	OK:?
315: erreportxatu	GENUS	Aurpegira eman, erantzuki egin .	OK:X
316: erreserbatu	GENUS	Gorde, begiratu, bazterrerera eman .	OK:?

317: esetsi	SIN	Ihes doanari, atzemateko edo harrapatzeko, jarraitu .	OK:X
319: ezagutarazi	GENUS	Ezagutzera eman .	OK:?
324: gorputzu	GENUS	Gorputza edo gorputz izatasuna hartu edo eman .	OK:?
339: jira-biratu	GENUS	Jira-birak egin .	OK:X
340: keinatu	GENUS	Zerbait egingo delako keinua egin .	OK:X
344: makatu	GENUS	Mailatu, zerbaiti edo norbaiti, kolpeak emanaz, sakonguneak egin .	OK:X
366: zeratu	GENUS	Zera izan, zera egin .	OK:X
373: aihertu	GENUS	"Norbaiti, zerbaiti". Kontra egin .	OK:X
379: erretratatu	GENUS	Erretratua egin .	OK:X
384: nahi_izan	GENUS	Zerbait egiteko, eginarazteko, lortzeko... asmo edo xede erabakia izan .	OK:X
390: zaunkatu	GENUS	Zaunka egin .	OK:X
397: munta_izan	GENUS	Axola izan .	OK:X

➤ **Adjektiboak: 6 kasu soilik baztertzen dira, guztiak sinonimoa gaizki identifikatu delako.**

15: basto	SIN	Arrunta, fina ez dena.	
	arrunt	1 (H1.A1.) Ohikoa, inolako berezitasunik ez duena.	
	arrunt	2 (H1.A2.) Nolakotasun apalekoa.	
OK: X			
56: laru	SIN	Laruena, larumina.	
	laru	1 (A1.) Hori ahula.	
	laru	2 (A2.) Laruena, larumina.	
OK: X			
140: gozoso	SIN	Gozozkoa; gozatsua.	
	gozo	1 (MA.A1.) Azukrearen edo eztiaren ahogozoa duena mikatz edo gazi-ren aurrez aurre.	
	gozo	2 (MA.A2.) Zentzumenei eta berezk. dastamenari atsegin edo gogoko zaiona; atsegina sortzen duena.	
	gozo	3 (MA.A3.) (ADIB.) Lur gozoa: emankorra edo lantzeko ona. Lur hain gozoa eta ugaria, non esnea eta ezitia darion.	
OK: X			
254: beheren	SIN	Apalena, mailarik apalenekoa.	
	apal	2 (H2.A1.) Garaitasun gabea, beheera.	
	apal	3 (H2.A2.) Bere burua gutxiesten eta beheratzen duena.	
OK: X			
260: erromaniko	SIN	Erromarra.	
	erromatar	1 (A1.) Antzinako Erromari dagokiona.	
	erromatar	2 (A2.) (ADIB.) Eliza Katoliko eta Erromatarra.	
	erromatar	3 (A3.) Erromari dagokiona; Erromako biztanlea.	
OK: X			
308: hoberen	SIN	Onena.	
	on	1 (H1.MA.A1.) Bere egitekoa ongi betetzen duena; espero den bezain egokia edo onuragarria dena.	
	on	2 (H1.MA.A2.) Zintzoa; moralak eskatzen duena betetzen duena.	
	on	3 (H1.MA.A3.) Ontasuna adierazten duena.	
OK: X			

B. Eranskina: Algoritmoen parametroen doiketa

Doitas./Estal.	Bazterte-hitzak	Lem./Hitz.
%85.34/%32.40	Baztertu2	Lem
%84.48/%32.40	Baztertu2	Lemhitz
%80.57/%38.82	Baztertu	Lemhitz
%80.14/%39.38	Baztertu	Lem
%77.36/%53.07	Ez-Baztertu	Lemhitz
%77.24/%52.79	Ez-Baztertu	Lem
%76.71/%40.78	Baztertu2	Hitz
%72.45/%46.64	Baztertu	Hitz
%69.15/%59.77	Ez-Baztertu	Hitz

B1 taula: “Teilakapena” algoritmoaren emaitzak parametrizazio desberdinekin

Doitas./Estal.	Bazterte-hitzak	Lem./Hitz.	Agerk. Def./Adib.	Agerk. Lem./Hitz.	Agerk. Ar./Mi.
%74.22/%98.60	Ez-Baztertu	lem	agerk_def	Agerk_lem	ar
%73.95/%93.29	Ez-Baztertu	hitz	agerk_def	Agerk_hitz	mi
%73.46/%100	Ez-Baztertu	lemhitz	agerk_def	Agerk_lem	mi
%73.18/%100	Ez-Baztertu	lemhitz	agerk_adib	Agerk_hitz	mi
%73.08/%98.60	Baztertu	lem	agerk_def	Agerk_lem	ar
%72.99/%86.87	Ez-Baztertu	hitz	agerk_adib	Agerk_hitz	ar
%72.90/%86.59	Baztertu	hitz	agerk_adib	agerk_hitz	ar
%72.75/%96.36	Ez-Baztertu	hitz	agerk_adib	agerk_hitz	mi
%72.62/%100	Baztertu	lemhitz	agerk_def	agerk_lem	mi
%72.62/%100	Baztertu	lem	agerk_def	agerk_lem	mi
%72.62/%100	Ez-Baztertu	lem	agerk_def	agerk_lem	mi
%72.54/%99.72	Baztertu	lemhitz	agerk_adib	agerk_hitz	ar
%72.52/%98.60	Ez-Baztertu	lemhitz	agerk_def	agerk_lem	ar
%72.34/%100	Ez-Baztertu	lemhitz	agerk_def	agerk_hitz	mi
%72.20/%92.45	Baztertu	hitz	agerk_def	agerk_hitz	mi
%72.06/%100	Baztertu	lemhitz	agerk_adib	agerk_hitz	mi
%71.98/%99.72	Ez-Baztertu	lemhitz	agerk_adib	agerk_hitz	ar
%71.59/%96.36	Baztertu2	lem	agerk_def	agerk_lem	ar
%71.59/%96.36	Baztertu	hitz	agerk_adib	agerk_hitz	mi
%71.50/%100	Baztertu	lemhitz	agerk_def	agerk_hitz	mi
%70/%92.17	Ez-Baztertu	lemhitz	agerk_def	agerk_hitz	ar
%70.94/%100	Baztertu2	lem	agerk_def	agerk_lem	mi
%70.94/%100	Ez-Baztertu	lem	agerk_adib	agerk_lem	ar
%70.58/%85.47	Baztertu2	hitz	agerk_adib	agerk_hitz	ar
%70.39/%100	Ez-Baztertu	lemhitz	agerk_adib	agerk_lem	ar
%70.39/%100	Ez-Baztertu	lem	agerk_adib	agerk_lem	mi
%70.15/%90.78	Baztertu2	hitz	agerk_def	agerk_hitz	mi
%70.12/%91.62	Baztertu	lemhitz	agerk_def	agerk_hitz	ar
%70.11/%100	Baztertu2	lemhitz	agerk_def	agerk_lem	mi
%70.11/%100	Baztertu	lem	agerk_adib	agerk_lem	ar
%69.94/%96.64	Baztertu2	lemhitz	agerk_def	agerk_lem	ar
%69.68/%98.60	Baztertu	lemhitz	agerk_def	agerk_lem	ar
%69.55/%100	Ez-Baztertu	lemhitz	agerk_adib	agerk_lem	mi
%69.29/%95.53	Baztertu2	hitz	agerk_adib	agerk_hitz	mi
%69.16/%70.67	Baztertu2	hitz	agerk_def	agerk_hitz	ar
%68.82/%99.44	Baztertu2	lemhitz	agerk_adib	agerk_hitz	ar
%68.43/%100	Baztertu	lemhitz	agerk_adib	agerk_lem	ar
%68.43/%100	Baztertu	lem	agerk_adib	agerk_lem	mi
%68.15/%100	Baztertu2	lemhitz	agerk_adib	agerk_lem	mi
%67.87/%100	Baztertu2	lemhitz	agerk_adib	agerk_hitz	mi
%67.87/%100	Baztertu	lemhitz	agerk_adib	agerk_lem	mi
%67.59/%100	Baztertu2	lemhitz	agerk_def	agerk_hitz	mi
%67.41/%86.59	Baztertu2	lemhitz	agerk_def	agerk_hitz	ar
%67.31/%100	Baztertu2	lem	agerk_adib	agerk_lem	mi
%66.75/%100	Baztertu2	lemhitz	agerk_adib	agerk_lem	ar
%66.75/%100	Baztertu2	lem	agerk_adib	agerk_lem	ar
%66.17/%75.97	Ez-Baztertu	hitz	agerk_def	agerk_hitz	ar
%64.57/%75.69	Baztertu	hitz	agerk_def	agerk_hitz	ar

B2 taula: “Agerkidetza-simplea” algoritmoaren emaitzak parametrizazio desberdinekin

Doitas./Estal.	Bazterte-hitzak	Lem./Hitz.	Agerk. Def./Adib.	Agerk. Lem./Hitz.	Agerk. Ar./Mi.
%72.06/%100	Baztertu2	lemhitz	agerk_def	agerk_lem	mi
%72.06/%100	Baztertu2	lem	agerk_adib	Agerk_lem	mi
%71.22/%100	Baztertu2	lem	agerk_def	Agerk_lem	mi
%70.94/%100	Baztertu2	lem	agerk_def	Agerk_lem	ar
%70.94/%100	Baztertu2	lem	agerk_adib	Agerk_lem	ar
%70.39/%100	Ez-Baztertu	hitz	agerk_adib	Agerk_hitz	mi
%69.55/%100	Baztertu2	lemhitz	agerk_adib	Agerk_lem	mi
%68.99/%100	Baztertu	hitz	agerk_adib	Agerk_hitz	mi
%68.99/%100	Baztertu	lemhitz	agerk_adib	Agerk_hitz	mi
%68.43/%100	Baztertu2	lemhitz	agerk_def	Agerk_lem	ar
%68.15/%100	Baztertu2	lemhitz	agerk_def	Agerk_hitz	mi
%68.02/%96.08	Baztertu	hitz	agerk_adib	Agerk_hitz	ar
%68.02/%96.08	Ez-Baztertu	hitz	agerk_adib	Agerk_hitz	ar
%67.87/%100	Ez-Baztertu	lemhitz	agerk_adib	Agerk_lem	mi
%67.59/%100	Baztertu	lemhitz	agerk_def	Agerk_lem	mi
%67.31/%100	Baztertu	lemhitz	agerk_adib	Agerk_lem	mi
%67.03/%100	Baztertu	lemhitz	agerk_def	Agerk_hitz	mi
%66.75/%100	Ez-Baztertu	lemhitz	agerk_adib	Agerk_hitz	mi
%66.48/%100	Baztertu2	lemhitz	agerk_adib	Agerk_lem	ar
%65.92/%100	Baztertu	lemhitz	agerk_def	Agerk_lem	ar
%65.92/%100	Baztertu	lem	agerk_def	Agerk_lem	mi
%65.64/%100	Ez-Baztertu	lemhitz	agerk_def	Agerk_hitz	mi
%65.36/%100	Baztertu	lem	agerk_adib	Agerk_lem	ar
%65.08/%100	Baztertu	lem	agerk_adib	Agerk_lem	mi
%65.08/%100	Ez-Baztertu	lemhitz	agerk_def	Agerk_lem	mi
%65.08/%100	Ez-Baztertu	lem	agerk_adib	Agerk_lem	mi
%64/%90.78	Ez-Baztertu	hitz	agerk_def	Agerk_hitz	ar
%64.80/%100	Baztertu2	lemhitz	agerk_adib	Agerk_hitz	mi
%64.80/%100	Baztertu	lemhitz	agerk_adib	agerk_hitz	ar
%64.80/%100	Baztertu	lem	agerk_def	agerk_lem	ar
%64.80/%100	Ez-Baztertu	lem	agerk_def	agerk_lem	mi
%64.52/%100	Baztertu	lemhitz	agerk_def	agerk_hitz	ar
%64.50/%90.50	Baztertu	hitz	agerk_def	agerk_hitz	ar
%64.40/%98.88	Baztertu	hitz	agerk_def	agerk_hitz	mi
%64.24/%100	Baztertu	lemhitz	agerk_adib	agerk_lem	ar
%64.24/%96.08	Baztertu2	hitz	agerk_adib	agerk_hitz	ar
%63.55/%98.88	Ez-Baztertu	hitz	agerk_def	agerk_hitz	mi
%63.12/%100	Ez-Baztertu	lemhitz	agerk_def	agerk_lem	ar
%63.12/%100	Ez-Baztertu	lemhitz	agerk_adib	agerk_lem	ar
%62.57/%88.82	Baztertu2	hitz	agerk_def	agerk_hitz	ar
%62.56/%100	Baztertu2	lemhitz	agerk_def	agerk_hitz	ar
%62.29/%100	Baztertu2	lemhitz	agerk_adib	agerk_hitz	ar
%62.01/%100	Baztertu2	hitz	agerk_adib	agerk_hitz	mi
%62.01/%100	Ez-Baztertu	lem	agerk_def	agerk_lem	ar
%61.73/%100	Ez-Baztertu	lem	agerk_adib	agerk_lem	ar
%61.45/%100	Ez-Baztertu	lemhitz	agerk_def	agerk_hitz	ar
%59.21/%100	Ez-Baztertu	lemhitz	agerk_adib	agerk_hitz	ar
%59.20/%98.60	Baztertu2	hitz	agerk_def	agerk_hitz	mi

B3 taula: "Agerkidetza-bektorea" algoritmoaren emaitzak parametrizazio desberdinekin

Doitas./Estal.	Bazterte-hitzak
%70.62/%98.88	Baztertu
%70.14/%99.16	Baztertu2
%69.94/%99.44	Ez-Baztertu

B4 taula: "Yarowsky 1992" algoritmoaren emaitzak parametrizazio desberdinekin

C. Eranskina: Izenen sinonimo multzoak

Synset-aren Tamaina	Synset Kopurua
3245	1
57	1
56	1
41	1
32	1
30	1
25	2
24	3
23	2
22	1
20	1
19	1
18	5
17	7
16	1
15	4
14	2
13	12
12	11
11	21
10	18
9	20
8	31
7	43
6	75
5	109
4	219
3	459
2	1210
1	7505

C1 taula: lehen hurbilpena jarraituz sorturiko multzoak.

Synset-aren Tamaina	Synset Kopurua
37	1
34	1
32	1
30	1
29	2
28	3
26	1
25	3
24	2
23	2
22	3
20	3
19	2
18	5
17	4
16	3
15	10
14	3
13	6
12	14
11	20
10	22
9	39
8	31
7	68
6	105
5	125
4	267
3	493
2	1179
1	9800

C2 taula: 2. hurbilpena jarraituz sorturiko multzoak.