

Izen+aditz konbinazioen azterketa elebiduna, hizkuntza-aplikazio aurreratuei begira

Bilingual analysis of noun+verb combinations and their functionality in advanced language applications

Uxoia Inurrieta, Itziar Aduriz¹, Arantza Díaz de Ilarraza, Gorka Labaka, Kepa Sarasola
IXA taldea, Euskal Herriko Unibertsitatea

usoa.inurrieta@ehu.es, a.diazdeilarraza|gorka.labaka|kepa.sarasola@ehu.es

(1) Department of Linguistics, University of Barcelona

itziar.aduriz@ub.edu

Laburpena

Hiztegi elebidunak oinarritzat hartuta, euskarazko eta gaztelaniazko izen+aditz konbinazioak izan ditugu aztergai lan honetan. Konbinazioen eta euren ordainen ezaugarri morfosintaktiko zein semantikoiei begiratu diegu, eta bi hizkuntzak parez pare jarri ditugu, zer alde eta antzekotasun duten aztertzeko. Artikulu honek agerian uzten du zeinen konplexuak diren era horretako egiturak eta, ondorioz, zeinen garrantzitsua den Hizkuntzaren Prozesamenduko aplikazioetan tratamendu egoki bat ematea, itzulpen automatikoan adibidez. Horrez gain, azterketatik lortutako emaitza guztiak interfaze publiko batean jarri ditugu, edonork bilaketak egin ahal izan dituzan guk landutako konbinazioen gainean.

Gako-hitzak

Hitz-konbinazioak, Hizkuntzaren Prozesamendua, Elhuyar hiztegiak

Abstract

This article deals with noun+verb combinations in bilingual Basque-Spanish and Spanish-Basque dictionaries. We take a look at morphosyntactic and semantic features of word combinations in both language directions, and compare them to identify differences and similarities. Our work reveals the high complexity of those constructions and, hence, the need to address them specifically in Natural Language Processing tools, for example in Machine Translation. All of our results are publicly available online, where users can query the combinations we have analysed.

Keywords

Word combinations, Natural Language Processing, Elhuyar dictionaries

1 Sarrera

Hizkuntza batean hitz-konbinazioak zuzen erabiltzeak agerian uzten du hiztunak hizkuntza hori zenbateraino ezagutzen duen, konbinazioek testua aberasten eta edertzen baitute gehienetan, eta hori egiteko gai denak derrigor ezagutu behar baitu ondo hizkuntza, gramatika-araueetatik eta oinarritzko hiztegitik haratago. Hizkuntza hori itzultzen edo tratamendu automatikorako gaitze-lanetan hasten garenean, ordea, aberastasun eta edertasun horiek arazo-iturri bihurtzen dira sarri.

Izan ere, era askotakoak izan badaitezke ere, hitz-konbinazio guztiak datoz bat zerbaitetan: normalean, ez dituzte hizkuntzaren ohiko arau morfosintaktiko zein semantikoak jarraitzen. Horregatik, hain zuzen ere, era horretako egiturek –izan kolokazio¹, lokuzio², edo enuntziatu fraseologiko³– buruhauste galantak sortzen dituzte hainbat alorretan: itzulpengintzatik hasi eta Hizkuntzaren Prozesamendura (HP), atzerriko hizkuntzen ikaskuntza ere tartean dela.

Lan honetan, izen+aditz konbinazio mota guztiak aztertu ditugu, momentuz tipologia kon-tuak alde batera utzita. Gure helburua euskararen eta beste hizkuntza batzuen artean zein alde eta antzekotasun dagoen ikustea da, gerora, hitz-konbinazioek HPko aplikazioetan sortzen dituzten arazoak konpontzen laguntzeko. Azterketa hau EHUKo IXA ikerketa-taldean egiten ari garenez, taldean bertan garatutako aplikazioak hobetzerantz bideratuko dugu gure ikerketa-lana, eta garrantzi berezia emango diegu gaztelaniatik euskararako Matxin⁴ itzultzaile automatikoari (ikus 3. atala).

Azterketa linguistikoa corpus paralelo eta alderagarrietan oinarritzeko asmoa dugu, baina,

¹haizea ibili; gogotik barre egin; euskaldun peto

²ahalak eta leherrak egin; adarra jo; aita ponteko

³nolako zura, halako ezpala; ez horregatik

⁴<http://www.opentrad.com/eu/>

horren aurretik, hiztegien gainean egin nahi izan dugu lehen urratsa. Horretarako, Elhuyarren gaztelania-euskara eta euskara-gaztelania hiztegietatik⁵ izen+aditz konbinazioak erauzi ditugu euren ordainekin batera, eta, bi hizkuntzak parez pare jarrita, hainbat ezaugarri begiratu diogu: morfologikoei, sintaktikoei eta semantikoei.

Hortaz, artikulu honetan, izen+aditz konbinazioak ardatz hartu eta euskararen eta gaztelaniaren bat-etortze mailari buruz arituko gara, tratamendu automatikoan zenbateraino lagun dezaketen ikusteko. Azterketaren emaitzen bidez, erakutsiko dugu ez dagoela kidetasun handirik bi hizkuntzen artean, egitura horiei dagokienez.

Gainera, landu ditugun konbinazio guztiak datu-base batean bildu ditugu informazio linguistikorekin batera, eta eskuragarri jarri ditugu interfaze publiko baten bidez (ikus 5. atala).

2 Aurrekariak

Hitz-konbinazioak, oro har, fraseologiaren alorrean kokatu ohi dira. Erabateko adostasunik ez dagoen arren, unitate fraseologikoak ohi baino sarriago elkarrekin agertzeko joera duten hitz-konbinaziotzat definitzen dira, eta hiru ezaugarri nagusi dituztela esaten da (Sanz Villar, 2011; Baldwin & Kim, 2010; Corpas Pastor, 2004): hitz batez baino gehiagoz osatuta daude, sintaktikoki finkoak dira, eta konbinazio osoaren esanahia ezin da osagaien esanahi indibidualetatik inferitu.

Gaiaren inguruan egindako lanak askotarikoak dira: azterketa teorikoek konbinazioak deskribatzen eta sailkatzen dituzte, idiomatikotasunaren zein ezaugarri morfosintaktikoen arabera (Rafel, 2004; Zabala, 2004; Corpas Pastor, 2001); beste lan batzuk, berriz, aplikazio jakinetara bideratuta egon ohi dira. Bigarren multzoan, azken hamarkadetan indar berezia hartu dute HPren alorrean egindako lanek (Seretan, 2013; Alonso Ramos, 1995), eta euskarak ere izan du bere lekua prozesu horretan.

Bi ikerketa-lan nagusi egin dira euskaraz: Urizarrena (2012) bata, eta Gurrutxagarena (2014) bestea. Lehenak euskarazko lokuzioen tratamendurako oinarri linguistikoak ematen ditu, eta bigarrenak, berriz, euskarazko aditz+izen konbinazioen detekzio eta karakterizazio automatikoa du aztergai. Ikuspegi elebidunetik, ordea, oraindik ez da ikerketarik egin, eta hutsune hori betetzera dator lan hau.

Izan ere, beste hizkuntzetan ere gehiago aztertu izan da konbinazioen detekzio eta erauz-

keta elebakarra, baina ikuspegi elebidunetik ere egin dira hainbat hurbilpen eta esperimendu. Lük eta Zhouk (2004), esaterako, corpus elebakarretatik erauzten dituzte kolokazioak, eta, gero, lexikoi elebidunak sortzeko, hizkuntza batean eta bestean lortutako emaitzak parekatzen dituzte. Champollion sistemak (Smadja, McKeown, & Hatzivassiloglou, 1996), berriz, corpus paraleleotatik erauzten ditu zuzenean ingelesezko eta frantsesezko kolokazioak eta ordainak.

Bestalde, Mel'cuk-en (1998) lana izan da, ziur asko, alor elebidunean egin diren hurbilpenen arteko ezagunena, non kolokazioak funtzio lexikalen bidez errepresentatzeko metodo bat proposatzen baita, kolokazioko osagaien artean dagoen erlazio semantikoaren arabera. Gerora, hainbat esperimendu egin dira hurbilpen horretan oinarrituta, kolokazioen inguruko informazioa itzultzaile automatikoetan gehitzeko (Heylen & Maxwell, 1994). Dena den, funtzio lexikaletan oinarritutakoez gain, badago itzulpen-prozesua beste transferentzia-metodo batzuen bidez egiten duen esperimentrik ere (Wehrli et al., 2009).

3 Itzulpen automatikora begira

Sarreraren esan bezala, ikerketa-lan honetatik aterako ondorioak HPko tresnak hobetzeko baliatzea da gure helburua. Horietako bat IXA taldean garatutako Matxin itzultzaile automatikoa da, gaztelaniatik euskarara itzultzen duena.

Sistemak erregela linguistikoak ditu oinarrian, eta itzulpen-prozesua hiru fasetan egiten du: analisia, transferentzia eta sorkuntza (Iñurrieta, 2013; Mayor et al., 2011). Lehen fasean, jatorrizko hizkuntzako itzulgaiaren analisi morfosintaktikoa egiten da. Bigarrenetan, jatorrizko hizkuntzaren egitura xede-hizkuntzara ekartzen da, eta, sistemaren lexikoi elebidunetik abiatuta, gaztelaniazko hitz bakoitzari euskarazko baliokide bat ematen zaio. Azkenik, hirugarren fasean, euskarazko arau sintaktiko eta morfologikoak aplikatu, eta itzulpena bera sortzen da.

Mayorren (2011) arabera, itzulpenen % 69 ulertzeko modukoak dira; beraz, esan daiteke nahiko emaitza onargarriak lortzen direla oro har. Hitz-konbinazioei dagokienez, ordea, sistema oraindik ez da itzulpen txukunak sortzeko gai, eta baliteke itzulpen okerrekin osatzen duten % 31 horretan konbinazioek garrantzia izatea. Izan ere, lexikoi elebidunean badaude hitz anitzeko sarrera batzuk –Elhuyar hiztegiako konbinazioak tarteko–, baina, batetik, ez dira asko; bestetik, ez dira beti ondo hautematen; eta, gainera, Matxinek hitz bakartzat hartzen ditu, ia informazio gehigarririk gabe.

⁵<http://hiztegiak.elhuyar.org/>

Konbinazio bat hitz anitzeko unitatetzat hauteman ahal izateko, osagai guztiek segidan eta ordena berean agertu behar dute esaldian, hala izan ezean, detekzioak huts egiten baitu (ikus 1. adibidea). Gainera, konbinazio osoari kategoria bakarra ematen zaio, eta ez konbinazio-ko osagai bakoitzari berea (ikus 2. adibidea). Izen+aditz konbinazioen kasuan, aditza lematizatu egiten da, baina izena eta izenarekin batera doazen determinatzaile eta preposizioak ez; bere horretan agertu behar dute, aldaketarik txikiena ere egin gabe.

Hori gutxi balitz, euskarazko sorkuntzarako ere ez dago hitz anitzekoentzako arau berezirik, eta, oro har, segida bat beste segida baten bidez ordeztzen da, bestelako informaziorik kontuan izan gabe (ikus 3. eta 4. adibideak).

Hori guztia aintzat hartuta, ez da harritze-koa emaitza traketsak lortzea nahiz eta itzulgaiak simple samarrak izan. Hona hemen adibide batzuk:

- (1) Ayer le **gastaron una broma**.
Atzo **broma egin** zioten.

La **broma** que le **gastaron** fue cruel.
***Gastatu** zioten **broma** krudela izan zen.

Lehen esaldia ondo itzuli da, *gastar una broma* konbinazioa jarraian agertzen delako osorik; bigarrenean, ordea, *broma* eta *gastar* ez daude bata bestearen ondoan, ez eta ordena berean ere, eta sistemak ez du konbinazioa hitz anitzeko unitatetzat hartu. Hori dela eta, euskaraz okerreko aditza aukeratu da: *gastatu*, *eginen* ordeztzen (*Egin zioten broma krudela izan zen*).

- (2) El colesterol **hace mal** al corazón.
Kolesterolak bihotzari **kalte egiten** dio.

Ha **hecho mal** el examen.
Azterketa ***kalte egin** du.

Bigarren adibide honetan, lehenengoaren kontrakoa gertatu da. *Hacer* eta *mal* jarraian ikusita, Matxinek unitate bakartzat hartu ditu bi hitzak, analisi sintaktikoari jaramonik egin gabe. Lehen esaldian ez dago arazorik, baina bai bigarrenean, beharrezkoa baitzen sistemak *mal* izentzat ez hartzea, baizik eta aditzondotzat, *kalteren* ordeztzen *gaizki* jarri ahal izateko (*Azterketa gaizki egin du*).

- (3) ¿**Tienes frío?**
Hotz zara?

Miren **tiene frío** el plato.
Mirenek platera **hotz *da**.

Hirugarren adibidean ikus daitekeenez, hitz anitzeko bat tartean dagoenean, sistema ez da gai subjektuaren eta aditzaren arteko konmuztadurarik egiteko. Horregatik, lehen esaldiaren itzulpena egokia bada ere, bigarrenarena ez: aditzak *ukan* motakoa behar luke *izan* motakoa beharrean (*Mirenek platera hotz du*).

- (4) Le **hizo daño**.
Min eman zion.

Se **hizo daño**.
Min *eman zen.

Hirugarrenean bezala, azken adibidean ere sorkuntzarako arau faltan datza arazoa. Matxinen lexikoi elebidunean, *hacer daño*ren baliokidetzat *min eman* bakarrik ageri da, eta, hori dela eta, sistemak ez du bereizten *min eman* eta *min harturen* artean. Informazio gehigarria behar du, *hacerse daño* bihurkaria denean *hartu* aditza erabili behar dela jakiteko eta dagokion aditza jartzeko (*Min hartu zuen*).

4 Hiztegien gaineko azterketa

Aipatu bezala, euskarazko eta gaztelaniazko hitz-konbinazioak aztertzeke eta alderatzeko ikerketan honetan, lehen urratsa hiztegien gainean egin dugu. Gure azterketa corpusetan oinarritu nahi badugu ere, egokia iruditu zaigu abiapuntutzat hiztegi elebidunak hartzea, konbinazio usuenak behintzat jasotzen baitituzte normalean.

Jakin badakigu hiztegiaren biltzen diren konbinazioak tipologia bakarrekoak izan ohi direla –lokuzioak⁶, oro har (Urizar, 2012)– eta beste mota batzuetako konbinazioak ez direla hainbeste landu izan hiztegiaren –kolokazioak⁷, kasu–. Dena den, esan bezala, pauso honetan hasitako lana osatzeko, corpusetara joko dugu geroa.

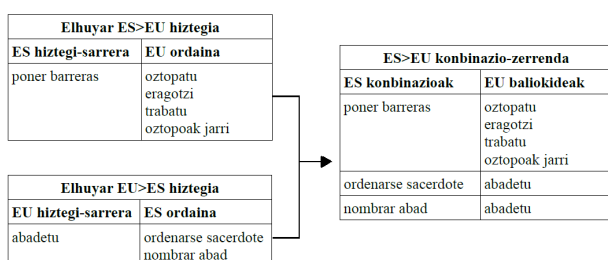
Oinarritzat hartu ditugun konbinazio-zerrendak Elhuyarren gaztelania-euskara eta euskara-gaztelania hiztegietatik erauzi ditugu. Sarrera guztiak hartu, eta Eustagger analizatzaile morfologikoaz (Aduriz et al., 1996) baliatu

⁶ *adarra jo, lan egin* edo *min eman* bezalakoak

⁷ *legeak urratu* edo *konpromisoa berretsiren* gisakoak

gara izen+aditz konbinazioak izan litezkeenak hautemateko. Prozesu automatiko gehienetan bezala, emaitza oker batzuk ere lortu ditugu benetako izen+aditz konbinazioekin batera, eta zerrenda eskuz gainbegiratu behar izan dugu.

Ahalik eta kasu gehien landu nahi genituzenez, hiztegi-sarrera diren izen+aditz konbinazioez gain, beste egitura batzuetako sarrerren ordaintzat ageri diren izen+aditz konbinazioak ere hartu ditugu. Hau da: gaztelaniazko *poner barreras* konbinazioa eta haren *oztopatu* ordainaz gainera, euskarazko *abadetu* sarrera eta haren ordaintzat ageri den *ordenarse sacerdote* ere erauzi ditugu, ordaina izen+aditz motakoa delako (ikus 1. irudia).



1. irudia: Konbinazio-zerrendak sortzeko prozesua.

Gaztelaniazko 2.650 konbinazio landu ditugu: gaztelania-euskara hiztegitik erauzitako 1.390 eta euskara-gaztelania hiztegitik erauzitako 1.260. Konbinazio gehienek ordain bat baino gehiago dutenez, euskarazko balioakideak askoz ere gehiago izan dira: 6.587.

Euskarazko konbinazioei dagokienez, berriaz, 2.954ko zerrenda osatu dugu, euskara-gaztelania hiztegiko 1.354 eta gaztelania-euskara hiztegiko 1.600 batuta. Horien gaztelaniazko balioakideak 6.392 dira guztira.

Datozen azpiataletan bildu ditugu horien guztien azterketatik ateratako emaitzak. Lehenik, 4.1 atalean, euskarazko eta gaztelaniazko konbinazioetako aditzak eta izenak nolakoak diren azalduko dugu, bi hizkuntzetako emaitzak parez pare jarrita. Bigarrenik, 4.2 atalean, beselako ezaugarri morfosintaktiko batzuk izango ditugu hizpide: zein kasu eta postposizio-marka hartzen dituzten euskarazko izenek, zein kategoriatako hitzez osatzen diren gaztelaniazko konbinazioak, zer gertatzen den egitura horiekin hizkuntza batetik bestera, eta zer portaera duten mugatasunak eta numeroak horrelako egituretan. Azkenik, 4.3 atalean, ezaugarri semantikoei begiratuko diegu, eta hizkuntza bateko eta besteko aditzak eta izenak elkarren artean balioakide ote diren ikusiko dugu.

4.1 Konbinazioetako aditzak eta izenak

Azterketa sakonagorik egin aurretik, gure lanaren oinarri diren zerrendak hartu, eta osagai nagusiei begiratu diegu: aditzei eta izenei. Hizkuntza bateko eta besteko konbinazioak hartuta, zein aditz/izen eta zenbateko maiztasunez erabili diren jakin nahi izan dugu, gero bi hizkuntzak parez pare jarri eta euren artean antzekotasunik ba ote dagoen ikusteko.

Fraseologiari buruzko lanetan maiz aipatu izanenez, hitz-konbinazioetan erabiltzen diren aditzak oso ohikoak izaten dira gehienetan, aditz "arinak", esanahi osorik gabeak. Ez da harritzekoa, beraz, zer aurkitu dugun: hizkuntza bateko eta besteko aditzik erabilienean artean, gehienak oso arruntak dira, eta, gainera, asko balioakideak dira euren artean. Hain zuzen ere, euskaraz gehien errepikatzen diren sei aditzen ordainak gaztelaniazko aditzetan bilatuta, zortzi aditzik erabilienean artean aurkitu ditugu: *egin* – *hacer*, *izan* – *ser/estar/tener*, *eman* – *dar*, *hartu* – *tomar*, *egon* – *estar*, *jarri* – *poner* (ikus 1. eta 2. taulak).

Aditza	Agerr.	Ehunekoa
egin	614	% 20,79
izan	296	% 10,02
eman	217	% 7,35
hartu	147	% 4,98
egon	133	% 4,50
jarri	92	% 3,11
Guztira	1.499	% 50,75

1. taula: Euskarazko konbinazioetan agerraldi gehien dituzten aditzak.

Aditza	Agerr.	Ehunekoa
hacer	233	% 8,79
dar	197	% 7,43
estar	122	% 4,60
poner	112	% 4,23
tener	103	% 3,89
echar	73	% 2,75
tomar	64	% 2,42
ser	56	% 2,11
Guztira	960	% 36,22

2. taula: Gaztelaniazko konbinazioetan agerraldi gehien dituzten aditzak.

Hala ere, agerraldiei begiratu gero, bat aise konturatzen da euskaraz askoz ere alde handiagoa dagoela gehien errepikatzen diren aditzen eta gaintzekoen artean. Izan ere, euskarazko konbinazioetan 310 aditz desberdin agertu dira guztira, baina konbinazio guztien % 50,75 sortzen dira 1. taulan bildutako seiekin bakarrik.

Gaztelaniaz ere badago aldea, baina ez da inondik inora ere euskarazkoa bezain handia; zor-

tzi aditzik erabilienean konbinazioen % 36,23 osatzen dituzte. Kontuan hartzekoa da, dena den, agertu diren aditzak ere gehiago direla: 453.

Bestalde, izenen agerraldiak ere kontatu ditugu, antzeko ondorioz atera ote genezakeen ikus-teko. Bagenekien izen desberdinen kopurua aditzena baino askoz ere altuagoa izango zela, baina, era berean, susmoa genuen izen batzuek beste batzuek baino joera handiagoa dutela konbinazioetan agertzeko. Eta ez genbiltzan oker, 3. eta 4. taulek agerian uzten dutenez.

Izena	Agerr.	Ehunekoa
buru	65	% 2,20
begi	54	% 1,83
aurre	43	% 1,46
kontu	33	% 1,12
atze	24	% 0,81
bide	23	% 0,78
esku	21	% 0,71
Guztira	263	% 8,91

3. taula: Euskarazko konbinazioetako izenik erabiliak.

Izena	Agerr.	Ehunekoa
mano	28	% 1,06
cabeza	27	% 1,02
ojo	21	% 0,79
vista	21	% 0,79
oído	18	% 0,68
cuenta	18	% 0,68
Guztira	133	% 5,02

4. taula: Gaztelaniazko konbinazioetako izenik erabiliak.

Hortaz, izenetan ere badago antzekotasunik bi hizkuntzen artean, gehien-gehien errepikatzen direnetako batzuk bat baitatuz euskaraz eta gaztelaniaz: *buru* – *cabeza*, *begi* – *ojo*, *esku* – *mano*, *kontu* – *cuenta*. Horieta asko, gainera, gorputz-atalen izenak dira, multzo horretako hitzak oso ohikoak baitira hitz-konbinazioetan, lokuzioetan batez ere.

4.2 Morfosintaxia

Behin konbinazioetan zein aditz eta izen agertzen diren ikusita, konbinazioen beste ezaugarri morfosintaktiko batzuk aztertzei ekin diogu. Hasteko, hizkuntza bateko eta besteko konbinazioak bereiz landu ditugu: euskarazko izenen kasu- eta postposizio-markak, eta gaztelaniazko osagaien kategoriak. Bigarrenik, bi hizkuntzetako egiturak parez pare jarri ditugu, kasu- eta postposizio-marken eta gaztelaniazko egituren artean loturarik ba ote dagoen argitu nahian. Eta, azkenik, mugatasunari eta numeroari ere begiratu diegu,

ezaugarri horiek hizkuntza batetik bestera zenbateraino gordetzen diren jakiteko.

4.2.1 Euskarazko konbinazioen ezaugarri morfologikoak: kasu- eta postposizio-markak

Lehenik, euskarazko konbinazioetako izenak hartu, eta etiketa bana jarri diegu eskuz, kasu- edo postposizio-markaren arabera. Kontaketak egitean, berehala konturatu gara alde nabarmena dagoela etiketa bateko edo besteko multzoen artean.

Urizarren (2012) eta Zabalaren (2004) lanetan aipatzen denez, izen batez eta aditz batez osatzen diren konbinazio gehien-gehienek absolutibo-marka izan ohi dute. Horixe bera ondorioztatu dugu guk ere, konbinazio guztien hiru laurden baino gehiago baitira, hain zuzen ere, multzo horretakoak: *denbora galdu*, *dei egin*, *gobada jo*, *itzal egin*, *hitza bete...*

Inesiboa daramaten izenak ere nahiko sarri ageri dira besteen aldean (*sutan egon*, *jokoan jarri*), eta adlatibodunek (*aurrera egin*, *eskura ordaindu*), instrumentaldunek (*eskuz jo*, *aurrez prestatu*) eta ablatibodunek (*burutik egon*, *hutsetik hasi*) jarraitzen diete.

Gehien errepikatu diren bost markak 5. taulan ikus daitezke, agerpen-kopuruaren eta ehunekoa-ren arabera.

Marka	Agerr.	Ehunekoa
absolutiboa (abs)	2250	% 76,18
inesiboa (ine)	364	% 12,32
adlatiboa (ala)	101	% 3,42
instrumentala (ins)	87	% 2,94
ablatiboa (abl)	85	% 2,88
beste batzuk	62	% 2,27

5. taula: Euskarazko konbinazioen markak.

Gainontzeko markak nahiko bakanak dira, eta ez dira konbinazio guztien % 1era ere heltzen: ergatiboa (*deabruak hartu*), datiboa (*amuari lotu*), adlatiboa + abutiboa (*leporaino egon*), soziatiboa (*suarekin jolastu*), genitiboa + absolutiboa (*gorrarena egin*), lekuzko genitiboa + absolutiboa (*hitzekoa izan*), adlatiboa + lekuzko genitiboa (*bururako gorde*), eta lekuzko genitiboa (*zorriarekin egon*).

4.2.2 Gaztelaniazko konbinazioen ezaugarri morfosintaktikoak: osagaien kategoriak

Gaztelaniari dagokionez, lau egituratako konbinazioak bildu ditugu. Batzuk izen eta aditz bakarrak osatuak dira, eta beste batzuek determinatzaile eta/edo preposizioak dituzte tartean.

Azpiko taulan ikusten denez, gehien errepi-
katzen den egitura *aditza + determinatzailea + izena* da (*dar un toque, ser una pena, hacer un favor*), eta beste hiru multzoen artean ez dago alde handirik: *aditza + izena* (*meter baza, tener afecto*), *aditza + preposizioa + izena* (*saber de memoria, tener a favor*), *aditza + preposizioa + determinatzailea + izena* (*dejar a un lado, caerse por su peso*).

Egitura	Agerr.	Ehunekoa
adi + det + ize	999	% 37,70
adi + ize	629	% 23,74
adi + prep + ize	577	% 21,77
adi + prep + det + ize	445	% 16,79

6. taula: Gaztelaniazko konbinazioen egiturak.

4.2.3 Gaztelaniazko ordainak euskarazko konbinazioaren arabera

Bi hizkuntzen ezaugarri morfosintaktikoak alderatzen hasteko, euskarazko konbinazioen gaztelaniazko ordain guztiak analizatu, eta hainbat multzotan sailkatu ditugu. Honako sailkapen hau sortu dugu:

- Aditz soila (**adi**): *descentrar, lanzar, descansar*.
- Aditza, determinatzailea eta izena (**adi + det + ize**): *matar el tiempo, hacer la colada, echar la llave*.
- Aditza eta izena (**adi + ize**): *ganar tiempo, tener noticia, hacer caso*.
- Aditza, preposizioa eta izena (**adi + prep + ize**): *arder en deseos, llevar por acompañante, comer con apetito*.
- Aditza eta adberbioa edo adberbio-sintagma (**adi + AdbS**): *salir adelante, sentar bien, llegar lejos*.
- Aditza eta adjektiboa edo adjektibo-sintagma (**adi + AdjS**): *estar claro, volverse loco/a, estar arrepentido/a*.
- Aditza, eta izen soila ez den izen-sintagma (**adi + IS**): *poner mala cara, levantar la tapa de los sesos, hacer buen tiempo*.
- Aditza, preposizioa, eta izen soila ez den izen-sintagma (**adi + prep + IS**): *dejar a medio camino, entrar en uso de razón, pagar en dinero contante*.
- Aditza eta partizipioa (**adi + part**): *pillar inadvertido, dejar frito, estar preocupado*.
- Aditza eta gerundioa (**adi + ger**): *salir corriendo, estar ardiendo, andar endredando*.

- Aurreko multzoetan sartzen ez direnak (**beste**): *estar por suceder, dar mucho que decir, entablar amistad con alguien*.

Kasu- eta postposizio-marka guztiak kontuan hartuta, gehien ageri den ordain mota aditz soila da nabarmen: ordain guztien % 58,07. Ez da harritzekoa, bai baitakigu euskaraz beste hizkuntza batzuetan baino ohikoagoak direla hitz batez baino gehiagoz osatutako aditz konplexu edo elkartuak (Zabala, 2004; Azkarate Villar, 1990), eta askotan itzultzen direla gaztelaniazko aditz soilen bidez.

Interesgarria da, hala ere, beste ordainen egiturei ere begiratutxo bat ematea. Aditz soilak alde batera utzita, lau egitura nabarmentzen dira besteen gaineratik: *adi + det + ize*, *adi + ize*, *adi + prep + ize* eta *adi + prep + det + ize*, hurrenez hurren. Egitura horiek dira, hain zuzen ere, gaztelaniazko konbinazioetan bildu ditugunak (ikus 4.2.2 atala), eta, agerraldien arabera antolatuta, ordena ere mantendu egiten da.

Egitura	Agerr.	Ehunekoa
adi	3711	% 58,07
adi + det + ize	790	% 12,36
adi + ize	544	% 8,51
adi + prep + ize	363	% 5,68
adi + prep + det + ize	275	% 4,30
beste	230	% 3,60
adi + AdjS	155	% 2,43
adi + IS	96	% 1,50
adi + AdbS	76	% 1,19
adi + prep + IS	67	% 1,05
adi + part	63	% 0,99
adi + ger	21	% 0,33

7. taula: Gaztelaniazko ordainen agerraldiak.

Bestalde, markaz markako azterketa eginez gero, absolutiboan dauden konbinazioen ordainek atentzia ematen dute, denen ia bi heren hartzen baitituzte aditz soilek bakarrik: *uzta bildu – cosechar; oreka galdu – desequilibrar...* Gainontzeko ordainetan, berriz, agerpen gehien dituztenak *adi + det + ize* eta *adi + ize* egiturak dira: *itxura egin – hacer el paripé; hotz izan – hacer frío*.

Horrez gain, izenek postposizio-markaren bat izateak ere badu eraginik ordainen egituretan (ikus 8. taula). Izan ere, sarri erabiltzen diren beste markei –ablatiboari, adlatiboari, inesiboari eta instrumentalari– begiratuz gero, aditz soilen ondoren preposiziodun egiturak nagusitzen direla ikusten da, absolutibodun konbinazioen ordainetan ez bezala: *armairutik atera – salir del armario; eskura ordaindu – pagar en mano; bistan egon – saltar a la vista; barrez ito – morir de risa*.

EU	ES egitura	Ehunekoak
	adi	% 63,86
abs	adi + det + ize	% 14,52
	adi + ize	% 9,81
abl	adi	% 29,50
	adi + prep + det + ize	% 22,30
	adi + prep + ize	% 13,67
ala	adi	% 37,28
	adi + prep + det + ize	% 19,74
	adi + prep + ize	% 13,60
ine	adi	% 34,64
	adi + prep + ize	% 19,35
	adi + prep + det + ize	% 14,73
ins	adi	% 60,32
	adi + prep + ize	% 18,25
	adi + prep + det + ize	% 8,73

8. taula: Gaztelaniazko ordainik ohikoenak markaren arabera.

Hori horrela, eta euskarazko postposizio-markak gehienetan preposizioen bidez itzultzen direla jakinik, batek pentsa lezake konbinazioen itzulpena ez dela hain irregularra zentzu horretan, eta, hein batean, zuzen legoke. Izan ere, kasuan kasuko azterketa eginda, ikusi dugu postposizio-marka batzuen baliokidetzat ageri diren preposizioak nahiko sarri direla edozein hitzunik espero litzakeenak. Ablatiboaren pare, esaterako, *de* eta *por* ageri dira ordainen % 86,89tan (*ahotik kendu – quitar de la boca; zeharretik irten – salirse por la tangente*); eta adlatiboaren ordez, berriz, *a* erabili da % 76,47tan (*belarrira esan – decir al oído*).

Dena den, baliokidetzat horiek hein batean baino ez dira erregularrak, postposizio-marka guztiak ez baita gauza bera gertatzen. Inesiboa, adibidez, % 57,7tan baino ez da *en*, *por* edo *so-bre* preposizioen bidez itzuli: *baxoerditan ibili – ir de poteo, txantxetan hartu – tomar a broma*.

4.2.4 Euskarazko ordainak gaztelaniazko konbinazioaren arabera

Gaztelaniazko ordainekin egin dugun bezala, euskarazkoak ere egituraren arabera multzokatu ditugu:

- Aditz soila (**adi**): *geldotu, neskazahartu, txunditu*.
- Absolutibodun izena eta aditza (**ize (abs) + adi**): *bizia arriskatu, ahotsa goratu, eskua jaso*.
- Postposizio-markadun izena eta aditza (**ize**

(**pos**) + **adi**): *buruan sartu, berriketan ibili, negarrari eman*.

- Adberbioa edo adberbio-sintagma eta aditza (**AdbS + adi**): *azkarrago ibili, alferrrik galdu*.
- Adjektiboa edo adjektibo-sintagma eta aditza (**AdjS + adi**): *nabaria izan, izugarria izan, argal mantendu*.
- Izen soila ez den izen-sintagma (absolutiboan) eta aditza (**IS (abs) + adi**): *kontu ezaguna izan, hitzaren jabe egon*.
- Izen soila ez den izen-sintagma (postposizio-markaren batekin) eta aditza (**IS (pos) + adi**): *bere onetik atera, bere kabuz utzi*.
- Aurreko multzoetan sartzen ez direnak (**beste**): *amaitutzat jo, tentuz ibiltzeko esan, hortzen artean hitz egin*.

Multzo bakoitzaren agerpen-kopuruak eta ehunekoak 9. taulan bildu ditugu.

Egitura	Agerr.	Ehunekoak
ize (abs) + adi	2321	% 35,24
adi	1550	% 23,53
ize (pos) + adi	876	% 13,30
beste	576	% 8,74
AdbS + adi	424	% 6,44
AdjS + adi	367	% 5,57
IS (abs) + adi	269	% 4,08
IS (pos) + adi	204	% 3,10

9. taula: Euskarazko ordainen agerraldiak.

Euskarazko ordain guztiak kontuan hartuz gero, argi eta garbi ikusten da bi egitura beste guztiak baino askoz ere gehiago errepikatzen direla: *ize (abs) + adi* motako konbinazioak lehenik, eta aditz soilak ondoren. Nolanahi ere, gaztelaniazko konbinazioak egituraren arabera bereizita, ikusi dugu determinatzaileen eta preposizioen agerpenak baduela zerikusia ordain motarekin.

Har ditzagun, esaterako, aditzez eta izenez osaturiko konbinazioak batetik, eta aditzez, determinatzailez eta izenez osaturikoak bestetik. Agerpen gehien dituzten ordainak *ize (abs) + adi* multzokoak eta aditz soilak dira bi kasuetan, eta gainontzeko ordain motak oso gutxi agertzen dira bi horien aldean. Hala ere, gaztelaniaz determinatzailezik ez duten konbinazioetan, *ize (abs) + adi* multzoko ordainen eta aditz soilen arteko aldea ez da hain nabarmena agerraldiak kontuan hartuz gero; determinatzailea duten konbinazioetan, aldiz, *ize (abs) + adi* egiturako ordainak aditz soilak baino ia hiru bider gehiago dira (ikus aldea 10. taulan).

ES	EU egitura	Ehunekoak
adi+det+ize	ize (abs) + adi	% 50,34
	adi	% 18,56
adi+ize	ize (abs) + adi	% 46,34
	adi	% 31,38

10. taula: Preposiziorik gabeko konbinazioen ordainik ohikoenak.

Horrez gain, hemen ere badago loturarik preposizioen eta postposizio-marken artean. Izan ere, aurreko bi kasuetan absolutibo-markadun izenak nagusi baziren ere, preposiziodun konbinazioen ordainetan ohikoagoak dira postposizio-markaren bat daramaten izenak. Hain zuzen ere, gaztelaniazko konbinazioak *adi + prep + ize* edo *adi + prep + det + ize* multzoetakoak direnean, sarrien agertzen den ordain mota *ize (pos) + adi* da.

ES	EU egitura	Ehunekoak
adi+prep+det+ize	ize (pos) + adi	% 29,78
	adi	% 25,51
	AdbS + adi	% 11,25
	ize (abs) + adi	% 9,31
adi+prep+ize	ize (pos) + adi	% 26,16
	adi	% 23,00
	ize (abs) + adi	% 14,42
	AdbS + adi	% 12,10

11. taula: Preposiziodun konbinazioen ordainik ohikoenak.

Aurreko atalean bezala, hemen ere preposizioen baliokidetzat agertzen diren postposizio-markak, batzuetan, aurreikusteko modukoak dira: *en* preposizioa duten konbinazioen ordainetatik % 87,68k inesibodun izen bat daramate (*estar en la inopia – ametsetan egon*); *con* preposizioaren ordezt instrumentala eta sozietiboa agertzen da ordainen % 87,5etan (*andar con cuidado – kontuz ibili, ir con el cuento – koplarekin etorri*); eta *por* preposizioa ere ablatiboaren eta inesiboaren bidez ordezkatu da % 88,56etan (*pasar por el tamiz – galbahetik pasatu, pasar por las armas – armetan iragan*).

Hemen ere, ordea, hein batean baino ez da erregulartasun hori gordetzen. A preposizioa, esaterako, % 56 kasutan bakarrik itzuli da adlatiboaren edo datiboaren bidez (*traer a la memoria – gogora ekarri*), eta gainontzeko ordain guztiek beste marka batzuk dituzte: *andar a la greña – istilutan ibili, saltar a la vista – begi-bistakoa izan*.

Azkenik, interesgarria da nabarmentzea adberbioak ere nahiko maiz agertzen direla preposiziodun konbinazioen ordainetan; 11. taulan ikus daitekeenez, *adi + prep + det + ize* multzoan, absolutibodun izenekin sortzen diren egi-

turak baino are sarriago agertzen dira adberbioekin sortzen direnak: *caer de su peso – argi egon, dar en el clavo – bete-betean asmatu*, etab.

4.2.5 Mugatasuna eta numeroa: euskarazko eta gaztelaniazko konbinazioen alderaketa

Ezaugarri morfosintaktikoak aztertzen eta alderatzen jarraitzeko, mugatasunari eta numeroari begiratu diegu, hizkuntza batetik bestera mantentzen ote diren jakiteko. Konbinazio-zerrenda osotik zati bat bakarrik hartu dugu, hain zuzen ere, ordaintzat ere konbinazioak dituzten konbinazioez osatua.

Euskaraz hiztegi-sarrerara diren konbinazioetan, mugatu singularrean (*umea izan, ahoa garbitu*) daude izenen ia erdia, eta mugagabearen (*ohar egin, zerraldo utzi*) eta mugatu pluralaren (*goraintziak eman, erroak bota*) agerraldien artean ez dago desberdintasun handirik. Bestalde, konbinazioen hamarren bat baino gehiago zailantzeko kasuak dira (*denbora egin, lotsa izan*), ezin baita zehatz jakin mugatu singularrean ala mugagabearen dauden.

Euskarazko konbinazioen mugatasunari eta numeroari buruzko informazioa 12. taulan bildu dugu, ehunekotan, bai eta ezaugarri hori gaztelaniaz zenbatetan gordetzen den ere.

	EU konbinazioak	ES gordea
s	% 48,48	% 54,23
mg	% 21,99	% 80,72
pl	% 18,08	% 23,44
*	% 11,46	–

12. taula: Mugatasuna eta numeroa euskaratik gaztelaniara.

Eskuineko zutabearen, euskarazko ezaugarria gaztelaniaz zenbatetan gorde den ikus daiteke. Deigarriena mugatu pluralaren kasua da beharbada, baliokideen laurdena baino ez baitago mugatu pluralean gaztelaniaz. Bestalde, mugagabearen kasuan kontrakoa gertatzen da, ordain gehienak mugagabeak baitira hizkuntzaz aldatuta ere.

Gaztelaniazko hiztegi-sarrerak abiapuntutzat hartuz gero, ordea, emaitzak aldatu egiten dira.

	ES konbinazioak	EU gordea
mg	% 50,25	% 27,09
s	% 40,56	% 61,94
pl	% 9,19	% 50,60

13. taula: Mugatasuna eta numeroa gaztelaniatik euskarara.

Batetik, hiztegi-sarrera gehien-gehienak mugagabeen daude (*dar alas, estar en auge*), eta mugatu singularrean ere asko (*perder el juicio, quitar de la cabeza*); mugatu pluralean, berriz, oso konbinazio gutxi daude (*parar los pies, subirse por las paredes*).

Bestetik, baliokideek mugatasuna eta numeroa zenbatetan gordetzen duten begiratu gero, mugagabeak atentzioa ematen du. Euskaratik gaztelaniara oso sarri gordetzen bazen ere, gaztelaniatik euskarara ordainen % 27,09tan baino ez da gorde. Mugatu pluralari dagokionez, ordea, baliokideen erdiak daude mugatu pluralean, eta ez, euskaratik gaztelaniara bezala, laurdena baino gutxiago.

4.3 Aditzak eta izenak hizkuntza batetik bestera

Gure azterketaren lehen atalean (ikus 4.1. atala), hizkuntza bateko eta besteko konbinazioen osagai nagusiak aztertu ditugu: aditzak eta izenak. Oraingoan, berriz, konbinazioen osagaiak euren ordainekin alderatu ditugu. Mugatasuna eta numeroa aztertzeke egin dugun bezala (ikus 4.2.5. atala), izenez eta aditzez osatutako ordainak bakarrik hartu ditugu, dagozkien hiztegi-sarrerekin batera. Gero, bi hizkuntzetako aditzak eta izenak parekatu, eta hiztegian bilatu dugu ea bata bestearen ordaintzat ageri diren.

Adibidez, *deabruetara bidali – mandar al infierno* bikotea hartuta, *deabru – infierno* eta *bidali – mandar* hiztegian baliokidetza jasota ote dauden begiratu dugu. Kasu horretan, aditza bakarrik mantendu da, *deabru* sarreraren ordainetan ez baita *infierno* agertzen.

Hasi aurretik, gure irudipena zen nahiko gutxitan agertuko zirela ordaintzat bai izena eta bai aditza, eta, oro har, gure ustea bete da, nahiz eta proportzioa desberdina izan euskaratik gaztelaniarako eta gaztelaniatik euskararako bikoteetan.

Gure kontaketatik atera ditugun portzentaiak 14. taulan jaso ditugu.

	ize	adi	biak	bat ere ez
eu-es	% 23,84	% 23,71	% 28,01	% 24,44
es-eu	% 19,14	% 22,84	% 21,79	% 36,23

14. taula: Izenen eta aditzen baliokidetza hiztegietan.

Taulako azken zutabeak atentzioa ematen du, izan ere, gaztelaniatik euskararako konbinazio-pareetan, herena baino gehiago dira ez izenik eta ez aditzik baliokide ez dutenak. Euskaraz, berriz, zifrarik altuena kontrako kasuena da, bai izena eta bai aditza ordaintzat dituzten pareena, alegia.

Nolanahi ere, zentzu batean zein bestean, argi ikusten da izen+aditz konbinazioen itzulpena oso gutxitan egin daitekeela hitzez hitz, osagaietako bat behintzat ezin baita normalean hartzen duen ordainez ordezkatu.

Horrez gain, kontuan izan behar da guk hitzak hiztegian agertzen diren ala ez bakarrik begiratu dugula baina, baiezko kasuetan, ez dugula zehaztu ordaina ohikoa den ala ez. Esaterako, *bidea urratu – abrir camino* parean, aditzak baliokidetza hartu ditugu, *urratu* sarreraren barruan *abrir* agertzen delako ordainen artean; euskaraz eta gaztelaniaz egiten dugunok, ordea, nekez erabiliko genituzke ordaintzat, *bide* eta *caminoren* alboan ez bada.

Hortaz, 14. taulak ere agerian uzten du hitz-konbinazioak zenbateraino diren irregularrak eta, hori kontuan hartuta, zenbateraino den beharrezkoa beste itzulpen-estrategia batzuk bilatzea, xede-hizkuntzan testu txukun bat sortu nahi bada.

5 Koloka: izen+aditz konbinazioen informazio linguistikoa biltzen duen datu-basea

Hiztegien gainean egin dugun azterketan, izen+aditz konbinazioei buruzko informazio ugari bildu dugu, eta datu-base bat sortu dugu informazio horrekin guztiarekin. *Koloka*⁸ interfaze publikoak aukera ematen du datu-base horretan bilaketak egiteko, eta erabiltzaileen eskura jartzen du euskarazko eta gaztelaniazko 5.604 konbinazioen eta 12.979 ordainen inguruko informazio linguistikoa.

Bilatzaileak hainbat aukera eskaintzen ditu. Batetik, bilaketa zein hizkuntzatatik abiatuta egin nahi den aukera daiteke: euskaratik gaztelaniara, ala gaztelaniatik euskarara. Bestetik, interfazeak aukera ematen du bilaketa konbinazio osoaren, aditzaren edo izenaren arabera egiteko. Eta, horrez gain, euskara-gaztelania zentzuan, euskarazko izenaren kasu- eta postposizio-marka zehaz daiteke, eta gaztelania-euskara zentzuan, berriz, konbinazioaren egitura.

Bigarren irudian ikusten denez, *adar* izena bilatuta, landu ditugun konbinazioen artean hitz horrekin sortu direnak erakusten dira, ordainekin batera. Euskarazko konbinazio bat gaztelaniara nola itzultzen den jakin nahi duten erabiltzaileek, ziur asko, nahikoa izango dute emaitza horrekin, baina informazio linguistikoa gehiago lortu nahi dutenek ere badute horretarako aukera, ordain bakoitzaren ondoan dagoen [+] ikonotxoa saktuta.

⁸<http://ixa2.si.ehu.es/koloka>

Euzkara - Gaztelania	adar	Izena	Marka gutxiak
adarra jo	tomar el pelo		+
adarrak aldatu	desmogar		+
adarrak aterata	descomar		+
	encomudar		+
adarrak berritu	desmogar		+
	encomudar		+
adarrak ipini	engafar		+
	poner los cuernos		+

2. irudia: *Koloka* interfazearen itxura, *adar* izena bilatuta.

Zubiak eraiki konbinazioa bilatuta, esaterako, bi ordain agertzen dira: *construir puentes* eta *tender puentes* (ikus 3. irudia). Bigarrenaren informazioa zabalduz gero, erabiltzaileak hainbat datu jasoko ditu: *zubiak eraiki* absolutiboan dagoela, eta *tender puentes*, berriz, aditz batez eta izen batez osatu dela; euskarazko konbinazioa mugatu pluralean dagoen bitartean, gaztelaniazkoa mugagabea dagoela; izenak hiztegiaren ordaintzat jasota daudela, baina ez aditzak; eta konbinazio hori Elhuyar euskara-gaztelania hiztegitik jaso dugula.

construir puentes			
tender puentes			
		zubiak eraiki	tender puentes
Marka/legitura morfos.		abz	adi + oz
Mugatuena/numera		pl	mg
Baliokidea		izenak ordaintzat ageri dira hiztegiaren Aditzak ez dira ordaintzat ageri hiztegiaren	
Zein hiztegitatik hartua		Elhuyar eu > es	

3. irudia: *Zubiak eraiki* bilaketaren emaitza.

6 Ondorioak eta etorkizuneko lanak

Azterketa linguistiko honen helburua hitz-konbinazioen konplexutasuna agerian uztea izan da, argi gera dadin zeinen garrantzitsua den Hizkuntzaren Prozesamenduko aplikazioetan horrelako egiturei tratamendu berezia ematea.

Gure ikerketa-lan nagusia corpusetan oinarritzeko asmoa badugu ere, hiztegi elebidunak erabili ditugu lehenik, konbinaziorik usuenak biltzen dituztelakoan.

Horretarako, hizkuntza bateko eta besteko izen+aditz konbinazioak eta euren ordainak aztertu ditugu hiru ataletan. Lehenik, euskarazko eta gaztelaniazko konbinazioetan ageri diren aditzak eta izenak begiratu diegu, eta ikusi dugu badagoela antzekotasunik euren artean –aditzen kasuan, bereziki–, nahiz eta gaztelaniazko aditz-zerrenda euskarazkoa baino dezente zabalagoa izan. Bigarrenik, euskarazko zein gaztelaniazko konbinazioen eta ordainen egiturak parekatu ditugu, eta, loturaren bat aurkitu badugu ere, gu-

re emaitzek argi uzten dute ez dagoela konbinazioen itzulpenetarako arau morfosintaktiko orokorrik sortzerik. Azkenik, ezaugarri semantikoei ere eman diegu begiratu bat, eta, ordaintzat konbinazio bat duten konbinazioetan, izenak izenekin eta aditzak aditzekin baliokide ote diren jakin nahi izan dugu. Atal horretan ere agerian geratu da hitzez hitz itzul daitezkeen konbinazioak oso bakanak direla.

Oro har, beraz, gure ustea bete da: antzekotasunak antzekotasun, izen+aditz konbinazioak oso aldakorak dira euskaratik gaztelaniara eta gaztelaniatik euskarara, eta behar-beharrezkoa da egitura horiei tratamendu berezi bat ematea, HPko tresnen bidez emaitza egokiak lortu nahi badira.

Azkenik, gure azterketaren emaitza guztiak sarean jarri ditugu eskuragarri, <http://ixa2.si.ehu.es/koloka> helbidean, edozein erabiltzailek bilaketak egiteko moduan.

Aurrera begira, hiztegiaren gainean hasitakoa corpusetara eramatea da gure helburua, orain arte aztertutako konbinazioak testu errealean zenbat erabiltzen diren ikusteko, eta gure konbinazio-zerrendak zabaltzeko. Hartara, gure beharretara hobeto egokituko den datu-base bat osatu ahal izango dugu.

Horrez gain, azterketa hau alderdi praktikora eraman nahi dugu, eta ikuspegi linguistikotik landu ditugun konbinazioetako bakoitzak itzulpen automatikoan zer motatako arazoak sortzen dituen aztertu.

Eskerrak

Lan hau Ekonomia eta Lehiakortasun Ministeriooko doktoretza aurreko diru-laguntza bati esker egin ahal izan dugu (BES-2013-066372), SKATeR proiektuaren barruan (TIN2012-38584-C06-02).

Eskerrak eman nahi dizkiogu Ruben Urizarri gurekin aholkulari-lanetan aritu izanagatik, eta baita Mikel Artetxeri ere, bilaketak egiteko interfazea prestatu izanagatik.

Horrez gain, Elhuyarren laguntza ere ezinbestekoa izan da azterketa honetarako, haiek eman baitigute gure lanaren oinarri izan den materiala.

Bibliografia

- Aduriz, Itziar, Izaskun Aldezabal, Iñaki Alegria, Xabier Artola, Nerea Ezeizai, & Ruben Urizar. 1996. Euslem: A lemmatiser/tagger for basque. *Proc. Of EURALEX'96*, pages 17–26.
- Alonso Ramos, Margarita. 1995. Hacia una definición del concepto de colocación: de jr firth a ia mel'cuk. *Revista de Lexicografía*, 1:9–28.

- Azkarate Villar, Miren. 1990. *Hitz elkartuak euskaraz*. Filosofi-Letren Fakultatea, Deustuko Unibertsitatea.
- Baldwin, Timothy & Suñam Kim. 2010. Multiword expressions. *Handbook of Natural Language Processing, second edition*. Morgan and Claypool.
- Corpas Pastor, Gloria. 2001. La traducción de unidades fraseológicas: técnicas y estrategias. *La lingüística aplicada a finales del siglo XX. Ensayos y propuestas*, Alcalá, Universidad de Alcalá, 2:779–787.
- Corpas Pastor, Gloria. 2004. *Diez años de investigación en fraseología: análisis sintáctico-semánticos, contrastivos y traductológicos*. Madrid: Vervuert.
- Gurrutxaga, Antton. 2014. *Idiomatikotasunaren karakterizazio automatikoa: izena+aditza konbinazioak*. Ph.D. thesis, UPV-EHU.
- Heylen, Dirk & Kerry Maxwell. 1994. Lexical functions and the translation of collocations. In *Proceedings of Euralex*.
- Iñurrieta, Uxo. 2013. Itzulpen automatikorako patroibilaketa. Master's thesis, UPV-EHU.
- Lü, Yajuan & Ming Zhou. 2004. Collocation translation acquisition using monolingual corpora. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 167. Association for Computational Linguistics.
- Mayor, Aingeru, Iñaki Alegria, Arantza Díaz De Ilarraza, Gorka Labaka, Mikel Lersundi, & Kepa Sarasola. 2011. Matxin, an open-source rule-based machine translation system for basque. *Machine translation*, 25(1):53–82.
- Mel'čuk, Igor A. 1998. Collocations and lexical functions. *Phraseology. Theory, Analysis, and Applications*, pages 23–53.
- Rafel, Joan. 2004. Los predicados complejos en español. In *Las fronteras de la composición en lenguas románicas y en vasco*, pages 445–534. Servicio de Publicaciones.
- Sanz Villar, Zuriñe. 2011. Alemanetik euskara itzulitako unitate fraseologikoen azterkerako jarraibideak. *Senez: itzulpen aldizkaria*, 41:125–139.
- Seretan, Violeta. 2013. On collocations and their interaction with parsing and translation. *Informatics*, 1(1):11–31.
- Smadja, Frank, Kathleen McKeown, & Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational linguistics*, 22(1):1–38.
- Urizar, Ruben. 2012. *Euskal lokuzioen tratamendu konputazionala*. Ph.D. thesis, UPV-EHU.
- Wehrli, Eric, Violeta Seretan, Luka Nerima, & Lorenza Russo. 2009. Collocations in a rule-based mt system: A case study evaluation of their translation adequacy. In *Proceedings of the 13th Annual Meeting of the European Association for Machine Translation*, pages 128–135.
- Zabala, Igone. 2004. Los predicados complejos en vasco. In *Las fronteras de la composición en lenguas románicas y en vasco*, pages 445–534. Deustuko Unibertsitatea.