



Itzulpen Automatikorako patroi-bilaketa

Egilea: Uxoia Iñurrieta Urmeneta

Tutoreak: Nora Aranberri, Arantza Díaz de Ilarraza eta Kepa Sarasola

hap

Hizkuntzaren Azterketa eta Prozesamendua masterreko titulua
lortzeko bukaerako proiektua

2013ko iraila

Sailak: Euskal Hizkuntza eta Komunikazioa, Lengoaia eta Sistema Informatikoak,
Konputazio Zientziak eta Adimen Artifiziala.

Laburpena

Proiektu honetan, erregeletan oinarritutako Itzulpen Automatikoko sistema bat hobetzeko patroi orokor batzuk sortu nahi izan ditugu, hiztegietako adibideetan oinarrituta. Patroiak sortzeko gure lehen hurbilpena izan denez, izen bereziak edo zenbakiak barne hartzen dituzten esaldi eta esaldi-zatiak bakarrik hartu ditugu. Gaztelania-euskara hiztegiko 132 adibide eta ingelesa-euskara hiztegiko 420 hartu, eta aztertu dugu ea itzulpen-sistema hobetuko luketen eta zenbat erabiliko liratekeen. Azkenean, gaztelania-euskara sistemarako 30 patroi eta ingelesa-euskara sistemarako 104 sortu ditugu.

Abstract

The aim of this project is to create some patterns that will improve a rule-based machine translation system. As this is our first approach, we have just taken some sentences and parts of sentences from dictionary examples (132 from a Spanish-Basque dictionary and 420 from an English-Basque dictionary), all of them containing entity nouns or numbers, and we have analyzed whether they would improve our system and how often they would be used. We have finally decided to create 30 patterns for the Spanish-Basque system and 104 for the English-Basque one.

Aurkibidea

Aurkibidea	3
Taulen aurkibidea	4
Irudien aurkibidea	5
1 Proiektuaren definizioa	6
2 Aurrekariak eta interesa	7
3 Baliabideak	10
4 Itzulpen-patroi eraginkorren bila	16
5 Emaidzak	31
6 Ondorioak eta etorkizunerako lanak	35
7 Erreferentziak	37
Eranskinak	39

Taulen aurkibidea

1	Gaztelaniatiko itzulpenen sailkapena	16
2	Ingelesetiko itzulpenen sailkapena.....	16
3	Lehen fasean baztertutako laginak, gaztelania-euskara	39
4	Lehen fasean baztertutako laginak, ingelesa-euskara	41
5	Agerraldi kopuruagatik baztertutako adibideak, gaztelania-euskara	43
6	Agerraldi kopuruagatik baztertutako adibideak, ingelesa-euskara	44
7	Patroiak sortzeko aukeratutako adibideak, gaztelania-euskara	52
8	Patroiak sortzeko aukeratutako adibideak, ingelesa-euskara	53
9	Euskarazko patroi automatikoen zuzenketak	56
10	Gaztelaniazko patroi automatikoen zuzenketak	56
11	Ingelesezko patroi automatikoen zuzenketak	56
12	Gaztelania-euskara sistemarako proposatzen ditugun patroiak	31
13	Ingelesa-euskara sistemarako proposatzen ditugun patroiak.....	31
14	Aldagairik gabeko segida-pareak, ingelesa-euskara	34

Irudien aurkibidea

1 Elhuyarren gaztelania-euskara hiztegiko sarrera bat: <i>preguntar</i>	10
2 Eiheraren demoa Interneten: “ <i>Amagoiaren lagunak Galizian egiten du lan</i> ”	11
3 Entitate-detektatzaileen emaitzak, parez pare	12
4 Matxinen itzulpen-prozesua	12
5 Transferentzian, gaztelaniazko egitura.....	13
6 Transferentzian, euskarazko egitura	14

1 Proiektuaren definizioa

Itzulpen Automatikoak (IA) gero eta garrantzi handiagoa du gaur egun munduko hainbat eta hainbat hizkuntzatan, eta helburu horretarako sorturiko tresnak etengabe ari dira garatzen. Estatistikan eta erregeletan oinarritutako sistemek zailtasunak izaten dituzte hainbat egitura itzultzeko, eta horregatik, gero eta ikerlan gehiago egiten ari dira bi metodo horietan arazoak ematen dituzten egiturak adibideetan oinarritutako sistemen bidez itzultzeko.

Lan honen helburua da, adibideetatik ateratako patroïak erabiliz, erregeletan oinarritutako Matxin itzultzaile automatikoak egiten dituen akatsetako batzuk zuzentzen laguntzea. Horretarako, Elhuyar hiztegiak gaztelania-euskara eta ingelesa-euskara hizkuntza-bikoteetarako ematen dituen adibideak landu ditugu, eta aztertu dugu horietako zeintzuk izan litezkeen baliagarriak gure itzulpen-sistamarako.

Lehen hurbilpen honetan, hiztegiko adibideetako batzuk baino ez ditugu hartu: entitateren bat daramaten esaldi edo esaldi-zatiak –gutxienez izen berezi bat edo zenbaki bat barne hartzen dutenak, alegia–. Adibide horiek aztertu eta gure IAko sistema hobetzeko erabilgarrienak iruditu zaizkigunak hautatu ditugu, hortik patroï batzuk ateratzeko. Adibideetan oinarritutako metodoak oso eraginkorrak izan daitezke zenbait akats zuzentzeko, baina kontuan izan behar da patroï gehiegi definitzeak nabarmen motelduko lukeela itzulpen-prozesua eta, hortaz, garrantzitsua dela adibide guztien artetik baliagarrienak bakarrik aukeratzea.

Patroïrik eraginkorrenak zeintzuk diren erabakitzeke, bi irizpide nagusi izan ditugu: batetik, aztertutako esaldien itzulpen automatikoaren egokitasuna, eta bestetik, esaldi horien erabilera-maiztasuna. Lehen pausoa jatorrizko esaldi guztiak Matxinen bidez itzultzea izan da, hiztegiko itzulpena makinarenarekin alderatzeko. Horrela, itzulpen automatiko onargarriak ematen dituzten adibideak alde batera utzi ditugu, emaitza okerrak ematen dituztenak bakarrik lantzeko, horiek zuzentzea baita gure xedea. Ondoren, gehien erabiliko diren patroïak zein diren jakiteko, jatorrizko esaldiak Interneteko bilatzaile batean sartu, eta emaitza gehien eman dituztenak bakarrik aukeratu ditugu.

Horrez gain, sortutako patroïak ahalik eta kasu gehienetan erabili ahal izateko, esaldietan ageri diren entitateak aldagai gisa definitzea erabaki dugu, eta hiru entitate-detektatzailez baliatu gara horretarako. Hala, patroïak sortzeko lana erdi-automatiki egin dugu.

2 Aurrekariak eta interesa

Itzulpen Automatikoak gero eta indar handiagoa du gaur egun mundu osoan zehar. Gero eta hizkuntza gehiago ari dira helburu horretara bideratutako sistemak garatzen (Hutchins, 2007), bai hizkuntza handiak, bai txikiak ere. Horren adierazgarri da euskaraz ere itzultzaile automatiko bat baino gehiago egotea: Eusko Jaurlaritzaren itzultzailea¹, Matxin², Opentrad³, Google⁴ eta EUSMT⁵ (Labaka, 2010). Gainera, aipagarria da Europako Batzordeak 2012ko otsailean argitaratutako Work-programme 2013 txostenean, informazio- eta komunikazio-teknologien atalean, hizkuntza-teknologiei buruzko lau ikerlerro nagusi ageri direla, eta Itzulpen Automatikoa dela horietako bat.

IAko sistema gehienek metodo estatistikoak (Farrús et al., 2009; Koehn, 2010) edo erregeletan oinarritutakoak (Mayor, 2008) erabiltzen dituzte. IXA taldean, egokiagotzat jo dute euren sistema, Matxin, erregeletan oinarritzea. Izan ere, metodo estatistikoek corpus erraldoiak behar izaten dituzte emaitza onargarriak lortzeko, eta euskarak ez du tamaina horretako corpusik oraindik. Arlo jakin bateko testuen corpus handi bat bilduz gero, domeinu horretan emaitza hobeak ematen dituzte estatistikan oinarritutako sistemek, baina, oro har, edozein gai itzultzeko, oraindik itzulpen hobeak lortzen dira erregeleen bidez. Gainera, oso zaila da metodo estatistikoek emaitza egokiak ematea euskarak bezain morfologia aberatsa duten hizkuntzetara itzultzean.

Gaur egun, aipatutako bi metodoez gain, gero eta ikerlan gehiago egiten dira adibideetan oinarritzen direnen inguruan ere (Díaz de Ilarraza, 2002; Salam et al., 2011). Nagaok (1984) dio europar hizkuntzek oinarri komun bat dutela gehienetan, eta hizkuntza horien arteko itzulpenak egin daitezkeela egitura gehiegi aldatu gabe. Ingelesaren eta japonieraren artean, ordea, ez da hala gertatzen –ezta beste edozein hizkuntzaren eta euskararen artean ere–, eta erregelen bidezko metodoak ez dira nahikoa:

“Sometimes (...) there is no good structural correspondence between each part of the sentences of the two languages. (...) A translation of this kind cannot be achieved by a mere detailed syntactic analysis of the original sentence. (...) the detailed analysis of a source language sentence is of no use for the translation between languages of completely different structure like English and Japanese.”

¹ <http://www.itzultzailea.euskadi.net>

² <http://matxin.elhuyar.org/>

³ <http://www.opentrad.com/>

⁴ <http://translate.google.com/#es/eu/>

⁵ <http://ixa2.si.ehu.es/openmt-demo/>

[Batuetan (...) ez dago kidetasunik bi hizkuntzetako esaldien egituren artean. (...) Era horretako itzulpen bat ezin da lortu jatorrizko esaldiaren sintaxia sakon aztertuta bakarrik. (...) jatorrizko hizkuntzako esaldiaren analisi xehea ez da erabilgarria egitura guztiz desberdina duten hizkuntzen arteko itzulpenetan, ingelesaren eta japonieraren kasuan gertatzen den bezala.]

Bestalde, adibideetatik jasotako informazioa ahalik eta gehien orokortzeko, gero eta ugariagoak dira patroietan informazio semantikoa gehitzeko saiakerak. Horretan aritu dira, besteak beste, Universität Zurich-eko eta Dublin City University-ko ikerketa-taldeak (Ebling et al., 2011), bai eta Osakako Oki Electric Industry enpresa ere (Kitamura et al., 2003).

Osakako taldeak adibideetan bakarrik oinarritutako sistema bat sortu du, itzulpen-memorieta informazioa erabiliz. Hala ere, itzulpenetako adibideak bere horretan erabili beharrean, patroi orokorrak sortu nahi izan dituzte, adibideak zatituta. Dublingo eta Zurich-eko taldeek, berriz, adibideetan oinarritutako bi sistema elkartu dituzte, bien artean ingelesetik alemanerako itzulpen-sistema hobe bat lor daitekeen ala ez aztertzeko. Lan horretan, alemana morfologia aberatseko hizkuntza denez, zailtasunak aurkitu dituzte esaldi-zatiak batzeko orduan, sistemak ez baititu kasu gramatikalak kontuan hartzen:

“The biggest shortcoming of EBMT is that it does not combine translations of phrases well. (...) It is particularly frequent when translating into a morphologically rich language.”

[AOIaren hutsune nagusia da ez dituela esaldien itzulpenak ondo konbinatzen. (...) Hori oso sarri gertatzen da morfologia aberatseko hizkuntzetara itzultzean.]

Hori dela eta, munduko sistema gehienak hainbat estrategia konbinatzen saiatzen dira Itzulpen Automatikoak dituen askotariko arazoak konpontzeko; izan ere, metodo bakoitza egokia da ataza jakin batzuetarako, baina metodo bakarrak ezin die IAren zailtasun guztiei aurre egin (Labaka et al., 2011). Matxinen sortzaileek eurek ere aitortzen dutenez, “etorkizuna hibridazioaren bidetik etorriko da seguruen” (Mayor et al., 2009).

Europar –eta mundu osoan, oro har– Itzulpen Automatikoa lantzen duten hainbat eta hainbat proiektu daude, eta IXA taldeak horietako batzuetan parte hartzen du (informazio gehiago, <http://ixa.si.ehu.es/Ixa/Proiektuak> helbidean):

- OpenMT2 proiektuan, hainbat motatako itzulpen-sistemen arteko hibridazioaren inguruan ari dira ikertzen.
- ENEus proiektuan, ingelesetik euskararako IA lantzen da.
- QTLeap proiektuan, Itzulpen Automatikoaren kalitatea hobetu nahi da, hizkuntza-teknologiaren alorreko beste baliabide batzuk erabilia: sakon

analizatutako corpus paraleloak, semantika lexikala, entitateak (NECR), Interneteko baliabide irekiak (Linked Open Data)...

- TACARDI proiektuak bi ikerlerro nagusi ditu:
 - Interneteko baliabideen erabilera, itzultzaile automatikoak hobetzeko, hiztegiak aberasteko (off-line) eta informazio eleanitza kontsultatzeko (on-line).
 - Testuinguruaren erabilera, dokumentu-mailan automatikoki itzultzeko, eta ez esaldi-mailan bakarrik. Testuzkoa ez den metainformazioa ere erabili nahi da.
- Ber2Tek proiektuan ere IAko aplikazio errealekin egiten da lan, beste hainbat gairen inguruan ere ikertzen bada ere.

3 Baliabideak

Lan honen oinarri izan diren esaldi-pareak Elhuyar fundazioaren hiztegi elebidunetatik erauzi ditugu, eta hiru entitate-detektatzailez baliatu gara lagin horiek lantzeko. Horrez gain, proiektu honetatik ateratako patroiak Matxin itzultzaile automatikoan integratuko direla pentsatuz egin dugu ikerlana. Datozen lerroetan, aipatutako hiru produktuen inguruan arituko gara.

3.1 Elhuyar hiztegi elebidunak

Elhuyar fundazioak hainbat tresna eta baliabide garatu ditu hizkuntza-teknologien alorrean⁶, horien artean, Interneten erabilgarri dauden hainbat corpus eta askotariko hiztegi elebakarrak eta elebidunak, paperezko edizioetan zein sarean eskuragarri. Lan honetarako, gaztelania-euskara eta ingelesa-euskara hiztegi elebidunak⁷ erabili ditugu.

Gaztelania-euskara hiztegiaren lehen edizioa 1996an argitaratu zen, 85.000 sarrera eta 16.000 azpisarrerarekin. Orduetik, ordea, hiztegia osatuz joan da, eta 3. edizioak, 2006ko azaroan argitaratuak, lehenak baino 3.000 sarrera eta 4.000 azpisarrera gehiago ditu. Horrez gain, 100.000 adibide ere eskaintzen ditu, hitz bakoitza nola erabiltzen den azaltzeko. Ingelesa-euskara hiztegiak, berriz, 34.000 hiztegi-unitate eta 12.000 adibide biltzen ditu. Sarrerek honako itxura hau dute, adibide eta guzti:

castellano vasco

ADIERAK

1 ♦ v.tr. ♦ ♦ ♦
galdetu, galdegin, itaundu

Si no lo sabes, pregúntaselo al maestro: ez badakizu, galdetu maisuari.

2 ♦ v.pnli. ♦ ♦ ♦
jakin nahi/gura izan; nork bere buruari galdetu/galdegin

*Todos nos preguntamos si ibamos a poder salir de allí: denon buruan zegoen handik ateratzerik izango ote genuen.
Todo el mundo se preguntaba cuál sería la solución del problema: denek jakin nahi zuten zein ote zen arazoaren irtenbidea.
Me pregunto si vale la pena: merezi ote duen galdetzen diot neure buruari.*

AZPISARRERAK

LOKUZIOAK

▶ [preguntar por](#)

1. irudia: Elhuyarren gaztelania-euskara hiztegiko sarrera bat: *preguntar*

⁶ Baliabideei buruzko informazio osoa: <http://www.elhuyar.org/hizkuntza-zerbitzuak/>.

⁷ Kontsultak <http://www.elhuyar.org/hizkuntza-zerbitzuak/EU/Hiztegi-kontsulta> helbidean egin daitezke.

Ikerlan hau patroiak aztertzeko lehen hurbilpen bat besterik ez denez, hiztegi-takoa adibide batzuk bakarrik hautatu ditugu: entitatearen bat dutenak. Izan ere, gure asmoa patroiak orokortzea da, entitateak aldagai bihurtu, eta izen berezi bat edo zenbaki bat dagoen tokian beste edozein entitate jarri ahal izateko. Automatikoki erauzi ditugu adibideak, eta hiru detektatzailez baliatu gara izen bereziak eta zenbakizko egiturak hautemateko.

3.2 Entitate-detektatzaileak

Entitateak hautemateko, hiru entitate-detektatzaile erabili ditugu: bat euskararako, beste bat gaztelaniarako, eta hirugarrena ingeleserako.

Euskarazko esaldientzat erabili dugun entitate-detektatzailea IXA taldean garatua da, eta denborazko egiturak, zenbakizkoak eta izen bereziak hautematen ditu. Lehenik, EUSLEM lematizatzaile-etiketatzailerak erabiltzen da (Aduriz et al., 1996) analisi morfologikoa egiteko. Ondoren, izen bereziak hautemateko, beste tresna bat erabiltzen da: *Eihera*, informazio morfologikoan oinarritutako arau linguistikoz osaturiko gramatika (Alegria et al., 2004):

Entitateak

Amagoiaren lagunak Galizian egiten du lan.

Pertsona Lekua Erakundea

Emaitza:

Amagoiaren lagunak Galizian egiten du lan .

2. irudia: *Eihera*ren demoa Interneten: “*Amagoiaren lagunak Galizian egiten du lan*”

Entitatearen ezaugarri nagusitzat letra larriak hartzen ditu, baina beste hainbat ezaugarri ere hartzen dira kontuan, besteak beste, kategoria eta azpikategoria gramatikalak. Entitatea non dagoen detektatzeaz gain, pertsona-izena, leku-izena edo erakunde-izena den zehaz dezake, eta postposizio-informazioa ere jasotzen du.

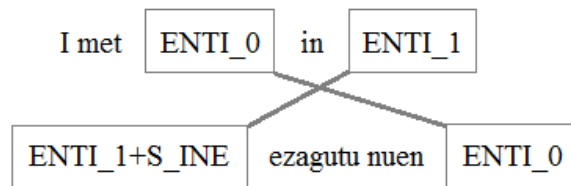
Guk horrelako patroiak sortu ditugu, entitatea eta bere postposizio-informazioa kontuan hartuta:

- (1) Euskarazko esaldia: *Izan zara inoiz Parisen?*
 Emaidza: *Izan zara inoiz ENTI_0+S_INE?*

Gaztelaniarako eta ingeleserako, berriz, Kataluniako Unibertsitate Politeknikoak garaturiko *Freeling* (Padró et al., 2012) eta Stanfordeko unibertsitatearen *Stanford CoreNLP* (Finkel et al., 2009) analizatzaileak erabiltzen dira. Hona hemen ingelesezko eta gaztelaniazko nolako patroiak aztertu ditugun:

- (2) Gaztelaniazko esaldia: *¿Has estado alguna vez en París?*
 Emaidza: *¿Has estado alguna vez en ENTI_0?*
- (3) Euskarazko esaldia: *¿Have you ever been to Paris?*
 Emaidza: *Have you ever been to ENTI_0?*

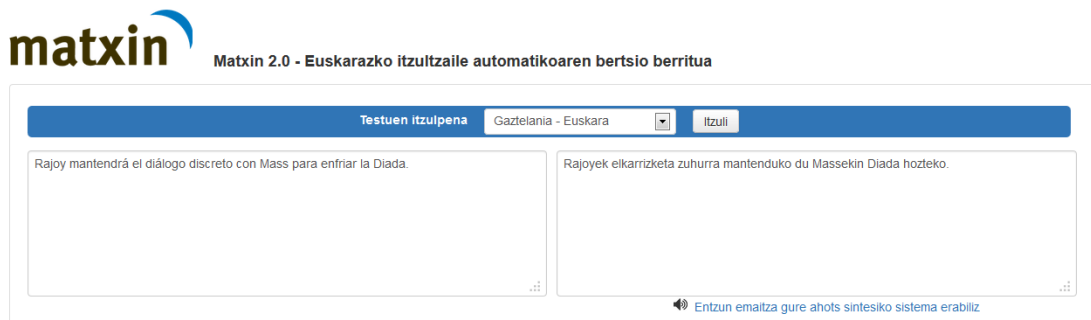
Ondoren, detektatzaileen bidez lortutako patroiak bikoteka jarri, eta itzulpenetako entitateak erlazionatzen dira:



3. irudia: Entitate-detektatzaileen emaitzak, parez pare

3.3 Matxin itzultzaile automatikoa

Matxin⁸ da euskararekin lan egiten duen eta publikoki erabilgarri jarri den lehen itzultzaile automatikoa. Erregeletan oinarritutako sistema bat darabil, eta transferentzia sintaktiko sakona egiten du. Oinarrian, hainbat tresna eta baliabide linguistiko ditu –horietako asko, itzulpen-sistema bera bezala, IXA taldean garatu dira–.



4. irudia: Matxin itzultzaile automatikoa

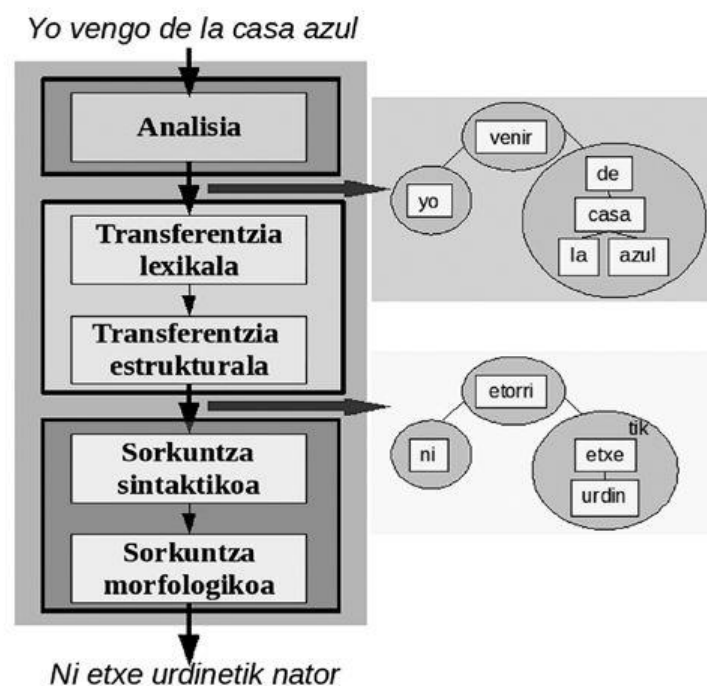
⁸ Interneteko helbidea: <http://matxin.elhuyar.org/translatetext>

IXA ikerketa-taldea 1998an hasi zen itzultzaile automatiko bat sortzeko lehen urratsak egiten, ingelesetik euskararako sistema bat garatzeko lehenik, eta gaztelaniatik euskararako bat sortzeko ondoren. 2007an egin zuten publiko Matxin, baina ordutik hona ere etengabe aritu dira sistemaren erroreak aztertzen eta hobekuntzak egiten, hainbat proiektutan parte hartuta.

Mayorrek (2009) dioenez, sistemaren ebaluazioetan, itzulpenen % 69a ulertzeko modukoa dela ondorioztatu da, eta honako hau aipatzen da sistema hobetzeko teknikei dagokienez:

“Gure sistemak dituen mugak gainditzeko beste hainbat teknika eta estrategia aztertzen eta erabiltzen hasiak gara: (...), corpusetan oinarritutako estrategien ikerketa eta *Matxin* estrategia horiekin konbinatzea edo hibridatzea.”

Matxin sisteman, itzulpen-prozesuak hiru fase ditu: analisia, transferentzia eta sorkuntza (Mayor et al., 2009). Honako irudi honetan, adibide baten bidez ikus daiteke sistemak nola funtzionatzen duen:



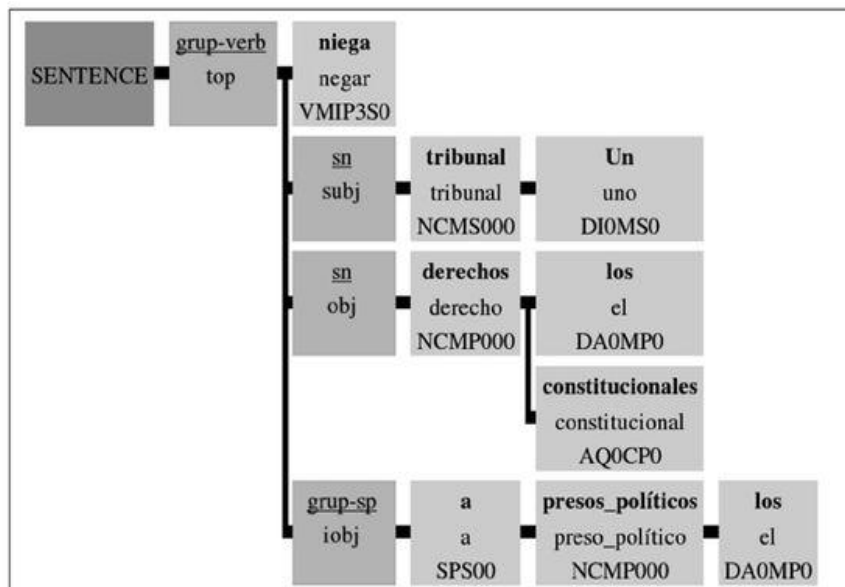
Datozen hiru azpiataletan, sistemaren fase bakoitzaren ezaugarriak aipatuko ditugu labur.

3.3.1 Analisia

Lehen irudian ikusten den bezala, analisia egiten da lehenik, eta UPC unibertsitatean espainierarako garatutako tresna batez baliatzen da Matxin horretarako: *FreeLing* (Atserias et al., 2006).

3.3.2 Transferentzia

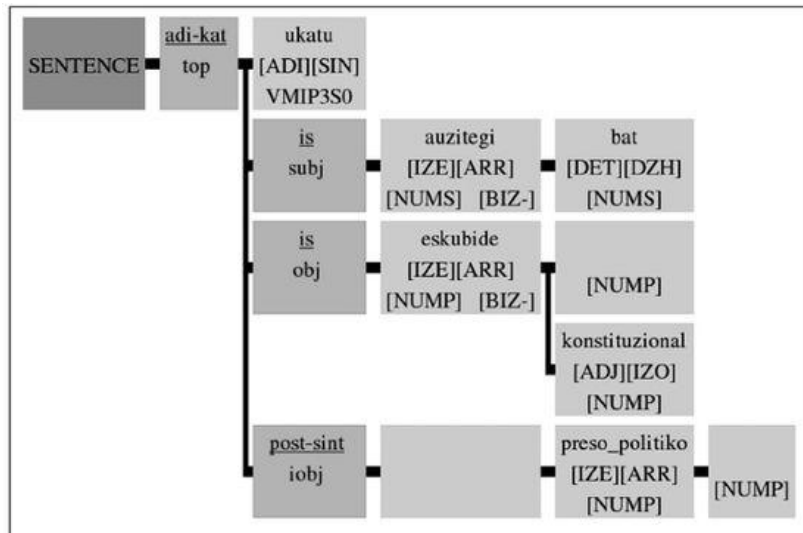
Bigarren fasean, transferentzia egiten da, bi mailatan: lexiko mailan eta egitura mailan. Lehenik, transferentzia lexikala egiten da, eta elementu bakoitza hainbat atributuren bidez gordetzen da: posizioa, forma, numeroa... Gainera, osagai guztiak chunk edo sasi-sintagmetan gordetzen dira (zatiketa postposizioa daraman unitateka egitea da helburua).



6. irudia: Transferentzian, gaztelaniazko egitura

Lexikoi elebidunak Elhuyar hiztegitik eta Euskalterm banku terminologikotik jasotako informazioa biltzen du, besteak beste, eta hor egiten da kontsulta, xedehizkuntzako atributuen balioak emateko. Anbiguotasun lexikalak ebazteko, lexikoi elebiduneko lehen adieraren lehen ordaina hautatzen da gehienetan, horrek izaten baitu itzulpen egokia emateko aukera gehien.

Transferentzia estrukturalen, berriz, balio lexikalik gabeko elementuen informazioa gordetzen da (preposizio eta artikuluen), eta jatorrizko hizkuntzako elementu horiek desagertu egiten dira. Aditz-kateak transferitzen direnean, egitura berri bat sortzen da bildutako informazio guztiarekin, baina egitura orokorra mantendu egiten da, beharrezko aldaketak sorkuntzan egiten baitira.



7. irudia: Transferentzian, euskarazko egitura

3.3.1 Sorkuntza

Behin transferentzia egindakoa, sistemak beharrezko informazioa izaten du xedehizkuntzako esaldia sortzeko. Batetik, sorkuntza sintaktikoa egiten da: elementuak ordenatu, eta postposizio-informazioa chunk bakoitzeko azken osagaiari gehitzen zaio. Ondoren, sorkuntza morfologikoa, elementu lexikal etiketatuak interpretatu, eta hitzei forma ematen zaie. Postposizio-informazioa duten hitzak bakarrik prozesatzen dira, besteetan hitzaren forma lemaren forma bera izaten baita; hau da, aditz-kateetan elementu guztiak lantzen dira, baina beste chunketan azken posizioan dagoena bakarrik. Sorkuntza morfologikorako erabiltzen den tresna IXA taldean garatutako *Morfeus* prozesadore morfologikoa da (Alegria eta Urkia, 2002).

Irudietan erabili den adibiderako, honako esaldi hau sortzen du Matxinek: *Auzitegi batek eskubide konstituzionalak ukatzen dizkie preso politikoei.*

4 Itzulpen-patroï eraginkorren bila

Ezertan hasi aurretik, Elhuyar fundazioaren gaztelania-euskara eta ingelesa-euskara hiztegietako adibideetako batzuk erauzi ditugu, eta horietatik entitateren bat –gutxienez izen berezi bat edo zenbaki bat– duten esaldiak hartu ditugu gure azterketari ekiteko, entitateek emango baitigute aldagaiak definitu eta patroï orokorrak sortzeko aukera.

Ondoren, hiru zati nagusitan banatu dugu gure lana. Lehenik, hiztegiko adibideak Matxinen bidez itzuli, eta hiztegiko itzulpenak eta makinarenak alderatu ditugu; bigarrenik, esaldien erabilera-maiztasuna aztertu dugu; eta azkenik, hautatutako adibideetatik patroïak sortu ditugu, erdi-automatikoki. Dena den, azterketa horren guztiaren ondorio izan diren patroï batzuk hobetu zitezkeela iruditu zaigunez, bukaeran aldaketa txiki batzuk egin dizkiogu hainbat patroïri.

4.1 Hiztegiko adibideen eta euren itzulpen automatikoen arteko alderaketa

Hasteko, hiztegiko adibideetatik gaztelaniazko eta ingelesezko esaldiak hartu, eta Matxinen bidez itzuli ditugu, hiztegiko itzulpenak eta makinarenak alderatu ahal izateko. Aintzat hartu behar da balorazio-lan hori norberaren ikuspuntuaren arabera dela, eta askotan ez dela erraza izaten zerbait zuzena den ala ez esatea edo, are gehiagotan, bi itzulpenen artean zein hobetsi erabakitzea. Ahalik eta ikuspegi zabalena izaten saiatu gara, baina baliteke beste ikerlari batzuk gure sailkapenarekin ados ez egotea.

Nahiko malgu jokatu dugu, bai baitakigu eskuzko itzulpenak eta automatikoki egindakoak ezin direla berdin ebaluatu, eta gainera, kontuan izan behar da gure sistema garapen-prozesuan dagoela oraindik. Hortaz, itzulpen automatikoak ulergarriak iruditu bazaizkigu, onargarritzat jo ditugu, nahiz eta, beharbada, guztiz zuzenak ez izan. Esate baterako, hitzen ordena zalantzazkoa izan arren esaldia arazorik gabe ulertzen bada, itzulpena onargarria dela erabaki dugu. Hona hemen adibide bat:

(4) Ingelesezko adibidea: *He has gone over to France.*

Hiztegiko itzulpena: *Frantziara joan da.*

Itzulpen automatikoa: *Hura joan da Frantziara.*

Bestalde, garrantzitsua da nabarmentzea egiturari eman diogula garrantzia bilaketarako, eta ez diogula entitate-izenen itzulpenari erreparatu, izen berezien itzulpen okerra ez baita patroïek konpondu nahi luketen arazoa. Izan ere, batzuetan, makinak ez du informazio nahikoa izaten entitate-izenak zuzen itzultzeko, eta, esaldiaren beraren itzulpena egokia izan arren, izen horretan egiten du, hain zuzen, akatsa. Esaterako:

- (5) Gaztelaniazko adibidea: *Dios maldijo la serpiente*.
 Hiztegiko itzulpena: *Jainkoak sugea madarikatu zuen*.
 Itzulpen automatikoa: *Diosek sugea madarikatu zuen*.

Adibide horretan, hiztegiko itzulpenaren eta itzulpen automatikoaren artean dagoen desberdintasun bakarra izen berezia da. Izena beste bat balitz, ordea, emaitza egokia izan liteke: *Miren maldijo la serpiente* → *Mirenek sugea madarikatu zuen*. Beraz, esaldiak osorik hartuta bat bestea baino zuzenagoa den arren, guk, xede honetarako, biak jo ditugu berdintzat, azken batean bietatik aterako baikenuke patroi berbera: *ENTI_0+S_ERG sugea madarikatu zuen*.

4.1.1 Sailkapena itzulpenen egokitasunaren arabera

Espero izatekoa den bezala, hiztegiko itzulpena makinarena baino nabarmen hobea da gehienetan, eta kasu horiei bakarrik erreparatu diegu gure bilaketarako, automatikoki ondo itzultzen diren esaldientzat patroiak sortzeak ez baitu sistema hobetuko. Beraz, adibideak alde batera utzi ditugu, honako kasu hauetan:

- Makinaren itzulpena hiztegiko itzulpena baino egokiagoa iruditu bazaigu.
- Hiztegiko euskarazko adibidea eta makinaren itzulpena berdinak badira, edo, berdin-berdinak ez izan arren, biak egokitzen jo baditugu.
- Hiztegiko itzulpena bestea baino hobea izan arren, itzulpen automatikoa ere onargarria iruditu bazaigu.

Honako bi taula hauetan, multzo bakoitzean zenbat lagin sailkatu ditugun ikus daiteke, gaztelaniatiko eta ingelesetiko itzulpen-adibideen artean:

	Adibide kopurua	Ehunekoa
Hiztegiko itzulpena makinarena baino hobea	68	% 51,51
Bi itzulpenak berdinak, edo biak egokiak	50	% 37,87
Hiztegiko itzulpena hobea, baina makinarena ere onargarria	10	% 6,82
Makinaren itzulpena hiztegikoa baino egokiagoa	4	% 3,03
Guztira	132	% 100

1. taula: Gaztelaniatiko itzulpenen sailkapena

	Adibide kopurua	Ehunekoa
Hiztegiko itzulpena makinarena baino hobea	372	% 88,57
Bi itzulpenak berdinak, edo biak egokiak	36	% 8,57
Hiztegiko itzulpena hobea, baina makinarena ere onargarria	12	% 2,85
Makinaren itzulpena hiztegikoa baino egokiagoa	0	% 0
Guztira	420	% 100

2. taula: Ingelesetiko itzulpenen sailkapena

Zenbakiak argi erakusten dutenez, Matxinek askoz ere emaitza onargarri gehiago ematen ditu gaztelaniatik itzultzean ingelesetik itzultzean baino: gaztelaniazko % 48,48 gutzira, ingelesezko % 11,42ren aldean. Ingelesetiko itzulpen-sistemak patroien behar handiagoa duela iruditu zaigu, eta horregatik hartu ditugu adibide gehiago ingeles-euskara hiztegitik gaztelania-euskara hiztegitik baino.

4.1.2 Baztertutako adibideak

Lehen fase honen ondorio gisa, gaztelania-euskarazko adibideetako 64 (% 48,48) eta ingeles-euskara adibideetako 48 (% 11,42) alde batera utzi ditugu, eta zenbaki bana eman diegu, baztertzearen zergatiaren arabera:

- Akatsa entitatean dago eta patroï bat sortzeak ez luke arazoa konponduko, edo adibide osoa izen berezi bat da [1].
- Makinaren itzulpena egokiagoa iruditu zaigu [2].
- Hiztegiko itzulpena eta automatikoa berdina dira, edo berdina ez izan arren, biak iruditu zaizkigu egokiak [3].
- Hiztegiko adibideetako itzulpena hobea izan arren, Matxinen ere onargarria da [4].

A eranskinean, baztertu ditugun adibide guztiak bildu ditugu, bi taulatan, euren itzulpen automatikoarekin eta baztertzearen zergatiari dagokion zenbakiarekin batera. Mota bakoitzean zenbat adibide sailkatu ditugun 4.1.1 ataleko 1. eta 2. taulatan ikus daiteke.

4.2 Adibideen erabilera-maiztasuna

Lehen bazterketa eginda ere adibideen zerrenda nahiko zabala zenez, bigarren ezaugarri bati erreparatu diogu: erabilera-maiztasunari. Gure helburua sistemaren akats larri samarrak zuzentzea da, erabiltzaileek ez dezaten oso emaitza traketsik jaso, itzulpen-sistemari esaldi nahiko oinarrizko bat emanda. Baina lehenago ere esan bezala, ez da komeni patroï gehiegi definitzea, horrek nabarmen motelduko bailuke itzulpen-prozesua. Horregatik, sarrien erabiliko diren patroïak zein diren aztertu nahi izan dugu.

4.2.1 Maiztasuna aztertze prozedura

Argi dago ezin dela zehatz-mehatz jakin zer itzuli nahi izango duten erabiltzaileek, baina corpus handietan bilaketak eginez gero, gutxi-gorabeherako ideia bat izatea lor daiteke. Guk corpus zatituta erabiltzea erabaki dugu, askotariko testuak eskaintzen baititu, gai anitzen ingurukoak eta hainbat erregistro mailatakoak. Hala ere, kontuan hartzekoa da Interneten dagoen informazioa etengabe ari dela aldatzen, eta une batean egindako bilaketen emaitzak desberdinak izan daitezkeela handik denbora gutxira.

Adibideetako esaldi-pareetatik jatorrizko hizkuntzakoak hartu ditugu –gaztelaniazkoak eta ingelesezkoak–, eta Google bilatzailean sartu ditugu, emaitza kopuruari begiratzeko. Googlen bilaketa egitea arriskutsua izan daiteke, landu gabeko corpusa denez, ezin baita aukeratu zer motatako hitzak bilatu nahi diren, eta hortaz, emaitza asko ez dira guk lortu nahi genituenak izaten (datozen azpiataletan, izandako arazoen inguruan arituko gara). Hala ere, anotatutako corpusak erabili beharrean Google erabiltzea erabaki dugu. Batetik, testu gehien biltzen dituen aukera da eta, beraz, ikuspegi zabalena izatea ahalbidetzen duena; eta bestetik, informazio gaurkotuena bertan aurki daiteke. Gainera, ez zen gure asmoa adibideak zehatz-mehatz zenbat erabiltzen diren jakitea, baizik eta kopuru orientagarri batzuk izatea, adibideak itzulpen-sistamarako erabilgarriak izango diren ala ez erabakitzen lagun diezaguten.

Gure azken helburua aldagai bat edo gehiago dituzten patroïak sortzea denez, ez ditugu esaldiak bere horretan bilatu, baizik eta patroïetan aldagai izango diren entitateak alde batera utzita. Esate baterako:

(6) Ingelesezko adibidea: *Paris and its environs*

Bilatu duguna: “*and its environs*”

Gure asmoa ez da *Paris and its environs* hitz-segidak zenbat emaitza ematen dituen jakitea; jakin nahi duguna da zenbatetan ageri den izen berezi bat *and its environs* hitzen aurretik. Googlek ez duenez ematen hori zuzenean bilatzeko aukerarik, entitate-izena alde batera utzi eta hurrengo hitz-segida bilatu dugu. Hiztegietan agertzen diren adibideek hitzen erabilera islatu nahi izaten dutenez, nahiko esaldi arruntak jasotzen dituzte normalean, eta pentsatzekoa da antzeko esaldiak ere nahiko maiz erabiltzen direla. Hala eta guztiz ere, batetik, hori ez da beti betetzen, eta bestetik, hori horrela izanda ere, askotan ez ditugu guk nahi dugun motako emaitzak bakarrik jaso.

Seigarren adibidearen kasuan, Googlek ematen dituen lehen hamar emaitzak bat datoz guk jaso nahi genuenarekin. Honako hauek agertzen dira gure adibidean *Paris* zegoen tokian: *New Orleans, Rome, Boston, Svolveer, Badajoz, Jerusalem, Chicago, Monrovia, Putney* eta *Jerusalem*. Horren arabera, ondoriozta dezakegu guk bilatutakoak toki-izen bati jarraituko diola gehienetan eta, hortaz, jaso dugun emaitza nahiko fidagarria dela. Beste batzuetan, ordea, bilaketa egitea ez da hain erraza izan (ikus 4.2.2 atala).

4.2.2 Zailtasunak eta konponbideak

Hiru arazo nagusi izan ditugu adibideen maiztasuna neurtzerakoan: patroïko aldagai bat zati finkoen artean agertzea, zenbakizko adierazpenak aldagaitzat jarri behar izatea, eta emaitzak fidagarriak ez izatea. Datozen lerroetan azalduko dugu zehatzago zer-nolako zailtasunekin egin dugun topo eta zer egin dugun horiek konpontzeko.

4.2.2.1 Aldagai bat zati finkoen artean

Esaldi eta esaldi-zati gehienetan, seigarren adibidean bezala, zer bilatu jakiteak ez digu aparteko zailtasunik eman; beste kasu batzuetan, ordea, bilaketa ez da hain erraza izan, entitatea ez baitzegoen esaldi edo esaldi-zatiaren hasieran edo bukaeran. Esaterako:

- (7) Gaztelaniazko adibidea: *la factoría de Seat en Manresa*

Bilatu duguna: “*la factoría de*en*”

Horrelako kasuetan, baldin eta aldagaia izen berezi bat bazen, entitatea izartxo batez ordeztu dugu, aurreko eta atzeko hitzen artean zuriunerik utzi gabe, bilatzaileak zuriunerik gabeko edozein karaktere-segida bila zezan.

4.2.2.2 Zenbakizko adierazpenak

Aldagaia izen berezi bat izan beharrean zenbaki bat bazen, Googleri adierazi diogu zenbakiak bakarrik bilatzea nahi genuela:

- (8) Gaztelaniazko adibidea: *una capacidad de 10 litros*

Bilatu duguna: “*una capacidad de 0.. litros*”

Adierazpen horren bidez, zehaztu daiteke 0 baino handiagoa den edozein zenbaki bilatu nahi dela. Esate baterako, 8. adibideko sintagma bilatuta, lehen hamar emaitzetan, honako hauek jasotzen dira 0.. adierazpena zegoen lekuan: 4.000, 60, 5, 60, 624, 100, 7, 16.000, 80 eta 13.500.

4.2.2.3 Emaitza ez fidagarriak

Izan dugun beste arazo bat da esaldi edo esaldi-zati batzuen zati finkoa –Googlen bilatu duguna, alegia– bakarrik ere erabil daitekeela, eta emaitzetan ez dela ia entitate-izenik agertzen. Horixe gertatzen da honako segida honekin:

- (9) Gaztelaniazko adibidea: *sistema de vídeo VHS*

Bilatu duguna: “*sistema de vídeo*”

Bilatzaileak bueltatzen dizkigun lehen hamar emaitzen artean, testu-zati bakarrean ageri da guk sartutako hitz-segida izen berezi baten aurretik: *sistema de vídeo Presentation Transcript*. Gainontzekoetan, ordea, ez da hala gertatzen: *esquema de un sistema de vídeo, sistema de vídeo vigilancia ciudadana...*

Horrelakoetan, Googleren emaitza kopurua ez da oso fidagarria gure azterketarako, jasotzen diren testu-zatiak ez baitira gure adibidea bezalakoak. Beraz, emaitza kopuruaz gain, ezaugarri hori ere kontuan hartu dugu (ikus 4.2.3 atala).

4.2.3 Aukeraketa

Lehenago ere esan dugun bezala, ezinezkoa da zehatz-mehatz jakitea adibideetatik ateratako patroïak zenbateraino izango diren erabilgarriak, eta, Googlen bilaketak egitean, gure asmoa zifra orientagarri batzuk izatea besterik ez da izan, oso zaila baita inongo laguntzarik gabe esatea gure esaldi eta esaldi-zatiak erabiltzen diren ala ez. Hori dela eta, lortutako zifren arabera egin dugu gure hautapena, jakin badakigun arren bilaketan emaitza kopurua ez dela beti fidagarria.

Bigarren fase honen ondorio gisa, beste bazterketa bat egin dugu. Bai gaztelania-euskara eta bai ingelesa-euskara adibideen artetik gutxi-gorabehera laurden bat hautatzea erabaki dugu: gaztelania-euskara hiztegiko 32 esaldi-pare (bilaketetan 700.000 emaitzatik gora eman dituztenak) eta ingelesa-euskara hiztegiko 99 (30.000.000 emaitzatik gorakoak).

Aukeratu ditugun laginen inguruan datozen ataletan arituko gara; emaitza kopuruagatik baztertutako adibideak, berriz, 5. eta 6. tauletan bildu ditugu (ikus B eranskina). Horrez gain, beste bederatzi adibide ere alde batera utzi ditugu –gaztelaniazko hiru eta ingelesezko sei–, emaitza kopurua oso fidagarria ez zelakoan:

- (10) Gaztelaniazko adibidea: *a Atenas vía Roma*
Hiztegiko itzulpena: *Atenasera, Erromatik barrena*
- (11) Gaztelaniazko adibidea: *Sistema de vídeo VHS*
Hiztegiko itzulpena: *VHS bideo-sistema*
- (12) Gaztelaniazko adibidea: *¿vas a Tolosa? Yo también*
Hiztegiko itzulpena: *Tolosara zoaz? Ni ere bai*
- (13) Ingelesezko adibidea: *a Strauss waltz*
Hiztegiko itzulpena: *Strauss-en balsa*
- (14) Ingelesezko adibidea: *are you going to Tolosa? So am I*
Hiztegiko itzulpena: *Tolosara zoaz? Ni ere bai*
- (15) Ingelesezko adibidea: *Bohemian crystal*
Hiztegiko itzulpena: *Bohemiako kristala*
- (16) Ingelesezko adibidea: *is that a Picasso?*
Hiztegiko itzulpena: *Picasso bat da hori?*
- (17) Ingelesezko adibidea: *Seat's Manresa plant*
Hiztegiko itzulpena: *Seat-en Manresako faktoria*
- (18) Ingelesezko adibidea: *to Athens via Rome*
Hiztegiko itzulpena: *Atenasera, Erromatik barrena*

Denek ematen dituzte emaitza asko, baina jasotzen ditugun testuetan ez da guk nahi duguna agertzen. Esaterako, 16. adibidean, “*is that a*” bilatuta, oso kasu gutxitan agertzen da segida hori izen berezi baten aurretik. Pentsatu dugu jasotako kopuruak ez zetzela bat benetan lortu nahi genituenekin, eta hobe zela kontuan ez hartzea.

Iruditu zaigu adibide horietako bat ere ez zela oso erabilgarria izango patroïak sortzeko, eta denak alde batera uztea erabaki dugu.

4.3 Patroïen sorkuntza

Patroïak sortzeko adibiderik egokienak zein diren aztertu ostean, taula bat osatu dugu hizkuntza-bikote bakoitzerako (ikus C eranskina), hautatu ditugun adibideak eta haien itzulpenak –bai hiztegikoa, bai automatikoa– jasotzeko. Ondoren, taula horiek oinarritzat hartuta, patroïak sortzeari ekin diogu.

4.3.1 Entitate-detektatzailearen zuzenketak

Adibideetatik patroïak sortzeko lana erdi-automatikoki egin dugu (erabilitako tresnei buruzko informazioa, 3.2 atalean). Entitate-detektatzaileak automatikoki markatu digu izen berezia edo zenbakia non zegoen, eta horrez gain, euskararen kasuan, baita haren postposizio-informazioa zein zen ere. Dena den, prozesu automatiko hori ere gainbegiratu egin behar da, ez baitu beti emaitza zuzena ematen. D eranskinean bildu ditugu egindako zuzenketa guztiak, eta akatsa zertan egin den: entitatearen detekzioan [1] ala postposizio-informazioaren detekzioan [2]. Datozen lerroetan adibide batzuk emango ditugu.

4.3.1.1 Entitatearen detekzioan egindako akatsak

Detektatzaileak ongi hautematen ditu entitateak gehienetan, baina, batzuetan, arazoak izaten ditu hitz bakarrekoak ez diren izen bereziak detektatzeko (euskarazko zortzi adibidetan eta ingelesezko bitan zuzendu dugu akats hori):

- (19) Hiztegiko adibidea: *San Joan sua*
 Patroï automatikoa: *ENTI_0 Joan sua*
 Patroï zuzendua: *ENTI_0 sua*

Beste lau kasutan, izen berezi bat ez da entitatetzat hartu (euskarazko adibide bat, gaztelaniazko beste bat, eta ingelesezko bi). Hona hemen horietako bat:

- (20) Hiztegiko adibidea: *el Tajo desemboca en el Atlántico*
 Patroï automatikoa: *el ENTI_0 desemboca en el Atlántico*
 Patroï zuzendua: *el ENTI_0 desemboca en el ENTI_1*

Azkenik, euskarazko bi adibidetan, bi izen berezi entitate bakartzat hartu dira, bata bestearen alboan zeudelako:

- (21) Hiztegiko adibidea: *Tajo Atlantikoan itsasoratzen da*
 Patroi automatikoa: *ENTI_0+S_INE itsasoratzen da*
 Patroi zuzendua: *ENTI_0+S_ABS ENTI_1+S_INE itsasoratzen da*

4.3.1.2 Postposizio-informazioa hautematean egindako akatsak

Postposizio-informazioa jasotzeko orduan, entitate-detektatzaileak zenbaki deklinatuekin izan du arazo gehien (zortzi patroï sortu ditu oker), ez baitu postposizio-informazioa entitatearen partetzat hartu. Adibidez:

- (22) Hiztegiko adibidea: *% 10eko beherapena*
 Patroi automatikoa: *% ENTI_0eko beherapena*
 Patroi zuzendua: *% ENTI_0+S_GEL beherapena*

Zenbakia ongi detektatu arren, ez da postposizio-informazioa jaso. Patroi horren bidez itzulpen zuzenak lor genitzake kasu batzuetan, baina baita okerrak ere (zenbakien deklinabideari buruzko informazioa Euskaltzaindiaren 8. arauan⁹ jasota dago). Izan ere, zenbakia aldatuz gero, honelako itzulpenak sor litezke: *% 3eko beherapena*. Beraz, komenigarria da entitateari dagokion informazioa gehitzea.

Beste bitan, detektatzailea nahastu egin da izen berezi baten kasuarekin. Hemengo adibide honetan, esaterako, izenaren amaiera datiboaren markaren berdina zelako gertatu da: *-ri*.

- (23) Hiztegiko adibidea: *Jauregiberri anderea*
 Patroi automatikoa: *ENTI_0+S_DAT anderea*
 Patroi zuzendua: *ENTI_0 anderea*

4.3.2 Patroiak hobetzeko proposamenak

Aurreko atalean aipatutako zuzenketez gain, beste aldaketa batzuk ere egitea erabaki dugu, patroïak zuzenak izan arren, eraginkorragoak izan zitezkeela iruditu baitzaigu.

4.3.2.1 Gaztelania-euskara patroïetarako proposamenak

Gaztelania-euskara adibideen artean, lau kasutan, iruditu zaigu lagin osoa hartu beharrean zati bat hartzeak maizago agertzen diren egiturak zuzentzeko aukera emango ligukeela. Datozen lerroetan azalduko dugu kasu horiek zein diren eta zer proposatu dugun guk, patroïen eraginkortasuna hobetze aldera.

⁹http://www.euskaltzaindia.net/index.php?option=com_euskaltzaindiarenarauak&view=frontpage&Itemid=424&lang=eu

- (24) Gaztelaniazko adibidea: *finalmente me admitieron en la **universidad de Zaragoza***
 Hiztegiko itzulpena: *azkenean **Zaragozako unibertsitatean** onartu ninduten*
 Gaztelaniazko patroï automatikoa: *finalmente me admitieron en la **universidad de ENTI_0***
 Euskarazko patroï automatikoa: *azkenean **ENTI_0+S_GEL unibertsitatean** onartu ninduten*

Beharbada, esaldi osoa ez da askotan agertuko bere horretan, baina itzultzaile automatikoak esaldiaren zati jakin batean egiten du akatsa (universidad de ENTI_0), eta zati hori, hain zuzen, oso sarri agertzen da bilaketetan. Beraz, egokiagoa iruditu zaigu patroia zati horretara mugatzea:

Gaztelaniazko patroï proposatua: *universidad de ENTI_0*
 Euskarazko patroï proposatua: *ENTI_0+S_GEL unibertsitatea*

Gauza bera gertatzen da beste adibide honekin ere:

- (25) Gaztelaniazko adibidea: ***llegué con 5 minutos de anticipación***
 Hiztegiko itzulpena: ***5 minutu lehenago** iritsi nintzen*
 Gaztelaniazko patroï proposatua: *con ENTI_0 minutos de anticipación*
 Euskarazko patroï proposatua: *ENTI_0 minutu lehenago*

Esaldi osoa kontuan hartuz gero, patroia erabili ahal izateko, derrigorrezkoa izango litzateke aditza lehen pertsona singularrean eta etorkizunean egotea. Beltzez dagoen zatia soilik hartuta, ordea, patroia erabili ahal izango da hitz horiek beste edozein aditzekin batera doazenean ere, beste pertsona eta denbora batean egon arren (*llegaron, saldréis, iremos...*).

Honako bi adibide hauekin ere antzeko zerbait gertatzen da. Arazoa aurreko esaldietakoen berbera da, baina konpontzeko, euskarazko esaldia gehixeago aldatu behar izan dugu:

- (26) Gaztelaniazko adibidea: *la puerta tiene **80 cm de ancho***
 Hiztegiko itzulpena: *atea 80 cm zabal da*
 Gaztelaniazko patroï proposatua: *ENTI_0 cm de ancho*
 Euskarazko patroï proposatua: *ENTI_0 cm-ko zabalera*
- (27) Gaztelaniazko adibidea: *la sala tiene **un aforo de 300 personas***
 Hiztegiko itzulpena: *aretoaren edukiera 300 pertsonakoa da*
 Gaztelaniazko patroï proposatua: *un aforo de ENTI_0 personas*
 Euskarazko patroï proposatua: *ENTI_0 pertsonako edukiera*

Patroiak ongi funtzionatuz gero, Matxinek bi itzulpen hauen antzekoak sortuko ditu: *ateak 80 cm-ko zabalera du eta aretoak 300 pertsonako edukiera du*. Itzulpenak hiztegikoak baino literalagoak dira, baina, era horretan, kasu gehiagotan erabili ahal

izango dira, beste subjektu batzuekin agertuta ere balioko baitute; adibidez: *la ventana tiene 50 cm de ancho*, edo *el bar tiene un aforo de 200 personas*.

Beste bi kasutan, berriz, hiztegiko adibidea bera iruditu zaigu aldatzekoa. Batean, hitz bat ez zaigulako oso egokia iruditu, eta bestean, hitz jakin bat erabiltzeak arazoak eman ditzakeela pentsatu dugulako. Azal ditzagun bi kasuak zehatzago:

(28) Gaztelaniazko adibidea: *el jorobado de Notre-Dame*

Hiztegiko itzulpena: *Notre-Dameko konkorra*

Gaztelaniazko patroi automatikoa: *el jorobado de ENTI_0*

Euskarazko patroi proposatua: *ENTI_0+S_GEL konkorduna*

Gure ustez, gaztelaniazko *jorobado* hitzaren ordain egokiagoa da *konkordun*, eta ez *konkor*. Izan ere, hiztegietan biak azaltzen badira ere, Googlen bilatuta argi ikusten da *Notre-Dameko konkorduna* erabiliagoa dela (17.900 emaitza) *Notre-Dameko konkorra* baino (5 emaitza), ziur asko, haurrentzako film ezagunaren izenburua hala dagoelako itzulita.

Bestalde, honako adibide honen patroi automatikoak arazoak eman litzakeela pentsatu dugu:

(29) Gaztelaniazko adibidea: *los menores de 14 años*

Hiztegiko itzulpena: *14 urtez azpiko adingabeak*

Gaztelaniazko patroi automatikoa: *los menores de ENTI_0*

Euskarazko patroi proposatua: *ENTI_0 urtez azpikoak*

Uste dugu *adingabe* hitza kentzea egokiagoa dela, bestela, itzulpen xelebreak sor litezke eta. Demagun *los menores de 30 años* dugula, 30 urtetik beherakoei erreferentzia egiteko. Sistemak *30 urtez azpiko adingabeak* itzultzea ez litzateke egokia izango, 25 urteko pertsona bat, adibidez, ez baita adingabea. Hortaz, *ENTI_0 urtez azpikoak* jartzea erabaki dugu, edozein zifrarekin ongi funtzionatuko duelakoan.

4.3.2.2 Ingelesa-euskara patroiatarako proposamenak

Gaztelania-euskara adibideetan bezala, pentsatu dugu ingelesa-euskara adibideetako patroi batzuetatik zati bat bakarrik hartzea egokiagoa zela. 24-27. adibideetan azaldu dugun moduan, askotan hobe izaten da esaldi-zati txikiak zuzenduko dituzten patrioiak sortzea esaldi osoak definitzea baino, gehiagotan erabili ahal izaten baitira. Beraz, patrioen eraginkortasuna handitzeko asmoz, honako adibide hauetatik zati bat baino ez dugu hartu:

(30) Ingelesezko adibidea: *finally I was admitted to the University of Zaragoza*

Hiztegiko itzulpena: *azkenean Zaragozako unibertsitatean onartu ninduten*

Ingeleseko patroi proposatua: *University of ENTI_0*

Euskarazko patroi proposatua: *ENTI_0+S_GEL unibertsitatea*

- (31) Ingeleseko adibidea: ***no, John is not here. Then he must be at the bar.***

Hiztegiko itzulpena: ***ez, John ez dago hemen. Orduan tabernan egongo da.***

Ingeleseko patroi proposatua: *no, ENTI_0 is not here*

Euskarazko patroi proposatua: *ez, ENTI_0+S_ABS ez dago hemen*

- (32) Ingeleseko adibidea: ***Hello? Is Ainara at home?***

Hiztegiko itzulpena: ***Bai, esan? Ainara etxean da?***

Ingeleseko patroi proposatua: *is ENTI_0 at home?*

Euskarazko patroi proposatua: *ENTI_0 etxean da?*

Horrez gain, beste bost patroitik ere aldaketa txiki bat egin dugu, itzulpena hobetu zitekeelakoan. Lehena honako hau da:

- (33) Ingeleseko adibidea: ***the banks of the Seine***

Hiztegiko itzulpena: *Sena ertzean*

Ingeleseko patroi proposatua: ***on the banks of the ENTI_0***

Euskarazko patroi automatikoa: *ENTI_0 ertzean*

Euskarazko adibideak *non?* galderari erantzuten dio, baina ingelesezkoak ez. Gure ustez, euskarazko patroiak inesibo-marka eramanez, beharrezkoa da ingelesezkoak *on* preposizioa izatea, hala, ingelesezko adibideak ere *where?* galderari erantzungo bailioke.

Beste patroi honetan, aldiz, lexiko mailako aldaketa bat egin dugu:

- (34) Ingeleseko adibidea: ***the rule of Henry VII***

Hiztegiko itzulpena: ***Henry VIIaren erregealdia***

Ingeleseko patroi automatikoa: *the rule of ENTI_0*

Euskarazko patroi proposatua: ***ENTI_0+S_GEN agintaldia***

Hiztegiko itzulpena guztiz egokia da, eta, esaldi osoa hartuz gero, ez dago ezer aldatu beharrik. Arazoa da *erregealdi* hitzak, patroi horretan, halabeharrez eskatzen duela errege baten izena aurretik, eta *the rule of* erabil daitekeela, esaterako, presidenteen izenekin ere. Beraz, *agintaldi* hitza jartzea erabaki dugu, zabalagoa denez, kasu gehiagotarako balioko baitu.

Beste itzulpen honetan, herrialde-izen bati edutezko genitiboa gehitzen zaio, eta, gure ustez, lekuzko genitiboa egokiagoa da. Gainera, Googlen ezezik, Ereduzko Prosa

Gaur corpusean¹⁰ ere sarriago agertzen da bigarren egitura (hamar aldiz gehiagotan, gutxi-gorabehera). Hortaz, *-ren* atzizkiaren orde, *-ko* jartzea erabaki dugu:

(35) Ingeleseko adibidea: *in the north of France*

Hiztegiko itzulpena: *Frantziaren iparraldea*

Ingeleseko patroï automatikoa: *in the north of ENTI_0*

Euskarazko patroï proposatua: *ENTI_0+S_GEL iparraldea*

Azkenik, hiztegiko bi itzulpen hauetan, euskarazko esaldia indartua da, eta pentsatu dugu patroïetarako egokiagoa zela esaldi neutroagoak erabiltzea:

(36) Ingeleseko adibidea: *Kevin loves you*

Hiztegiko itzulpena: *Kevinek zu maite zaitu*

Ingeleseko patroï automatikoa: *ENTI_0 loves you*

Euskarazko patroï proposatua: *ENTI_0+S_ERG maite zaitu*

(37) Ingeleseko adibidea: *he ended up in Madrid*

Hiztegiko itzulpena: *non eta Madrilen bukatu zuen*

Ingeleseko patroï automatikoa: *he ended up in ENTI_0*

Euskarazko patroï proposatua: *ENTI_0+S_INE bukatu zuen*

Itzulpenek berek ez dute inongo arazorik, baina, testuinguruaren arabera, arraroak izan daitezke. Demagun bi galdera hauek ditugula 36. adibideko esaldiarekin batera: *nor maite du Kevinek?* eta *nork maite zaitu?* Ingelesez, elementu bat indartu daiteke ahoz, baina idatzizko erantzuna berdina da bi galderetarako: *Kevin loves you*. Euskaraz, ordea, arraro samarra da bigarren galderari *Kevinek zu maite zaitu* esaldiaz erantzutea. Beraz, ahalik eta formarik neutroena erabiltzen saiatu gara.

37. adibidearekin ere gauza bera egin dugu. Iruditu zaigu *non eta Madrilen* esateak harridura-kutsua ematen diola esaldiari eta, gehienetan, egitura hori espero ez den zerbaiti buruz hitz egiterakoan erabiltzen dela. Kasu horretan ere forma neutroago bat proposatu dugu, testuingurua edozein dela ere arazorik eman ez dezan.

4.3.3 Alde batera utzitako adibideak

Adibideak aztertzen ari ginela, iruditu zaigu patroïetako batzuk ez zirela oso egokiak eta hobe zela baztertzea. Honako honetan, esaterako, iruditu zaigu ingelesezko adibidetik sortuko genukeen patroïa orokorregia dela euskarazko itzulpen bakarra emateko:

(38) Ingeleseko adibidea: *at 30*

Hiztegiko itzulpena: *30 urterekin*

¹⁰ Kontsultak egiteko: <http://www.ehu.es/euskara-orria/euskara/ereduzkoa/>

At preposizioa era askotara itzul daiteke, baita atzetik zenbaki bat duenean ere. Adinak adierazteko balio izateaz gain, ordu eta temperaturekin ere erabiltzen da, besteak beste; adibidez: *water freezes at 0 degrees and boils at 100*. Esaldiaren bukaeran patroia erabiliz gero, sistemak itzulpen bitxi honen antzeko zerbait sor lezake: *ura 0 gradutan izozten da eta 100 urterekin irakiten du*.

Beste lau adibide hauetan, euskarazko patrioiak gramatikalki okerrak izan litezke, segun eta zein testuingurutan erabiltzen diren:

(39) Ingeleseko adibidea: *with the exception of Martin*

Hiztegiko itzulpena: **Martin** izan ezik

(40) Ingeleseko adibidea: *located in Paris*

Hiztegiko itzulpena: **Parisen kokatua**

(41) Ingeleseko adibidea: *made in China*

Hiztegiko itzulpena: **Txinan egina**

Esate baterako, 39. adibidea zuzena izan liteke honako kasu honetan: *everybody went home, with the exception of Martin*. Izan ere, euskaraz honela itzul daiteke: *denak etxera joan ziren, Martin izan ezik*. Aurretik beste zerbait izanez gero, ordea, itzulpen okerra sor liteke: *she has been to every country in Europe with the exception of Portugal*. Euskarazko itzulpenean, beharrezkoa da inesiboa markatzea: *Europako herrialde guztietan izan da, Portugalen izan ezik*. Hala dio Euskaltzaindiaren 16. arauak *izan ezik* adierazpenari buruz:

“Erabil bedi formula osoa, *izan ezik*, eta ez *ezik* bakarrik, eta aurreko sintagmak har beza dagokion kasu marka.”

40. eta 41. adibideekin ere antzeko zerbait gertatzen da. Erreferente bakarraz ari bagara, egokia da *Parisen kokatua* edo *Txinan egina* esatea, baina ez erreferente bat baino gehiago ditugunean. Jo dezagun eskultura multzo batez ari garela; ingelesez *made in China* esango genuke, baina euskarazko esaldiak pluralgilea behar luke: *Txinan eginak*.

Honako beste adibide hau ere aldakorra da testuinguruaren arabera:

(42) Ingeleseko adibidea: *to return in 4 days*

Hiztegiko itzulpena: *4 egun barru itzuli*

Liburutegiko liburu bat hartu eta ohar batean adibideko esaldia jartzen badu, euskarazko itzulpena egokia litzateke: *4 egun barru itzuli*. Esaldi luzeagoetan, ordea, euskaraz ezingo genuke patroia beti erabili: *she told him to return in 4 days (4 egun barru itzultzeko esan zion)*.

Azkenik, hiru kasutan, euskarazko bi adibide zeuden ingelesezko adibide bakarrarentzat. Makinak arazoak emango lituzke ingelesezko patroi berberarentzat itzulpen bat baino gehiago balitu; beraz, kasu horietan, bi aukeretako bat hautatu dugu, bestea alde batera utzita.

4.3.4 Patroi gehigarriak

Adibideen artean, datak hirutan agertu zaizkigu, eta horietako bat erabilera-maiztasunik handieneko adibideen zerrendan sailkatu dugu:

(43) Ingelesezko adibidea: *5 December 2003*

Hiztegiko itzulpena: *2003ko abenduaren 5a*

Iruditu zaigu datak adierazteko egitura hori sarri erabiltzen dela, eta komenigarria litzatekeela patroi gehiago ere sortzea. Beraz, hilabete bakoitzeko patroi bat definitu dugu (ikus 5. atala).

Gauza bera egin dugu puntu kardinalak barne hartzen dituzten hiru adierazpenekin ere. Hona hemen bi adibide (hirugarrenak 44. adibidekoaren egitura berbera du):

(44) Ingelesezko adibidea: *in the west of Spain*

Hiztegiko itzulpena: *Espainiako mendebaldean*

(45) Ingelesezko adibidea: *south of France*

Hiztegiko itzulpena (zuzenduta; ikus 35. adibidea, 4.3.2.2 atalean): *Frantziako hegoaldea*

Hiztegiko adibideen artean, behin baino gehiagotan ageri dira horrelako egiturak, eta gainera, oso ohikoak direla iruditu zaigu. Hori dela eta, patroi bakarra sortu beharrean, puntu kardinal bakoitzeko bana sortu dugu (ikus 5. atala).

4.3.4.1 Aldagairik gabeko segidak

Orain arte sortu ditugun patroi guztiek aldagairen bat dute hiztegiko esaldi eta esaldizatiatan izen berezi bat edo zenbaki bat zegoen tokian. Honako bi adibide hauek, ordea, desberdinak iruditu zaizkigu:

(46) Ingelesezko adibidea: *glory to the Father, the Son and the Holy Spirit*

Hiztegiko itzulpena: *loria Aitari, eta Semeari, eta Izpiritu Sainduari*

(47) Ingelesezko adibidea: *Merry Christmas and a Happy New Year!*

Hiztegiko itzulpena: *Zorionak eta Urte Berri On!*

Patroi gisa, hau da, entitate-izenak kontuan hartu gabe, ez dute emaitza asko ematen bilaketetan, baina bere horretan bilatuta oso sarri agertzen dira. Iruditu zaigu

egokia izan litekeela segida-pare finko batzuk sortzea kasu horietarako, aldagaidun patroïak izan beharrean.

4.3.5 Entitate moten zehaztapena

Azkenik, patroïetan entitateak markatzeaz gain, entitate horiek zer motatakoak diren ere zehaztu dugu, patroï batzuk okerrak izan baitaitezke entitate mota jakin batekin erabiltzen ez badira. Adibidez:

- (48) Ingelesezkotroïa: *the adventures of ENTI_0*
Euskarazkotroïa: *ENTI_0+S_GEN abenturak*

Patroï hori pertsona-izenentzat pentsatuta dago. Izan ere, toki-izen bat jarritz gero, itzulpen okerra sortuko luke, honelako zer bait: *Ingalaterraren abenturak*, *Ingalaterrako abenturak* beharrean.

Eihera entitate-detektatzaileak hiru entitate mota bereizteko aukera ematen duenez, guk ere hiru mota horien arabera sailkatu ditugu gure patroïetako entitateak, eta horrez gain, zenbakiak ere markatu ditugu. Beraz, lau entitate mota bereizi ditugu guztira: erakunde-izenak [E], leku-izenak [L], pertsona-izenak [P] eta zenbakiak [Z]. Eraitzen atalean, zutabe bat gehitu dugu tauletan entitateen mota zehazteko (ikus 5. ataleko 12. eta 13. taulak).

5 Emaitzak

Egindako azterketaren ondoren, gaztelania-euskara itzulpen-sistamarako baliagarri izan daitezkeen 30 patroï (adibide guztien % 22,73) proposatzen ditugu, denak hiztegiko adibideetatik zuzenean hartuak. Ingelesa-euskara sistamarako, berriz, beste 104 patroï: adibideetatiko 88 (adibide guztien % 20,95) eta geuk gehituta, 16.

Ondoko taulotan bildu ditugu proposatzen ditugun patroï guztiak:

Gaztelaniazko patroia	Euskarazko patroia	Entitate mota
¡viva ENTI_0!	gora ENTI_0+S_ABS!	L/P/E
con ENTI_0 minutos de anticipación	ENTI_0 minutu lehenago	Z
el canal de ENTI_0	ENTI_0+S_GEL kanala	L
el canal de la ENTI_0	ENTI_0+S_GEL kanala	L
el desierto de ENTI_0	ENTI_0+S_GEL basamortua	L
el ENTI_0 aniversario de la boda	ezkontzaren ENTI_0. urtemuga	Z
el ENTI_0 desemboca en el ENTI_1	ENTI_0+S_ABS ENTI_1+S_INE itsasoratzen da	0: L 1: L
el estrecho de ENTI_0	ENTI_0+S_GEL itsasartea	L
el jorobado de ENTI_0	ENTI_0+S_GEL konkorduna	L
el principio de ENTI_0	ENTI_0+S_GEN printzipioa	P
el reino de ENTI_0	ENTI_0+S_GEL erresuma	L
ENTI_0 cm de ancho	ENTI_0 cm-ko zabalera	Z
ENTI_0 está en cinta	ENTI_0+S_ABS haurdun dago	P
hora de cierre: ENTI_0	itxiera-ordua: ENTI_0	Z
la dama de ENTI_0	ENTI_0+S_GEL dama	L
la factoría de ENTI_0 en ENTI_1	ENTI_0+S_GEN ENTI_1+S_GEL faktoria	0: E 1: L
la hoguera de ENTI_0	ENTI_0 sua	E
la orquesta filarmónica de ENTI_0	ENTI_0+S_GEL orkestra filarmonikoa	L
la paga de ENTI_0	ENTI_0+P_GEL ordainsaria	E
la selva de ENTI_0	ENTI_0+S_GEL basoa	L
la situación en ENTI_0	ENTI_0+S_GEL egoera	L
la unidad de los pueblos de ENTI_0	ENTI_0+S_GEL herrien batasuna	L
las cataratas de ENTI_0	ENTI_0+S_GEL ur-jauziak	L
los menores de ENTI_0 años	ENTI_0 urtez azpikoak	Z
paso de ENTI_0	ENTI_0+S_GEL lepoa	L
reducción de un ENTI_0%	% ENTI_0+S_GEL beherapena	Z
un aforo de ENTI_0 personas	ENTI_0 eserlekuko edukiera	Z
un peso de ENTI_0 kilos	ENTI_0 kiloko pisua	Z
una capacidad de ENTI_0 litros	ENTI_0 litroko edukiera	Z
universidad de ENTI_0	ENTI_0+S_GEL unibertsitatea	L

12. taula: Gaztelania-euskara sistamarako proposatzen ditugun patroïak

Ingeleseko patroia	Euskarazko patroia	Entitate mota
a certain ENTI_0 called	ENTI_0 delako batek deitu du	P
a poem by ENTI_0	ENTI_0+S_ERG egindako olerkia	P
a speed of ENTI_0 mph	ENTI_0 milia orduko abiadura	Z

a subsidiary of ENTI_0	ENTI_0+S_GEN filiala	E
a war with ENTI_0	ENTI_0+S_GEN aurkako gerra	L
a weight of ENTI_0 kg	ENTI_0 kiloko pisua	Z
above ENTI_0 degrees	ENTI_0 gradutik gora	Z
an angle of ENTI_0 degrees	ENTI_0 graduko angelua	Z
an increase of ENTI_0%	% ENTI_0+S_GEL igoera	Z
are you going to ENTI_0?	ENTI_0+S_ALA zoaz?	L
are you related to ENTI_0?	ENTI_0+S_GEN ahaidea zara?	P
around ENTI_0 people	ENTI_0 lagun inguru	Z
as ENTI_0 put it	ENTI_0+S_ERG esan zuen bezala	P
at ENTI_0 mph	ENTI_0 milia orduko abiadura	Z
back in ENTI_0	ENTI_0+S_INE	Z
between ENTI_0 and ENTI_1 kilos	ENTI_0 eta ENTI_1 kilo bitartean	0: Z 1: Z
between ENTI_0 and ENTI_1 years	ENTI_0 eta ENTI_1 urte bitartean	0: Z 1: Z
beyond ENTI_0 years	ENTI_0 urtetik gora	Z
children of ENTI_0 and over	ENTI_0 urtetik gorako haurrak	Z
children under the age of ENTI_0	ENTI_0 urtetik beherako haurrak	Z
Dr ENTI_0, I presume?	zu ENTI_0 doktorea izango zara, ezta?	P
east of ENTI_0	ENTI_0+S_GEL ekialdea	L
ENTI_0 and ENTI_1 met in the final	ENTI_0+S_ABS eta ENTI_1+S_ABS elkarren aurka aritu ziren finalean	0: P 1: P
ENTI_0 and the surrounding area	ENTI_0 eta bere inguruak	L
ENTI_0 April ENTI_1	ENTI_1+S_GEL apirilaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 August ENTI_1	ENTI_1+S_GEL abuztuaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 December ENTI_1	ENTI_1+S_GEL abenduaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 desert	ENTI_0+S_GEL basamortua	L
ENTI_0 falls	ENTI_0+S_GEL ur-jauziak	L
ENTI_0 February ENTI_1	ENTI_1+S_GEL otsailaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 hours a day	eguneko ENTI_0 ordu	Z
ENTI_0 is in his room	ENTI_0 bere gelan dago	P
ENTI_0 January ENTI_1	ENTI_1+S_GEL urtarrilaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 July ENTI_1	ENTI_1+S_GEL uztailaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 June ENTI_1	ENTI_1+S_GEL ekainaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 loves you	ENTI_0+S_ERG maite zaitu	P
ENTI_0 March ENTI_1	ENTI_1+S_GEL martxoaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 May ENTI_1	ENTI_1+S_GEL maiatzaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 November ENTI_1	ENTI_1+S_GEL azaroaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 October ENTI_1	ENTI_1+S_GEL urriaren ENTI_0+S_ABS	0: Z 1: Z
ENTI_0 September ENTI_1	ENTI_1+S_GEL irailaren ENTI_0+S_ABS	0: Z 1: Z

ENTI_0 venue	ENTI_0+P_GEN egoitza	E
ENTI_0% off	% ENTI_0+S_GEL beherapena	Z
give me ENTI_0 minutes	emaizkidazu ENTI_0 minutu	Z
have you ever been to ENTI_0?	izan zara egundaino ENTI_0+S_INE?	L
he ended up in ENTI_0	ENTI_0+S_INE bukatu zuen	L
how far is ENTI_0 from ENTI_1?	zer distantzia dago ENTI_0 eta ENTI_1 artean?	0: L 1: L
I didn't go and neither did ENTI_0	ni ez nintzen joan eta ENTI_0+S_ABS ere ez	P
I live in ENTI_0	ENTI_0+S_INE bizi naiz	L
I lived in ENTI_0 for a time	ENTI_0+S_INE bizi izan nintzen denboraldi batez	L
I lived in ENTI_0 for a while	ENTI_0+S_GEN bizi izan nintzen denboraldi batez	L
I love ENTI_0, especially in the spring	izugarri gustatzen zait ENTI_0+S_ABS, batez ere udaberrian	L
I love ENTI_0, especially in the summer	ENTI_0 gustuko dut, udan gehienbat	L
I make it ENTI_0	ENTI_0 direla esango nuke	Z
I met ENTI_0 in ENTI_1	ENTI_1+S_INE ezagutu nuen ENTI_0+S_ABS	0: P 1: L
in ENTI_0	ENTI_0+S_INE	Z/L
in ENTI_0, ENTI_1, or wherever	ENTI_0+S_INE, ENTI_1+S_INE edo auskalo non	0: L 1: L
in the east of ENTI_0	ENTI_0+S_GEL ekialdea	L
in the north of ENTI_0	ENTI_0+S_GEL iparraldean	L
in the south of ENTI_0	ENTI_0+S_GEL hegoaldea	L
in the west of ENTI_0	ENTI_0+S_GEL mendebaldean	L
is ENTI_0 at home?	ENTI_0 etxean da?	P
I've never been in ENTI_0	ez naiz behin ere ENTI_0+S_INE izan	L
I've never been to ENTI_0	ez naiz inoiz ENTI_0+S_INE egon	L
Lady of ENTI_0	ENTI_0+S_GEL dama	L
mainland ENTI_0	ENTI_0 kontinentala	L
member state of the ENTI_0	ENTI_0+S_GEL estatua	E/L
Mister ENTI_0	ENTI_0 jauna	P
more than ENTI_0	ENTI_0tik gora	Z
Mr ENTI_0	ENTI_0 jauna	P
Mrs ENTI_0	ENTI_0 andrea	P
no, ENTI_0 is not here	ez, ENTI_0+S_ABS ez dago hemen	P
north of ENTI_0	ENTI_0+S_GEL iparraldea	L
on the banks of the ENTI_0	ENTI_0 ertzean	L
on the eve of the ENTI_0	ENTI_0+S_GEN bezperan	E
once we went to ENTI_0	behin ENTI_0+S_ALA joan ginen	L
quotations from ENTI_0	ENTI_0+S_GEN aipuak	P
round ENTI_0 o'clock	ENTI_0+P_ABS aldera	Z
she was born in ENTI_0	ENTI_0+S_INE jaio zen	L
south of ENTI_0	ENTI_0+S_GEL hegoaldea	L
The Adventures of ENTI_0	ENTI_0+S_GEN abenturak	P
the border between ENTI_0 and ENTI_1	ENTI_0+S_GEN eta ENTI_1+S_GEN arteko muga	0: L 1: L
the children under ENTI_0	ENTI_0 urtez azpiko adingabeak	Z
the Diary of ENTI_0	ENTI_0+S_GEN egunkaria	P
the ENTI_0 AIDS Foundation	ENTI_0+S_GEL HIES Fundazioa	L

the ENTI_0 flows into the ENTI_1	ENTI_0+S_ABS ENTI_1+S_INE itsasoratzen da	0: L 1: L
the ENTI_0 goes round the ENTI_0	ENTI_0+S_ABS ENTI_1+S_GEN inguruan biratzen da	0: E 1: E
the ENTI_0 metro	ENTI_0+S_GEL metroa	L
the fall of the ENTI_0	ENTI_0+S_GEN gainbehera	E
the rule of ENTI_0	ENTI_0+S_GEN agintaldia	P
the works of ENTI_0	ENTI_0+S_GEN lanak	P
they left for ENTI_0	ENTI_0+S_ABZ joan dira	L
they say ENTI_0 is rich	ENTI_0 aberatsa omen da	P
to be ENTI_0 by birth	jatorriz ENTI_0+S_GEL_ABS izan	L
university of ENTI_0	ENTI_0+S_GEL unibertsitatea	L
upwards of ENTI_0 years	ENTI_0 urtetik gora	Z
we did ENTI_0 in ENTI_1 weeks	ENTI_0 ENTI_1 astean ikustatu genuen	0: L 1: Z
we walked ENTI_0 kilometres	ENTI_0 kilometroko bidea egin genuen oinez	Z
west of ENTI_0	ENTI_0+S_GEL mendebaldea	L
what is the distance between the ENTI_0 and the ENTI_1?	zein da ENTI_0+S_ABL ENTI_1+S_ABU_GEL distantzia?	0: E 1: E
what is the population of ENTI_0?	zenbat biztanle ditu ENTI_0+S_ERG?	L
when I left ENTI_0	ENTI_0+S_ABL kanporatu nintzenean	L
which of you is ENTI_0?	zuetariko nor da ENTI_0+S_ABS?	P
you saw the ENTI_0?	ENTI_0 berbera ikusi duzula?	P

13. taula: Ingelesa-euskara sistemarako proposatzen ditugun patroïak

Horrez gain, 4.3.4.1 atalean azaldu dugun moduan, aldagairik gabeko bi segida-pare ere sor daitezkeela pentsatu dugu. Honako hauek dira:

Ingeleszko patroïa	Euskarazko patroïa
glory to the Father, the Son and the Holy Spirit	loria Aitari, Semeari eta Izpiritu Sainduari
Merry Christmas and a Happy New Year	Zorionak eta Urte Berri On

14. taula: Aldagairik gabeko segida-pareak, ingelesa-euskara

6 Ondorioak eta etorkizunerako lanak

Proiektu honen helburu nagusia zen, hiztegietako adibideak oinarritzat hartuta, aztertzea ea zein esaldi-pare izan zitezkeen baliagarriak Itzulpen Automatikorako patroi orokor batzuk sortzeko. Patroiak gerora erregeletan oinarritutako sistema bat aberasteko erabili nahi direnez, hasteko, sistemak oker itzultzen dituen esaldi eta esaldi-zatiak bakarrik hartu nahi izan ditugu (gaztelania-euskara hiztegiko adibideen % 51,51 eta ingelesa-euskara hiztegikoen % 88,57).

Horrez gain, kontuan izanik patroi gehiegi erabiltzeak itzulpen-sistema asko motelduko lukeela, adibideen erabilera-maiztasuna aztertu dugu, gaztelaniazko eta ingelesezko esaldi eta esaldi-zatiek Google bilatzailean ematen zituzten emaitzen arabera. Hala, adibide guztien % 25 inguru hautatu ditugu (gaztelania-euskarazko 32 eta ingelesa-euskarazko 102).

Azkenik, patroiak sortzeko, hiru entitate-detektatzailez baliatu gara, izen bereziak eta zenbakiak aldagaitzat marka zitzaten. Automatikoki sortutako patroiak hobetze aldera, hainbat zuzenketa eta aldaketa egin dizkiegu, eta azkenik, gaztelania-euskara sistemarako 30 patroi eta ingelesa-euskara sistemarako 88 sortu ditugu. Horrez gain, adibideetako batzuk oso baliagarriak izan litezkeelakoan, horietatik beste 16 patroi ere ateratzea erabaki dugu, denak ingelesetiko itzulpen-sistemarako (104 patroi dira, beraz, guztira).

Bestalde, aldagairik gabeko bi segida-pare ere definitu ditugu, bere horretan asko itzul litezkeelakoan. Aurrera begira, ondo legoke maiz erabiltzen diren eta hitzez hitz itzuli ezin diren adierazpenentzako patroi gehiago ere definitzea.

6.1 Etorkizunerako lanak

Itzulpen Automatikorako emankorrak iruditu zaizkigun patroiak hautatu ditugu, eskuz nagusiki, zenbait hizkuntza-tresna ere erabili ditugun arren. Hurrengo urratsa patroi horiek itzulpen-sistemetan integratzea eta ondo funtzionatzen duten aztertzea izango da, horixe izan baitzen lan honi ekiteko motibazioa. Itzulpen-prozesuan, patroiak jatorrizko analisía egin aurretik hautematea da gure asmoa. Hala, sistemak hitz-segida bat patroi batekin lotzen badu, Matxini esango dio zati hori ez itzultzeko erregelen bidez.

Dena den, aintzat hartu behar da patroiak emaitza onak emateko, entitate-detektatzaileen lanak beti zuzenak izan behar direla. Errealitatean, ordea, hori ez da gertatzen, tresna automatiko horiek errore-tasa bat izaten baitute. Beraz, baliteke

patroïek guk lortu nahi ditugun emaitzak ez ematea, eta kasu horretan, egokitzapen batzuk egin beharko dira, akatsak non egin diren aztertuta.

Landutako adibide guztien artean % 22 inguru jo ditugu Itzulpen Automatikorako patroï baliagarriak sortzeko iturburu egokitzat, eta eskura ditugun adibide guztiak aztertzea komeni dela ondorioztatu dugu. Lan horri ekin beharko zaio eta, ahal bada, aukeraketa hori modu erdi-automatikoan egiten saiatuko gara, automatikoki bereiziz emankorrak izan daitezkeen patroïak eta gainontzekoak.

Gure lan honetan, izen bereziren bat edo zenbakiren bat zuten adibideak aztertu, eta osagai horiek aldagaien bidez orokortu ditugu, patroï orokortuak gehiago erabiliko direlakoan. Hala ere, entitaterik gabeko patroï ugari ere sortu daitezke eskura ditugun adibideetatik, eta lagungarriak izan daitezke horiek ere itzultzaile automatikoen kalitatea hobetu ahal izateko.

Bestalde, aipagarria da lan honetan ez ditugula patroïak makinaren akatsen arabera sailkatu, baina interesgarria litzateke, etorkizunera begira, hori ere egitea. Izan ere, Matxin oraindik garapen-prozesuan dagoenez, baliteke erregela berriak sortuz lan honetan aztertutako itzulpen oker batzuk konpontzea eta, ondorioz, hemen proposatutako patroï batzuk, gerora, beharrezkoak ez izatea. Une honetan, sortutako patroïak erabilgarriak izango direlakoan gaude, baina sistemak aurrera egin ahala, aztertu beharko da ea horietako batzuk soberan dauden.

7 Erreferentziak

- Aduriz, I., Aldezabal, I., Alegria, I., Artola, X., Ezeiza, N. and Urizar, R. (1996). EUSLEM: A Lemmatiser / Tagger for Basque. In: *Proc. EURALEX'96, Part 1*, 17-26. Góumlteborg, Sweden.
- Alegria, I., Arregi, O., Ezeiza, N., Fernandez, I. and Urizar, R. (2004). Design and Development of a Named Entity Recognizer for an Agglutinative Language. In: *First International Joint Conference on NLP (IJCNLP-04)*, Hainan Island, China.
- Alegria, I. and Urkia, M. (2002). *Morfologia konputazionala. Euskararen morfologiaren deskribapena*. UEU.
- Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., and Padró, M. (2006). Freeling 1.3: Syntactic and semantic services in an open-source nlp library. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy.
- Díaz de Ilarraza, A., Mayor, A., Sarasola, K. (2002). Adibideetan Oinarritutako Itzulpen Automatikoa (AOIA): azterketa bibliografikoa eta euskararekin lan egiteko proposamenak. *Senez*, 25 (pp. 17-35).
- Ebling, S., Way, A., Volk, M. and Kumar Naskar, S. (2011). *Combining semantic and syntactic generalization in example-based machine translation*. In: *The 15th conference of the European Association for Machine Translation (EAMT 2011)*, Leuven, Belgium.
- España-Bonet, C., Labaka, G., Díaz de Ilarraza, A., Mårquez, L. and Sarasola, K. (2011). Hybrid Machine Translation Guided by a Rule-Based System. In *MT Summit XIII: the Thirteenth Machine Translation Summit organized by the Asia-Pacific Association for Machine Translation AAMT* (pp. 553-561). Xiamen, China.
- Farrús, M., Iglesias, G., Artola, X., Díaz de Ilarraza, A., Labaka, G., Lersundi, M., Mayor, A., Sarasola, K. (2008). Strategies for sustainable MT for Basque: incremental design, reusability, standarization and open-source. In *Proceedings of the IJCNLP-08* (pp. 235-243). Hyderabad, India.
- Finkel, J. R., and Manning, C. D. (2009, May). Joint parsing and named entity recognition. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 326-334). Association for Computational Linguistics.
- Hutchins, J. (2007). Machine translation: A concise history. *Computer aided translation: Theory and practice*.
- Kitamura, M. and Murata, T. (2003) Practical Machine Translation on System allowing Complex Patterns. In: *Proceedings of MT Summit IX*.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University. Press, Cambridge.
- Labaka, G. (2010). *EUSMT: Incorporating Linguistic Information into SMT for a Morphologically Rich Language. Its use in SMT-RBMT-EBMT hybridation*. UPV/EHU.

- Labaka, G., Díaz de Ilarraza, A., España-Bonet, C., Sarasola, K. and Màrquez, L. (2011). Deep evaluation of hybrid architectures: simple metrics correlated with human judgments. In *International Workshop on Using Linguistic Information for Hybrid Machine Translation LIHMT* (pp. 50-57). Universitat Politècnica de Catalunya, Barcelona.
- Mayor, A. (2008). *Matxin: Erregeletan oinarritutako itzulpen automatikoko sistema baten eraikuntza estaldura handiko baliabide linguistikoak berrerabiliz*. UPV/EHU.
- Mayor, A., Alegría, I., Díaz de Ilarraza, A., Labaka, G., Lersundi, M. and Sarasola, K. (2009). Matxin, euskararako lehenengo itzultzaile automatikoa. *SENEZ Itzulpen Aldizkaria*, 37, (pp. 197-220).
- Nagao, M. (1984) *A framework of a mechanical translation between Japanese and English by analogy principle*. Alick Elithorn & Ranan Banerji (eds.) *Artificial and Human Intelligence*, Amsterdam, North-Holland.
- Padró, L., Collado, M., Reese, S., Lloberes, M., and Castellón, I. (2012). *Freeling 2.1: Five years of open-source language processing tools*. UPC.
- Salam, K. M. A., Yamada, S. and Nishino, T. (2011). Example-Based Machine Translation for Low-Resource Language Using Chunk-String Templates. In: *13th Machine Translation Summit*, Xiamen, China.

Eranskinak

A eranskina: lehen fasean baztertutako adibideak

Gaztelaniazko lagina (Elhuyar hiztegia)	Euskarazko lagina (Elhuyar hiztegia)	Itzulpen automatikoa (Matxin)	Zergatia
Baja Navarra	Nafarroa Beherea	Nafarroa beherea	1
Bellas Artes	Arte Ederrak	Arte ederrak	1
Bilbao está bien comunicado	Bilbo ondo komunikaturik dago	Bilbao ondo dago komunikatuta	1
Comisión Europea	Europako Batzordea	Europako batzordea	1
Compañía de Jesús	Jesusen Lagundia	Jesusen lagundia	1
Departamento de Medio Ambiente	Ingurumen Saila	Ingurumen-saila	1
Dios maldijo la serpiente	Jainkoak sugea madarikatu zuen	Diosek sugea madarikatu zuen	1
el Espíritu Santo	Espiritu Santua	Espiritu santua	1
el Gran Cañón del Colorado	Coloradoko Arroila Handia	Gran Cañón De El Colorado	1
el padre Donostia	aita Donostia	Aita Donostia	1
el Peñón de Gibraltar	Gibraltarko Haitza	Peñón De Gibraltar	1
el Tratado de Roma	Erromako Ituna	Erromako ituna	1
Facultad de Derecho	Zuzenbide Fakultatea	Zuzenbide-fakultatea	1
Fondo Monetario Internacional	Nazioarteko Diru Funtsa	Diru-funts Internacional	1
Hechos de los Apóstoles	Apostoluen eginak	Apostoluen Eginak	1
la Alta Navarra	Nafarroa Garaia	Nafarroa Garai	1
la Luna gira alrededor de la Tierra	Ilargia Lurraren inguruan biratzen da	Lunak Tierraren inguruan biratzen du	1
los Diez Mandamientos de Dios	Jainkoaren Hamar Aginteak	10 Mandamientos De Dios	1
los Reyes Magos	Errege Magoak	Errege magoa	1
Organización de las Naciones Unidas	Nazio Batuen Erakundea	Nazio Batuen erakundea	1
Oriente Próximo	Ekialde Hurbila	Ekialde hurbila	1
San Ignacio	Inazio Donea	San Ignacio	1
Sancho el fuerte	Antso Azkarra	Gogorra Sancho	1
Santa Ana	Santa Ana	Santa Ana	1
Secretaría General	Idazkaritza Nagusia	Idazkaritza nagusia	1
trompas de Falopio	Falopioren tronpak	Falopioren tronpak	1
VIII. Congreso de Euskaltzaindia	Euskaltzaindiaren VIII. Biltzarra	VIII. Congreso De Euskaltzaindia	1
cuando salí fuera de Francia	Frantziatik kanporatu nintzenean	Frantziatik kanpora atera nintzenean	2
estado miembro de la Unión Europea	Europar Batasuneko estatua	Europar Batasuneko estatu kidea	2
Mikel está viendo la televisión	Mikel telebista ikusten dago	Mikel telebista ikusten ari da	2
un tal Stalin	Stalin delako morroi	Stalin delako bat	2

	batek		
a orillas del Sena	Sena ertzean	Senaren ertzean	3
delta del Nilo	Niloren delta	Niloren delta	3
el día de Nochebuena es el 24 de diciembre	Gabon-eguna abenduaren 24a da	Gabon-eguna abenduaren 24a da	3
el diario de Ana Frank	Ana Frank-en egunkaria	Ana Franken egunkaria	3
el ministro de Sanidad	Osasun-ministroa	Osasun Ministroa	3
en la zona de Tolosa	Tolosako partean	Tolosako partean	3
fórmula 1	1 formula	1 formula	3
fuera de Euskal Herria	Euskal Herritik landa	Euskal herritik kanpora	3
hoy he quedado con Idoia	gaur Idoiarekin elkartzekoa naiz	Gaur Idoiarekin geratu naiz	3
la aparición de la Virgen	Ama Birjinaren agerkundea	Virgenen agerpena	3
la decadencia del Imperio romano	Erromatar Inperioaren gainbehera	Inperio erromatarraren gainbehera	3
la descendencia de Jacob	Jakoben ondorea	Jakoben ondorea	3
la ONU enviará observadores	NBEk begiraleak bidaliko ditu	NBE-k behatzaileak bidaliko ditu	3
Las aventuras de Tom Sawyer	Tom Sawyer-en abenturak	Tom Sawyerren abenturak	3
las obras de Mitxelena	Mitxelenaren lanak	Mitxelenaren lanak	3
le dicen Manolo	Manolo esaten diote	Manolo esaten diote	3
los ingleses que se marcharon a América	Ameriketara joan ziren ingelesak	Amerikara joan ziren ingelesak	3
mar Mediterráneo	Mediterraneo itsasoa	Mediterraneo itsasoa	3
pasando por Toulouse	Tolosan gaindi	Tolosatik pasatzean	3
Pedro es un vinagre	Pedro ozpina da	Pedro ozpin bat da	3
reproductor de DVD	DVD irakurgailua	DVD-irakurgailua	3
según Arquímedes...	Arkimedesen arabera...; Arkimedesek dioenez...	Arkimedesen arabera...	3
sucesor de San Pedro	San Pedroren ondorengoa	San Pedroren ondorengoa	3
¡girar el timón 10 grados a babor!	lema 10 gradu ababorrera okertu!	Lema biratzea 10 gradu ababorrera!	4
conocí a Mikel en Bilbao	Bilbon ezagutu nuen Mikel	Mikel ezagutu nuen Bilbaon	4
el norte de Francia	Frantziako iparraldea	Frantziaren iparraldea	4
encontrarse en Roma	Erroman elkarrekin gertatu	Erroman aurkitzea	4
la caída del Imperio romano	Erromatar Inperioaren erorialdia	Imperio erromatarraren erorketa	4
la habitación mide 2 metros cuadrados	gelak 2 metro karratu ditu	Gelak bi metro karratu neurtzen ditu	4
sita en Paris	Parisen kokatua	Parisen dagoen	4
su nombre es Izaskun pero la llaman Ixaxkun	Izaskun deitzen da baina Ixaxkun deitzen diote	Haren izena Izaskun da baina Ixaxkun deitzen dute	4
todos los precios llevan incluido el Impuesto sobre el Valor Añadido	salneurri guztiek BEZa barne dute	Prezio guztiek Balio Erantsiaren Gaineko Zerga hartua daramate	4

Zezeaga juega en la zaga	Zezeagak atzealdean jokutzen du	Zezeagak atzean jokutzen du	4
--------------------------	---------------------------------	-----------------------------	---

3. taula: Lehen fasean baztertutako adibideak, gaztelania-euskara

Ingeleseko lagina (Elhuyar hiztegia)	Euskarazko lagina (Elhuyar hiztegia)	Itzulpen automatikoa (Matxin)	Zergatia
Alexander the Great	Alexandro Handia	Handia Alejandro Alejandromagno	1
Basque Studies Society	Eusko Ikaskuntza	@basque study societyak	1
Cape Horn	Hornos lurmuturra	@cape horna	1
Cicero's oratory	Zizeronen oratoria	Ciceroren oratoria	1
English Channel	Mantxako kanala	@english channela	1
Father Donostia	aita Donostia	@father donostia	1
in the Tolosa area	Tolosako partean	@tolosa aldean	1
John introduced himself	Jonek bere burua aurkeztu zigun	Komunak bere burua aurkeztu zuen	1
miss China	miss Txina	@miss china	1
Mount Everest; Mt Everest	Everest mendia	@mount Everesta @mt Everesta	1
Panama Canal	Panamako kanala	Kanal Panama	1
Saint Anne; St. Anne	Santa Ana	St. Anne Deun @anne;	1
Saint Thomas Fair	Santo Tomasko feria	Deun @tomas Fairra	1
Sancho the Strong	Antso Azkarra	Indartsua @sancho	1
Society of Jesus	Jesusen Lagundia	Jesuseko Gizartea	1
St. Ignatius	Inazio Donea	St. Ignatius	1
the 8th Congress of the Euskaltzaindia	Euskaltzaindiaren VIII. Biltzarra	@euskaltzaindia @8th Congressa	1
The Boston Philharmonic Orchestra	Bostongo Orkestra Filarmonikoa	@boston Philharmonic Orchestra	1
the Grand Canyon of Colorado	Coloradoko Arroila Handia	Coloradoko @grand Canyona	1
the Indian Ocean	Indiako ozeanoa	@indian Oceana	1
the Mediterranean Sea	Mediterraneo itsasoa	@mediterranean Sea	1
the River Thames	Tamesis ibaia	@river Thamesa	1
the Rock of Gibraltar	Gibraltarko Haitza	@gibraltar Arroka	1
the Ten Commandments	Jainkoaren Hamar Aginteak	Aginduak Hamar	1
the Treaty of Rome	Erromako Ituna	Erromako Hitzarmena	1
the Tropic of Capricorn	Kaprikornio tropikoa	@tropic Of Capricorna	1
the Wise Men	Errege Magoak	@wise Mena	1
United Arab Emirates	Arabiar Emirerri Batuak	@united arab emiratesak	1
United Nations	Nazio Batuen Erakundea	Nazio Batuen erakundeak	1
in Rome	Erroman	Erroman	3
inhabitants of the Netherlands	Herbehereetako herritarrak	Herbehereetako biztanleak	3
Jacob's descendants	Jakoben ondorea	Jakoben ondorengoak	3
situation in Iraq	Irakeko egoera	Irakeko egoera	3
the north of France	Frantziako iparralde	Frantziako iparralde	3
to meet in Rome	Erroman elkarrekin	Elkartu Erroman	3

	gertatu		
we saw the Argentinian plains	Argentinako zabaldiak ikusi genituen	Guk lautada Argentinarrak ikusi genituen	3
European countries	Europako lurraldeak	Herri europarrak	4
he amassed a fortune in America	Ameriketan dirutza egin zuen	Hark dirutza bat metatu zuen Amerikan	4
he has gone over to France	Frantziara joan da	Hura joan da Frantziara	4
next, we visited Paris	ondoren, Paris bisitatu genuen	Guk Paris bisitatu genuen gero,	4
the Basque colony in Paris	Parisko euskaldunen kolonia	Parisko kolonia Euskalduna	4
the flight alone costs 1,000 euros	hegaldiak soilik 1.000 euro balio du	Hegaldia 1,000 euro kostatzen da bakarrik	4
the French Way	Frantziar bidea	Bide Frantsesa	4
the kingdom of Castile	Gaztelako erresuma	Gaztelako erreinua	4
the show will be on in Barcelona very soon	ikuskizuna laster izango da Bartzelonan	Ikuskizuna izango da Bartzelonan oso laster	4
this road goes to Rome	bide honek Erromara darama	Bide hau doa Erromara	4
to establish in London	Londresen kokatu	Finkatu Londresen	4
we've travelled around France	Frantziatik zehar bidaiatu dugu	Guk bidaiatu dugu Frantziatik zehar	4

4. taula: Lehen fasean baztertutako adibideak, ingelesa-euskara

B eranskina: agerraldi kopuruagatik baztertutako adibideen taulak

Gaztelaniazko esaldia (Elhuyar)	Euskarazko esaldia (Elhuyar)	Googleren emaitza kopurua
Creo que Mikel es tonto	Mikel tontoa dela iruditzen zait	598.000
¿has estado alguna vez en Marruecos?	Izan zara egundaino Marokon?	598.000
60 euros y pico	60 euro eta koska	583.000
No he estado nunca en París	Ez naiz behin ere Parisen izan	459.000
En América amasó una fortuna	Ameriketean dirutza egin zuen	402.000
Los siete sabios de Grecia	Greziako zazpi jakintsuak	387.000
Gloria al Padre, al Hijo y al Espíritu Santo	Loria Aitari, eta Semeari, eta Izpiritu Sainduari	275.000
Aquí yace Mikel Goiburu	Hemen datza Mikel Goiburu	234.000
Cocer al baño María	Maria bainuan egosi; Maria bainutan egosi	192.000
Los festivales musicales de Edimburgo	Edinburgh-eko musika-jaialdiak	103.000
El pueblo más antiguo de Europa	Europako populu zaharrena	83.300
La cuenca del río Deba	Deba ibaiaren arroa	17.800
Pasé unas vacaciones estupendas en Cuba	Oporraldi ederra egin nuen Kuban	5.080
En las hermosas llanuras de Mesopotamia	Mesopotamiako ordoki ederretan	4.360
La temperatura en verano oscila entre los 20 y los 35 grados	Udan tenperatura 20 eta 35 gradu artean dabil	2.360
el cuadrado de 3 es 9	3ren karratua 9 da	2.100
5 más 2 igual a 7	5 gehi 2 berdin 7	1.870
8 por 4 igual a 32	8 bider 4 berdin 32	1.460
2 por 2 igual a 4	2 bider 2 4	1.460
el cubo de 2 es 8	2ren kubo 8 da	942
20 menos 8 igual a 12	20 ken 8 berdin 12	561
7 menos 2 igual a 5	7 ken 2 berdin 5	561
12 dividido por 2 igual a 6	12 zati 2 berdin 6	280
El símbolo del hierro es Fe	Burdinaren sinboloa Fe da	169
En esta botella caben 5 litros	Botila honek 5 litro hartzen ditu	70
ETA ha reivindicado el último atentado	ETAk azken atentatua errebindikatu du	55
Tendremos que suplir a Martín con otro	Martinen orde z besteren bat ipini beharko dugu	47
La Rioja es famosa por la excelencia de sus vinos	Errioxa famatua da bere ardoen bikaintasuna dela medio	43
La colonia de los vascos en París	Parisko euskaldunen kolonia	9
Se fue a América, donde amasó una fortuna	Ameriketara joan, eta han aberastu egin zen	7
Los quechuas viven en los Andes	Kitxuak Andetan bizi dira	4
Fue mécido en Getxo, donde murió...	Mediku izan zen Getxon, non hil baitzen	4
Mikel y los demás irán mañana	Mikel eta bihar joango dira	0

5. taula: Agerraldi kopuruagatik baztertutako adibideak, gaztelania-euskara

Ingeleseko esaldia (Elhuyar)	Euskarazko esaldia (Elhuyar)	Googleren emaitza kopurua
I lived in London as a child	Haurretan Londresen bizi nintzen	27.200.000
the Roman Empire stretched from North Africa to Britain	Erromatar Inperioa Afrikaren iparraldetik Britainia Handiraino hedatzen zen	26.700.000
London is my home town	Londres da nire jaioterria	25.900.000
is Mikel in?	Mikel etxean dago?; Mikel hor dago?	25.270.000
Paris and its environs	Paris eta bere inguruak	24.900.000
the Straits of Gibraltar	Gibraltarko itsasartea	24.100.000
we had a good time in Benidorm	Benidormen primeran ibili ginen	23.400.000
The Quechua people live in the Andean region	Kitxuak Andean bizi dira	23.300.000
100 km per hour	100 km orduko	21.300.000
b for Bilbo	Bilboren b	21.200.000
passing through Toulouse	Tolosan gaindi	21.100.000
600 dollars per month	600 dolar hilabeteko	21.000.000
she's been to China	Txinan ibili da	20.900.000
this is where we met Tom	Hementxe topatu genuen Tom	19.200.000
many of us have been to America	Gutako asko egon gara Ameriketara	18.500.000
long live Wales!	Gora Gales!	17.800.000
The United Kingdom took possession of the Falkland Islands again	Erresuma Batuak Malvinak beretu ditu berriro ere	17.700.000
3 degrees below zero	3 gradu zero azpitik	17.100.000
Amy can't go and I can't either	Amy ezin da joan eta ni ere ez	15.200.000
John F. Kennedy was assassinated in 1963	John F. Kennedy 1963an erail zuten	14.300.000
I sat down between Mikel and Jon	Mikel eta Jonen artean eseri nintzen	14.300.000
Tim is behind me	Tim atzean dut	13.900.000
she sat next to John	Johnen ondoan eseri zen	11.000.000
how far is it to Rome?	Zer distantzia/zenbat dago Erromara?	9.600.000
those were brought from China	Haiek Txinatik ekarriak dira	8.930.000
Luisa is pregnant	Luisa haurdun dago	8.560.000
I've saved almost £100 so far	Orain arte ia 100 libera aurreztu ditut	8.550.000
Finland is cold	Finlandian hotz egiten du	7.890.000
the English people who went to America	Ameriketara joan ziren ingelesak	7.120.000
it rained all over England	Euria egin zuen Ingalaterra osoan zehar	6.530.000
beyond 2007	2007tik aurrera	6.230.000
Edinburgh's music festivals	Edinburgh-eko musika-jaialdiak	5.470.000
he fled to Australia after the war	Australiara ihes egin zuen gerra ostean	5.430.000
Laura doesn't like it and neither do I	Laurari ez zaio gustatzen eta ezta niri ere	5.380.000
the Guggenheim Museum is a landmark of Bilbao	Guggenheim Museoa Bilboko oso toki ezaguna da	5.330.000
2 weeks back	Orain dela 2 aste	5.290.000
the Colossus of Rhodes	Rodasko kolosoa	5.270.000
hello, is Bob there please?	Kaixo, Bob hor dago?	4.700.000
Jesus died out of love	Jesus maitasunez hil zen	3.840.000

Jon got married last year	Iaz ezkondu zen Jon	3.630.000
the bonfire of the Eve of St. John	San Joan sua	3.630.000
Maria showed us her house	Mariak bere etxea erakutsi zigun	3.570.000
they named her Ruth	Ruth ipini zioten izena	3.210.000
she's been dating Mikel	Mikelekin ibili da	3.010.000
Tom's elder sister	Tom-en arreba nagusia	3.010.000
how many times have you been to Elorrio?	Zenbatetan joan zara Elorriora?	2.650.000
the distribution of wealth in Europe	Aberastasunen banaketa European	1.940.000
they call him Manolo	Manolo esaten diote	1.680.000
the apparition of the Virgin Mary	Ama Birjinaren agerkundea	1.670.000
here lies Mikel Goiburu	Hemen datza Mikel Goiburu	1.670.000
it's 3 metres in height	3 metroko goiera du	1.630.000
the unity of the peoples of America	Ameriketako herrien batasuna	1.530.000
warmest thanks to Mr Smith	Gure eskerrik beroena Smith jaunari	1.530.000
Mikel seems silly to me	Mikel tontoa dela iruditzen zait	1.480.000
can you tell Tom from his twin brother?	Bereizten dituzu Tom eta bere anaia bikia?	1.430.000
Mikel is watching television	Mikel telebista ikusten dago	1.400.000
to empty in Biscay	Bizkaian itsasoratu	1.250.000
Osaka is Japan's second-largest city	Osaka da Japoniako hiri handienen artean bigarrena	1.120.000
the seven wise men of Greece	Greziako zazpi jakintsuak	1.070.000
99% of the workers were out	Langileen % 99k greba egin zuten	1.050.000
water freezes at 0°C	Ura 0 °C-an izozten da	1.030.000
Leire introduced herself	Leirek bere burua aurkeztu zigun	975.000
I couldn't connect with the Internet	Ezin izan nintzen Internetera konektatu	872.000
many people converged on Bilbao	Jende piloak elkartu zen Bilbon	800.000
Europe's most ancient people	Europako populu zaharrena	765.000
more than 50 people died	50 pertsonatik gora hil ziren	712.000
ETA has claimed responsibility for the latest attack	ETAk azken atentatua erreibindikatu du	680.000
name all the states of the USA	Esan AEBko estatu guztien izenak	676.000
the film was shot on location in England	Filmaren kanpoaldeak Ingalaterran filmatu ziren	650.000
two point six (2.6)	Bi koma sei (2,6)	635.000
the journey up from Munich	Munichetik iparralderako bidaiak	623.000
Rontgen discovered X-rays	Rontgen-ek X izpiak aurkitu zituen	597.000
old Mr Brown and young Mr Brown	Mr Brown aita eta Mr Brown semea	555.000
this bottle takes 5 litres	Botila honek 5 litro hartzen ditu	524.000
Amaia bought it herself	Amaiak berak erosi du	496.000
a 10% discount	% 10eko beherapena	486.000
have you ever been to Paris? No, never	Izan zara inoiz Parisen? Ez, inoiz ez	450.000
somebody called Stalin	Stalin delako morroi batek	397.000
I'm very excited about the tripto Jamaica	Erabat hunkituta nauka Jamaikarako bidaiak	380.000
I had a wonderful holiday in Cuba	Oporraldi ederra egin nuen Kuban	367.000
Charles Dickens is a nationally known writer	Charles Dickens idazlea nazio osoan ezaguna da	284.000
the square of 3 is 9	3ren karratua 9 da	278.000
Pello arrived very angry	Pello hagitz haserre etorri zen	256.000
on the beautiful plains of Mesopotamia	Mesopotamiako ordoki ederretan	253.000

turn to page 19	Ireki liburua 19. orrialdean	237.000
closing time: 19.00	Itxiera-ordua: 19:00	233.000
easy ways to aid Africa	Afrika laguntzeko bide errazak	229.000
Britain's National Rail Network	Britainia Handiko Trenbide Sare Nazionala	217.000
he has been conned out of 10,000 euros	10.000 euro xixkatu dizkiote	207.000
indian philosophical systems	Indiako sistema filosofikoak	202.000
40 degrees in the shade	40 gradu gerizpean	201.000
a seating capacity of 2,000	2.000 eserlekuko edukiera	192.000
apparitions in Fatima	Fatimako agerraldiak	178.000
we did the trip in 4 hours	4 ordutan egin genuen bidaia	167.000
the Academy's headquarters in Cambridge	Akademiak Cambridgen duen egoitza	166.000
50th wedding anniversary	Ezkontzaren 50. Urtemuga	157.000
a 6-credit course	6 kredituko ikastaroa	156.000
a 60-gallon drum	60 galoiko bidoia	140.000
remind me to phone Alan before I go out	Gogorarazi irten aurretik Alani hots egin behar diodala	133.000
20 degrees below zero	20 gradu zero azpitik	111.000
she's partial to Bach	Bach oso gogoko du	109.000
5 plus 2 equals 7	5 gehi 2 berdin 7	97.600
polls close at 8 p.m.	Bozketa gaueko 8:00etan amaitzen da	77.000
the 4th of May	maiatzaren 4a	72.100
today I'm meeting up with Idoia	Gaur Idoiarekin elkartzekoa naiz	68.500
Jim has always loved gardening	Jim-i betidanik gustatu izan zaio lorezaintza	66.800
Mary is the nominee for mayor	Mary da alkategaia	62.000
Christmas Eve is on 24 December	Gabon-eguna abenduaren 24a da	58.200
on the 14th of November	azaroaren 14an	58.100
Liz had an easy labour	Liz-ek erditze erraza izan zuen	57.400
we've been saving up to go to Australia	Australiara joateko aurrezten aritu gara	50.000
he got 2 years	2 urte ezarri dizkiote	49.200
Sally exploded suddenly	Sally sutan jarri zen bat-batean	48.500
Tom and Nancy roomed together at college	Tom eta Nancy elkarrekin bizi ziren unibertsitatean	48.200
a capacity of 10 litres	10 litroko edukiera	48.000
that smell reminds me of France	Usain horrek Frantzia oroitarazten dit	44.000
Round May 10	Maiatzaren 10aren inguruan	43.900
nudes painted by Rubens	Rubens-ek pintaturiko biluziak	42.100
you got it wrong; it's Maria who's coming, not Marina	Ez duzu ongi ulertu; Maria etorriko da, ez Marina	40.800
it only took me 5 minutes	5 minutu baino ez nituen behar izan	38.000
2 times 2 equals 4	2 bider 2 4	30.200
on my way home I bumped into Maite	Etzerako bidean Maite aurkitu dut	29.100
20 minus 8 equals 12	20 ken 8 berdin 12	24.400
7 minus 2 equals 5	7 ken 2 berdin 5	24.400
the Stock Exchange crash of 1929	1929ko burtsaren kraka	22.200
12 divided by 2 equals 6	12 zati 2 berdin 6	20.800
it's minus 5°C	5°C ditugu zero azpitik	18.400
the koala is unique to Australia	Koala Australian baino ez da	18.300

	aurkitzen	
I arrived 5 minutes early	5 minutu lehenago iritsi nintzen	15.900
a poem supposedly written by Beaudelaire	Ustez Beaudelairek idatzitako poema	13.900
2 euros a kilo	2 euro kiloa	9.430
see illustration on page 20	Ikus 20. Orrialdeko ilustrazioa	8.240
to be ready in 5 minutes	5 minututan prest egon	7.060
3 multiplied by 2 equals 6	3 bider 2 sei da	6.980
8 multiplied by 4 equals 32	8 bider 4 berdin 32	6.980
see note 5, page 410	ikus 5. oharra, 410. orrian	6.400
the cube of 2 is 8	2ren kuboa 8 da	5.280
we live half an hour outside London	Londonetik ordu-erdira bizi gara	1.930
500 metres from my house	Nire etxetik 500 metrora	1.920
the river is 300 miles in length	Ibaiak 300 miliako luzera du	1.450
4 squared equals 16	4 ber 2 berdin 16	1.440
we've lived here since 1994	Hemen bizi gara 1994tik	1.340
to average about 9 hours	9 orduko batezbestekoa egin	1.200
the committee was constituted in 1984	Batzordea 1984an sortu zen	1.190
my watch is 5 minutes slow	Nire erlojua 5 minutu atzeratuta dabil	969
summer temperatures range from 20 to 35 degrees	Udan tenperatura 20 eta 35 gradu artean dabil	945
300 odd euros	300 euro eta koska	892
she'll be 30 tomorrow	bihar 30 urte egingo ditu	725
the town is 8 km off	Herria 8 kilometrora dago	693
6 subtracted from 9 is 3	9 ken 6 berdin 3	597
salaries were increased by 5%	Soldatak % 5 gehitu zituzten	594
the bus seats 20	Autobusak 20 eserleku ditu	557
5 kilometres away from here	Hemendik 5 kilometrora	533
Kevin started us all off laughing	Kevinek barrez leherrarazi gintuen	476
do not write more than 200 words	Ez idatzi 200 hitz baino gehiago	442
he has 20 men under him	20 gizon ditu bere ardurapean	368
in Istanbul, East and West fuse together	Istanbulen, Ekialdeak eta Mendebaldeak bat egiten dute	303
I'm 5 years his senior	Hura baino 5 urte zaharragoa naiz	300
he has two big houses in this country, not to mention his villa in France	Bi etxe handi ditu herrialde honetan, Frantzia duen etxaldez gain, jakina	295
set the alarm for 7 o'clock	Jarri alarma 7etarako	293
she is somewhere around 40	40 urte inguru ditu	278
he remained in power for 20 years	20 urtez mantendu zen boterean	227
it must be delivered within 72 hours	72 orduren barruan eman behar dizute eskura	205
the hall can seat 300 people	Aretoareen edukiera 300 eserlekukoa da	198
the plane left Heathrow at 11:30	Hegazkina 11:30etan atera zen Heathrow-tik	143
turn 10 degrees to port!	Lema 10 gradu ababorrera okertu!	138
the bill was carried by 58 votes	Lege-proiektua 58 botorekin eman zen ontzat	133
I'd put her at about 35	35 bat urte jotzen dizkiot	130
he was put on standby for the flight to New York	New York-eko hegaldirako itxaron-zerrendan jarri zuten	116
I find it astonishing that Mike should be	Harritzen nau Mike zurekin hain	112

so rude to you	zakarra izateak	
we've been working for 12 hours straight	12 orduz jarraian aritu gara lanean	105
the prices vary from 1 euro to 10 euros	Prezioak euro 1 eta 10 euro bitartean dabilta	92
at 30 euros per kilogram	30 eurotan kiloa	89
60 euros and something	60 euro eta koska	70
it's the biggest theme park outside the United States	Estatu Batuetatik kanpo dagoen parke tematikorik handiena da	54
the length of each class is 45 minutes	Eskola bakoitzak 45 minutuko iraunaldia du; eskola bakoitzak 45 minutu irauten ditu.	53
Kazakhstan alone contains more than a hundred nationalities	Kazakhstanek berak bakarrik ehun nazio baino gehiago ditu	52
the room measures 2 square metres	Gelak 2 metro karratu ditu	51
the coinage was reformed under Elizabeth I	Moneta-sistema Elizabeth I.aren garaian aldatu zen	40
a 15 amp electrical current	15 ampereko korrante elektrikoa	35
next Thursday is 10 April	Hurrengo ostegunean apirilak 10 ditu	31
2,000 political prisoners had been freed	2.000 preso politiko libre utzi zituzten	31
we have collected 500 euros	500 euro osatu ditugu	18
the return ticket costs 60 euros	Joan-etorriko txartelak 60 euro balio ditu	10
she doesn't like them and nor does Mikel in Madrid and the villages round about	Hark ez ditu maite; Mikelek ere ez Madrilen eta inguruko herrietan	10 9
parental consent to marry is needed if you are under 18	Ezkontzeko beharrezkoa da gurasoen oniritzia 18 urtetik beherakoen kasuan	8
Mikel and the others are going tomorrow	Mikel eta bihar joango dira	8
I judge him about 14	14 bat urte jotzen dizkiot	7
be set to leave by 10 o'clock	Prest egon zaitez 10etan irteteko	7
to exile the prisoners to Australia	Presoak Australiara atzerriratu	7
she settled in Vienna after her father's death	Aitaren heriotzaren ondoren Vienan kokatu zen	7
her name is Izaskun, but they called her Ixaxkun	Izaskun deitzen da baina Ixaxkun deitzen diote	6
we rode to London in a jeep	Jeep batean joan ginen Londonera	6
my house is worth 300,000 euros	Nire etxeak 300.000 euro balio ditu	5
use 12-mm-thick wood	Erabili 12-mm-ko lodieradun zura	5
we can go either to Venezuela or to Argentina	Edo Venezuelara edo Argentinara joan gaitzke	5
there's an age difference of 2 years between us	Gure artean 2 urteko aldea daukagu	4
when a student in London, I lived in a flat	Londoneen ikasle nintzela, pisu batean bizi nintzen	4
Mikel incurred the hatred of his father	Mikelek bere aitaren gorrotoa irabazi zuen	4
Titanic sank after she hit an iceberg	Iceberg baten aurka jo ondoren hondoratu zen Titanic	4
Daniel provoked his friend to break the law	Legea haustera bultzatu zuen Danielek bere laguna	4
I don't want to press the point, but you do owe me \$200	Ez dut gaia berriro atera nahi, baina 200 \$ zor dizkidazu	3

this hall is 50 metres long	Areto hau 50 metro luze da	3
the sale netted 30,000 euros	Salmentarekin 30.000 euro irabazi genituen	3
we'll hold your reservation for 24 hours	Erreserba 24 ordutan mantenduko dizugu	3
this book is the fruit of 15 years' research	Liburu hau 15 urteko ikerketaren emaitza da	3
I'm on now till 8 tomorrow morning	Lana dut bihar goizeko 8ak arte	3
to request a 2,000-euro loan	2.000 euroko kreditua eskatu	3
literature is on the wane in Europe	Europak, literaturan, beheraldia du	3
we'll have to replace Martin with someone else	Martinen ordez besteren bat ipini beharko dugu	3
he worked as a doctor in Getxo, where he died	Mediku izan zen Getxon, non hil baitzen	3
Amy got off with John yesterday	Amy-k John-ekin ligatu zuen atzo	3
in 1640 the Portuguese revolted against the Spanish	1640an portugesak espainolen aurka matxinatu ziren	2
a 30minute consultation costs £50	30 minutuko ontsultak 50 £ balio du	2
to receive £10 000 in compensation	10.000 £-ko konpentsazioa jasotzea	2
he went to America, where he amassed a fortune	Ameriketara joan, eta han aberastu egin zen	2
we commute from Ondarroa to Usurbil every day	Egunero Ondarroatik Usurbilera bidaiatzen dugu lanera	2
observers will be sent by UN	NBEk begiraleak bidaliko ditu	2
they're prospecting for oil in the North Sea	Petrolio bila dabiltza Ipar Itsasoan	2
that railway line was electrified in 1955	Trenbide hori 1955ean elektrifikatu zen	1
the door is 80 cm in width	Atea 80 cm zabal da	1
the spring term finishes on 29th June	Bigarren hiruhilekoa ekainaren 29an amaitzen da	1
she's on 25,000 pounds a year	Urtean 25.000 libera irabazten ditu	1
we usually have lunch at 12 noon	Eguerdiko 12etan bazkaltzen dugu normalean	1
this product retails at 5 euros	Produktu honen jendearentzako salmenta-prezioa 5 eurokoa da	1
you may stay in the US even if your visa has expired	Estatu Batuetan gera zaitezke, bisa iraungita ere	1
Mary recoiled from shaking his hand	Maryk atzera egin zuen hari eskua ematetik	1
it fell to Nerea to welcome the guests	Nereari egokitu zitzaion gonbidatuei ongietorria egitea	1
I'll do the cake for John's birthday	Nik egingo dut Johnen eguneko pastela	1
he wagered 220 euros on this horse	20 euro jokatu zituen zaldi honen alde	0
the winner will receive 200 dollars' worth of gold	Irabazleak 200 dolar jasoko ditu urretan	0
how do you get by on only 100 euros per month?	Nola konpontzen zara hilean 100 euroekin?	0
that'll be 60 euros altogether, please	60 euro dira guztira, mesedez	0
he earns 60 euros a day, twice as much as	60 euro irabazten ditu egunean, bere	0

his father	aitak halako bi	
buy three and make a saving of 55 cents	Erosi hiru eta aurreztu 55 zentimo	0
I've got a book for 50 cents	50 zentimoren truke liburu bat erosi dut	0
they tried out the drug out on 50 patients	Botika 50 gaixorengan probatu zuten	0
she has 50,000 pound' worth of antiques	Haren antigoaleko altzariak 50.000 liberako balioa dute	0
we allowed 20 euros a day for food	Janaritarako eguneko 20 euro gorde genituen	0
the work will cost £10 000 plus	Lana 10.000 liberatik gora kostatuko da	0
she did 5 years of piano	5 urtez ibili zen pianoan (pianoa ikasten=	0
he had 4 classes back to back	4 klase bata bestearen atzetik izan zituen	0
women make up 56% of the student numbers	Ikasle-kopuruaren % 56 emakumeek osatzen dute	0
increases of 12% and above	% 12tik gorako igoera	0
the King of Scots was subordinate to the King of England	Eskoziarren erregea Ingalaterrako erregearen mendeko zen	0
Gibran was born in Lebanon but was naturalized as an American subject	Gibran Libanon jaio zen baina Ameriketara bertakotu zen	0
everyone imposes on Tom's good nature	Tom-en ontasunaz dena baliatzen dira	0
you can always depend on/upon the tram arriving on time in Germany	Alemanian tranbia beti garaiz iritsiko dela fida zaitezke	0
Claire took a jacket that wasn't hers	Clairek berea ez zen jaka hartu zuen	0
Rioja is famous on account of its excellent wines	Errioxa famatua da bere ardoen bikaintasuna dela medio	0
he thumbed all across Europe	Europa osoan barrena bidaiatu zuen auto-stop eginez	0
John wasn't goaded by their insults	John ez zuten zirikatatu haien irainek	0
John's strange behaviour aroused our suspicions of him	Johnen jokaera arraroak susmoa eragin zigun	0
the best tribute we can pay to Leizarraga is reading his work	Leizarragari egin dakioken omenaldirik onena bere obrak irakurtzea da	0
Lizaso junior follows the same attachment to sing out verses, as does his father	Lizaso semeak aitaren bertsozaletasuna jarraitzen du	0
Maria has got many suitors, but she doesn't like any of them	Mariak ezkongai asko dauzka, baina bat bera ere ez du gogoko	0
Tom hasn't left yet; I saw him here this morning	Tom ez da oraindik joan; gaur goizean hemen ikusi dut	0
they've based the new company in China	Enpresa berria Txinan kokatu dute	0
neither John nor Liz enjoy/enjoys vegetables	Ez Johnek eta ez Lizak dituzte gustuko barazkiak	0
he never agreed with the opinions of Darwin and his colleagues	Ez zen sekula Darwin eta kideen iritzien alde agertu	0
he interpreted with great mastery works by Frederic Chopin	Maisutasun handiz jo zituen Frederic Chopin-en lanak	0
tele-evangelists are very popular in the USA	Tele-ebanjelistak oso ezagunak dira AEBn	0
the train calls first at Clapham Junction	Trenak lehendabizi Clapham	0

	Junction-en egiten du geldialdia	
in the summer she travelled all over Europe	Udan Europa osoa korritu zuen	0
a very nice guy asked after you. That would be Tom.	Mutil jator bat etorri da zutaz galdezka. Tom izango zen.	0
the part of Navarre on the banks of the Ebro	Nafarroako erribera	0
Roger stole the money I thought as much	Rogerrek lapurtu du dirua. Pentsatzen nuen, bai	0

6. taula: Agerraldi kopuruagatik baztertutako adibideak, ingelesa-euskara

C eranskina: patroiak sortzeko erabili ditugun adibideak

Gaztelaniazko adibidea (Elhuyar)	Euskarazko adibidea (Elhuyar)	Matxinen itzulpena
¡viva Gales!	gora Gales!	Gales biziduna!
crystal de Bohemia	Bohemiako kristala	Bohemiaren kristala
el 50 aniversario de la boda	ezkontzaren 50. urtemuga	ezkontzako 50 urteurrena
el canal de la Mancha	Mantxako kanala	Mancharen kanala
el canal de Panamá	Panamako kanala	Panamaren kanala
el desierto de Sáhara	Saharako basamortua	Sahararen basamortua
el estrecho de Gibraltar	Gibraltarko itsasarte	Gibraltarren itxia
el jorobado de Notre-Dame	Notre-Dameko konkorra	Notre-dameren gogaitua
el principio de Arquímedes	Arkimedesen printzipioa	Arkimedesen hasiera
el reino de Castilla	Gaztelako erresuma	Gaztelaren erresuma
el Tajo desemboca en el Atlántico	Tajo Atlantikoan itsasoratzen da	Tajo Atlánticon itsasoratzen da
en una ocasión fuimos a París	behin Parisera joan ginen	Aukera batean Parisera joan ginen
finalmente me admitieron en la universidad de Zaragoza	azkenean Zaragozako unibertsitatean onartu ninduten	Azkenean Zaragozaren unibertsitatean onartu ninduten
hora de cierre: 19.00	itxiera-ordua: 19:00	Itxieraren ordua: 19,00
la dama de Amboto	Anbotoko dama	Ambotoren damera
la factoría de Seat en Manresa	Seat-en Manresako faktoria	Seaten Manresarengan faktoria
la hoguera de San Juan	San Joan sua	San Juanen sua
la orquesta filarmónica de Boston	Bostongo Orkestra Filarmonikoa	Bostonen orkestra filarmonikoa
la paga de Navidad	Gabonetako ordainsaria	Navidaden paga
la puerta tiene 80 cm de ancho	atea 80 cm zabal da	Ateak 80cm du zabalez
la sala tiene un aforo de 300 personas	aretoaren edukiera 300 eserlekukoa da	Aretoak 300 pertsonaren edukiera bat du
la selva de Irati	Iratiko basoa	Iratiren basoa
la situación en Iraq	Irakeko egoera	Iraken egoera
la unidad de los pueblos de América	Ameriketako herrien batasuna	Herrien Amerikaren unitatea
las cataratas de Iguazú	Iguazuko ur-jauziak	Iguazúen ur-jauziak
llegué con 5 minutos de anticipación	5 minutu lehenago iritsi nintzen	5min-rekin iritsi nintzen aurrerapenetik
los menores de 14 años	14 urtez azpiko adingabeak	14 urtez txikiak
Luisa está en cinta	Luisa haurdun dago	Luisa zintan dago
paso de Ibañeta	Ibañetako lepoa	Ibañetaren urratsa
reducción de un 10%	% 10eko beherapena	% 10ren murriztapena
un peso de 200 kilos	200 kiloko pisua	200 kilotako 1CUP
una capacidad de 10 litros	10 litroko edukiera	10l gaitasun bat

7. taula: Patroiak sortzeko aukeratutako adibideak, gaztelania-euskara

Ingeleseko adibidea (Elhuyar)	Euskarazko adibidea (Elhuyar)	Matxinen itzulpena
5 December 2003	2003ko abenduaren 5a	5 December 2003
5% off	% 5eko beherapena	5 % urruti
8 hours a day	eguneko 8 ordu	8 hour a dayak
a certain Mikel Larrinaga called	Mikel Larrinaga delako batek deitu du	@mikel Larrinaga ziur batek deitu zuen
a poem by Nathaniel Hawthorne	Nathaniel Hawthornek egindako olerkia	@nathaniel Hawthorneak poema bat
a speed of 60 mph	60 milia orduko abiadura	Mpheko anfetamina bat 60
a Strauss waltz	Strauss-en balsa	@strauss baltse bat
a subsidiary of General Motors	General Motors-en filiala	@general Motorsetako menpeko bat
a war with Italy	Italiaren aurkako gerra	Italiarekin gerra bat
a weight of 200 kg	200 kiloko pisua	Kilogramoko pisu bat 200
above 20 degrees	20 gradutik gora	Graduen gainetik 20
an angle of 45 degrees	45 graduko angelua	Graduetako angelu bat 45
an increase of 6%	% 6ko igoera	6 % igoera bat
are you related to Margaret?	Margareten ahaidea zara?	Zuk kontatzen zaitu @margaret?
around 100 people	100 lagun inguru	Jendeetatik zehar 100
as Shakespeare put it	Shakespearek esan zuen bezala	Hura sartu zen @shakespeare bezala
at 100 mph	100 milia orduko abiaduran	Mphean 100
at 30	30 urterekin	30
back in/into 1973	1973an	Eskugain 1973 in/intoa
between 2 and 3 years	2 eta 3 urte bitartean	Urteetan 2 eta 3
between 3 and 4 kilos	3 eta 4 kilo bitartean	Kiloak 3 and 4
beyond 100 years	100 urtetik gora	Urteez haraindi 100
children of 14 and over	14 urtetik gorako haurrak	14 umeak eta hemen
children under the age of 12	12 urtetik beherako haurrak	Adinaren azpian 12 umeak
Dr Smith, I presume?	zu Smith doktorea izango zara, ezta?	@dr smithak Nik uste naiz,?
finally I was admitted to the University of Zaragoza	azkenean Zaragozako unibertsitatean onartu ninduten	Azkenik Ni onartu ninduen @university Of Zaragozari
give me 2 minutes	emaizkidazu 2 minutu	2 minutuak eman iezazkidazu Niri
glory to the Father, the Son and the Holy Spirit	loria Aitari, eta Semeari, eta Izpiritu Sainduari	Aitara loria, Jainkoaren semea eta Santua Adorea
have you ever been to Morocco?	izan zara egundaino Marokon?	Inoiz zu izan izan Ahuntz-larrura?
he ended up in Madrid	non eta Madrilen bukatu zuen	Hark bukatu zuen up Madrilen
—hello? —is Ainara at home?	—bai, esan? —Ainara etxean da?	-- kaixo? Da etxean @ainara --?
how far is Madrid from Paris?	zer distantzia dago Madril eta Paris artean?	Parisko Madril da nola urruti?
I didn't go and neither did Mary	ni ez nintzen joan eta Mary ere ez	Ni ez nindoan eta neither Maria egin zuen
I live in Australia	Australian bizi naiz	Nik biziduntzen dut Australian
I lived in Egypt for a time	Egipton bizi izan nintzen denboraldi batez	Nik biziduntu nuen Egipton denbora batentzat

I lived in London for a while	Londonen bizi izan nintzen denboraldi batez	Nik biziduntu nuen Londresen aldi batentzat
I love Maule, especially in the summer	Maule gustuko dut, udan gehienbat	Ni @maule maite naiz batez ere in the summer
I love Rome, especially in the spring	Izugarri gustatzen zait Erroma, batez ere udaberrian	Ni Erroma maite naiz batez ere in the spring
I make it 50	50 direla esango nuke	Nik egiten dut hark 50
I met Mikel in Bilbao	Bilbon ezagutu nuen Mikel	Nik Bilboko @mikel elkartu nuen
Iguazu Falls	Iguazuko ur-jauziak	@iguazu fallsa
in 1976	1976an	1976
in Paris, Berlin, or wherever	Parisen, Berlinen edo auskalo non	Edonon Parisen, Berlina, edo
in the north of France	Frantziaren iparraldean	Iparraldean Frantzia
in the west of Spain	Espainiaren mendebaldean	Mendebaldean Espainia
I've never been in Paris	ez naiz behin ere Parisen izan	Ni izan Parisen
I've never been to Australia	ez naiz inoiz Australian egon	Ni izan Australiara
Kevin loves you	Kevinak zu maite zaitu	@kevin zu maite da
Lady of Anboto	Anbotoko dama	@anboto andrea
located in Paris	Parisen kokatua	Parisen locateda
made in China	Txinan egina	Egin Portzelanan
mainland Greece	Grezia kontinentala	Grezia kontinentea
member state of the European Union	Europar Batasuneko estatua	Europar batasuneko kide egoera
Merry Christmas and a Happy New Year!	zorionak eta Urte Berri On!	Gabonak Alaiak eta Zorioneko Urte Berri bat!
Michael is in his room	Michael bere gelan dago	Mikel da haren gelan
Mister Smith	Smith jauna	@mister smitha
more than 60	60tik gora	60 gehiago
Mr Adams	Adams jauna	@mr adamsa
Mrs Adams	Adams andrea	@mrs adamsa
Mrs Jauregiberri	Jauregiberri anderea	@mrs jauregiberria
—no, John is not here —then he must be at the bar	—ez, John ez dago hemen —orduan tabernan egongo da	Da ez hemen hura izan orduan Komuna tabernan ez --, --
Olympic Games venue	Joko Olinpikoen egoitza	Topalekua @olympic gamesak
on the 14th of November	azaroaren 14an	@the 14th Of Novemberrean
on the eve of the French Revolution	Frantziako Iraultzaren bezperan	Bezperan Frantsesa Iraultza
once we went to Paris	behin Parisera joan ginen	Gu gindoazen behin Parisera
quotations from James Joyce	James Joyce-ren aipuak	@james Joyceko aipuak
round 10 o'clock	10ak aldera	10 o'clock biribila
Sahara desert	Saharako basamortua	Basamortua Sahara
she was born in 1950	1950ean jaio zen	Hura eramana zen 1950
Smith and Jones met in the final	Smith eta Jones elkarren aurka aritu ziren finalean	Arotza eta @jones elkartu ziren azkenean
South of France	Frantziako hegoaldea	@south of francea
The Adventures of Tom Sawyer	Tom Sawyer-en abenturak	@tom Sawyerreko Abenturak
the banks of the Seine	Sena ertzean	Arraste-inguraketako sareko

		bankuak
the border between Germany and Denmark	Alemaniarren eta Danimarkaren arteko muga	Ale-maniaren artean ertza eta Danimarka
the children under 14	14 urtez azpiko adingabeak	14 azpian umeak
the Diary of Anne Frank	Ana Frank-en egunkaria	@anne Frankeko Egunkaria
the fall of the Roman Empire	Erromatar Inperioaren erorialdia	@roman Empirea fall ofa
the fall of the Roman Empire	Erromatar Inperioaren gainbehera	@roman Empirea fall ofa
the Moon goes round the Earth	Ilargia Lurraren inguruan biratzen da	Argizagia doa itzuliak Ludia
the Paris metro	Parisko metroa	Metro Paris
the rule of Henry VIII	Henry VIII.aren erregealdia	@HENRY VIIIKO erregela
the San Francisco AIDS Foundation	San Francisco-ko HIES Fundazioa	@san Francisco Aids Foundationa
the Tagus flows into the Atlantic	Tajo Atlantikoan itsasoratzen da	Tajok isurtzen du Atlantikora
the works of Mitxelena	Mitxelenaren lanak	@mitxelena lanak
they left for Morocco	Maroko-rantz joan dira	Haiek utzi zuten @marocco
they say John is rich	John aberatsa omen da	Haiek esaten dute Komunak da aberatsa
to be English by birth	jatorriz Ingalaterrakoa izan	Izan zaitez erditzeak Ingelesa
to return in 4 days	4 egun barru itzuli	Itzuli egunetan 4
upwards of 30 years	30 urtetik gora	Urteetako goranzko 30
we did Cuba in 3 weeks	Kuba 3 astean ikustatu genuen	Guk asteetako Kuba egin genuen 3
we walked 20 kilometres	20 kilometroko bidea egin genuen oinez	Gu 20 kilometroak ibili ginen
what is the distance between the Sun and the Earth?	zein da Eguzkitik Lurrerainoko distantzia?	Zer da eguzkiaren artean distantzia eta Ludia?
what is the population of Maule?	zenbat biztanle ditu Maulek?	Zer da @maule biztanleria?
what? you saw the Julia Roberts?	zer? Julia Roberts berbera ikusi duzula?	What? Zuk @julia Robertsa ikusi zenuen?
when I left France	Frantziatik kanporatu nintzenean	Nik Frantzia utzi nuen
which of you is Mike?	zuetariko nor da Mike?	Zu zein da Mikrofona?
with the exception of Martin	Martin izan ezik	Salbuespenarekin Enara
Zarautz and the surrounding area	Zarautz eta bere inguruak	@zarautz eta inguratu aldea

8. taula: Patroiak sortzeko aukeratutako adibideak, ingelesa-euskara

D eranskina: entitate-detektatzailearen zuzenketak

Hiztegiko adibidea	Euskarazko patroi automatikoa	Euskarazko patroi zuzendua	Akatsa
General Motors-en filiala	General ENTI_0+S_GEN filiala	ENTI_0+S_GEN filiala	1
Ilargia Lurraren inguruan biratzen da	ENTI_0+S_GEN inguruan biratzen da	ENTI_0+S_ABS ENTI_1+S_GEN inguruan biratzen da	1
itxiera ordua: 19:00	itxiera-ordua: ENTI_0:00	itxiera-ordua: ENTI_0	1
James Joyce-ren aipuak	ENTI_0 Joyce-ren aipuak	ENTI_0+S_GEN aipuak	1
Mikel Larrinaga delako batek deitu du	ENTI_0 Larrinaga delako batek deitu du	ENTI_0 delako batek deitu du	1
Nathaniel Hawthornek egindako olerkia	ENTI_0 Hawthornek egindako olerkia	ENTI_0+S_ERG egindako olerkia	1
San Joan sua	ENTI_0 Joan sua	ENTI_0 sua	1
Tajo Atlantikoan itsasoratzen da	ENTI_0+S_INE itsasoratzen da	ENTI_0+S_ABS ENTI_1+S_INE itsasoratzen da	1
Tom Sawyer-en abenturak	ENTI_0 Sawyer-en abenturak	ENTI_0+S_GEN abenturak	1
zein da Eguzkitik Lurrerainoko distantzia?	zein da ENTI_0+S_ABL Lurrerainoko distantzia	zein da ENTI_0+S_ABL ENTI_1+S_ABU_GEL distantzia?	1
zer? Julia Roberts berbera ikusi duzula?	zer? ENTI_0 Roberts berbera ikusi duzula?	zer? ENTI_0+S_ABS ikusi duzula?	1
% 5eko beherapena	% ENTI_0eko beherapena	% ENTI_0+S_GEL beherapena	2
% 6ko igoera	% ENTI_0ko igoera	% ENTI_0+S_GEL igoera	2
10ak aldera	ENTI_0ak aldera	ENTI_0+P_ABS aldera	2
1950ean jaio zen	ENTI_0ean jaio zen	ENTI_0+S_INE jaio zen	2
1973an	ENTI_0an	ENTI_0+S_INE	2
2003ko abenduaren 5a	ENTI_1ko abenduaren ENTI_0a	ENTI_1+S_GEL abenduaren ENTI_0+S_ABS	2
60tik gora	ENTI_0tik gora	ENTI_0+S_ABL gora	2
jatorriz Ingalaterrakoa izan	jatorriz ENTI_0+S_ABS izan	jatorriz ENTI_0+S_GEL_ABS izan	2
Jauregiberri anderea	ENTI_0+S_DAT anderea	ENTI_0 anderea	2

9. taula: Euskarazko patroi automatikoen zuzenketak

Hiztegiko adibidea	Gaztelaniazko patroi automatikoa	Gaztelaniazko patroi zuzendua	Akatsa
el Tajo desemboca en el Atlántico	el ENTI_0 desemboca en el Atlántico	el ENTI_0 desemboca en el ENTI_1	1

10. taula: Gaztelaniazko patroi automatikoen zuzenketak

Hiztegiko adibidea	Ingeleseko patroi automatikoa	Ingeleseko patroi zuzendua	Akatsa
Iguazu Falls	ENTI_0	ENTI_0 falls	1
Mrs Jauregiberri	ENTI_0	Mrs ENTI_0	1
the Tagus flows into the Atlantic	the ENTI_0 flows into the Atlantic	the ENTI_0 flows into the ENTI_1	1
what is the distance between the Sun and the Earth?	what is the distance between the Sun and the ENTI_0?	what is the distance between the ENTI_0 and the ENTI_1?	1

11. Taula: Ingeleseko patroi automatikoen zuzenketak