



Euskarazko egitura sintaktikoen azterketa testuen  
sinplifikazio automatikorako: aposizioak, erlatibozko  
perpausak eta denborazko perpausak

Itziar Gonzalez Dios  
Tutorea: M<sup>a</sup> Jesús Aranzabe

hap

Hizkuntzaren Azterketa eta Prozesamendua Masterreko titulua  
lortzeko bukaerako proiektua

2011ko iraila

**Sailak:** Lengoaia eta Sistema Informatikoak, Konputagailuen Arkitektura eta Teknologia,  
Konputazio Zientziak eta Adimen Artifiziala, Euskal Filologia, Elektronika eta Telekomunikazioak.

ABSTRACT

In this work we will expose the current situation of text simplification in NLP. Besides, due to the difficulty that long sentences cause at machine translation, we will analyse a corpus of long sentences in Basque in order to make syntactic simplification proposals. Although those simplification suggestions are limited to some specific sentences (appositions, relative and temporal sentences), this linguistic approach opens a research line for the Basque language in NLP text simplification.

key words: text simplification, Basque, linguistic approach, syntactic simplification

## **I Edukien aurkibidea**

1 Sarrera eta motibazioa .....	4
2 Aurrekariak.....	8
2.1 Sistemak, prototipoak eta proposamenak.....	8
2.2 Testu-sinplifikazioa egiten duten produktuak.....	10
2.3 Sinplifikazioa beste aplikazioetan: laburpen automatikoak eta irakurketa-mailaren ebaluazioa.....	12
3 Euskarazko Corpusaren azterketa .....	13
3.1 Sintaxi mailako fenomenoak .....	14
3.1.1 Aposizioak eta azalpenezko egiturak.....	15
3.1.2 Erlatibozko perpausak.....	15
3.1.3 Perpaus markatuak.....	16
3.1.4 Egitura sintagmatiko luzeak.....	16
3.2 Estilo mailako fenomenoak.....	17
3.2.1 Puntuazio-markak gaizki.....	18
3.2.2 Polisindetonak.....	18
3.2.3 Estilo nahasketa.....	19
3.2.4 Hurrenkera logikoa eta hitz-hurrenkera kanonikoa.....	19
4 Sinplifikazio-proposamenak .....	21
4.1 Aposizioak eta azalpenezko egiturak .....	22
4.2 Erlatibozko perpausak .....	26
4.3 Denborazko perpausak.....	29
4.4 Sinplifikazio erregelen hurrenkera.....	32
4.5 Sinplifikatzeko irizpideak, laburbilduz.....	33
5 Baliabideak eta tresnak.....	34
5.1 Muga Gramatika .....	34
5.2 Idazlanen Autoebaluaziorako Sistema.....	36
6 Ondorioak eta etorkizuneko lana.....	37
Bibliografia.....	39
Gramatikak, estilo-liburuak eta hizkuntza errazteko gomendioak.....	42
Ixa taldeko tresnak eta baliabideak.....	42

## **II Taulen aurkibidea**

1. taula: Itzulpen automatikoaren errore-tasaren ebaluazioa esaldien luzeraren arabera.....6
2. taula: PorSimples proiektuan sinplifikatzeko erabili duten eskema.....21

## **III Irudien aurkibidea**

1. irudia: SIMPLIFICA: esaldi baten sinplifikazio sintaktikorako aukerak.....11
2. irudia: MuGak gramatika estatistikoaren adibidea.....35
3. irudia: Ikasle batek egindako idazlanaren ebaluazioa IAS sistemaren bitartez.....37

## **1 Sarrera eta motibazioa**

Ikerketa-lan honetan euskararen tratamendu automatikorako baliagarria izango den idatzizko testuen sinplifikazioa aztertuko da. Lan hau Ixa taldean<sup>1</sup> egiten da, hizkuntzaren tratamendu automatikoan lan egiten duen Euskal Herriko Unibertsitateko ikerkuntza-taldea. Euskararen gaineko ikerketa aplikatua da taldearen xede nagusia, baina beste hizkuntzen inguruko ikerketa eta produktuen garapena ere sartzen dira gure eginkizunen artean. Lan horretan, diziplinarteko taldea osatzen dugu, batez ere Hizkuntzalaritza eta Informatika uztartuz.

Testuak sinplifikatzea edo erraztea, ingelesez *Text Simplification*, testu zail bat errazago edo ulergarriago bihurtzea da, jatorrizko testuaren esanahia mantenduz. Errazteko prozesu honek hiru atal ditu: lehendabiziko atala, egitura sintaktiko konplexu bat egitura errazagoa bihurtzea da. Bigarren atala, berriz, lexiko-semantikaren atala da, hau da, maiztasun gutxiko hitzak edo “hitz zailak” ezagunagoengatik edo maiztasun handiagokoengatik ordezkatzeko da. Eta azkenik, hirugarren atala, sinplifikaturiko testu horien egokitzapena da, alegia, behin prozesu hori guztia gauzaturik sortzen diren testu erraztuak egokiak izatea lortu behar da. Sinplifikazioa hiru atal hauek osatzen duten arren, sintaxia erraztea da helburu nagusia.

Sinplifikazioaren lana egin ahal izateko oinarrizko baliabidea testuak dira. Testu zailak eta testu errazak identifikatzen ditugu hiztunok, eta Aristegietak (2009) dioen moduan, testu bat sintaktikoki ulertteraza dela esaten dugu baldin eta lehendabiziko aldiz irakurtzean ulertzen badugu. Aldiz, testu bat ulertzeko behin baino gehiagotan irakurri behar badugu testua zaila iruditzen zaigu, batik batik egitura konplexuak dituelako edo hiztegia ezezaguna delako. Lan honetan, gainera, testu zailak lehen irakurraldian ulertzen ez diren testuak izateaz gain, esaldi konplexuak dituzten testuak dira. Esaldi konplexuak mendeko perpausez osaturiko esaldiak dira, horiek irakurketa-maila baxua duen jendearentzat zailak baitira.

Sinplifikazioan hasierako lan gehienak ingeleserako egin dituzte, hein batean beste hizkuntzetarako bideak irekiz (Chandrasekar, 1996). Hizkuntza honetarako, adibidez, 2.1 atalean azalduko diren sinplifikatzaile sistemak eta prototipoak proposatu dira. Era berean, erakunde ofizialetatik eta komunikabideetatik testu errazak edo ez hain konplexuak sortzeko irizpideak ere

---

<sup>1</sup> <http://ixa.si.ehu.es/Ixa>

## ***HAP Masterra 2010/2011 ikasturtea***

jaso dira (*Plain language* programak, estilo-liburuak...). Beste hizkuntza batzuetarako, adibidez, Brasilgo portugesezko, analfabetismoaren tasa altua dela-eta, portugesezko testuak errazteko PorSimples (2008) proiektu sendoa lan egiten ari dira. Bertan hainbat produktu sortu dituzte 2.2 puntuan ikusiko dugun moduan; produktu horiek sortzeko azterketa linguistiko sendoak burutu dituzte eta esku artean dugun lan honen bideak ireki dituzte. Egun PorSimples proiektuan oinarrituz, arabiarerako ere sinplifikatzailea sortzen hasi dira.

Bestalde, herrialde batzuetan, administrazioaren aldetik batez ere hizkera errazagoa bultzatzeko kanpainak egiten dituzte. Honen ildotik sorturikoa da Plain Language (2011), AEBetako gobernu federaletik emandako irizpideak (*Federal Plain Language Guidelines*) biltzen dituenak. Erresuma Batuan lehentasunen artean sartu dute *Technology Foresight Programmen* eta *EPSRC Speech and Language Programmen* bitartez interneteko testuak ulerterrazak izatea. Hego Amerikako herrialde batzuetan ere aurrekontuak eta informazio administratiboa hizkuntza errazago batean interneten jartzen hasi dira, jende gehiagorengana irits daitezzen. Beraz, testu errazak beharrezkoak diren baliabideak direla ikus dezakegu.

Irizpide linguistikoei dagokienez, euskararako ere baditugu estilo-liburuetan (Berriakoa, Egunkariakoa, IVAPekoa) eta idazkera-liburuetan (Zubimendi, 1993) hizkuntza errazteko eta ulerterrazagoa izateko aholkuak, baita eskoletatik eta irakaskuntza mundutik lor ditzakegun gomendioak ere, adibidez Durangoko berritzeguneko (2003) irakasleek egindako aholku-txostena. Estilo- eta idazketa-liburu hauetan guztietan, oro har, esaldi sinple, argi eta laburrak idaztea gomendatzen da.

Testuak erraztea edo sinplifikatzea gaur egun irakaskuntzan eta administrazioan ez ezik Hizkuntzaren Prozesamenduan (HP) ere puri-purian dagoen gaia da. Izan ere, testu errazek abantaila ugari eskaintzen dizkiete, bai gizakiei, baita makinei ere. Gizakien kasuan, hainbat azpitaldera bidera daitezke sinplifikazioak:

- 1) Atzerriko hizkuntzak ikasten ari direnak: ikasketa-prozesu horretan, egiturak pixkanaka ikasten direnez, ikasketa-prozesua aurreratua izan arte, ez dituztelako egitura guztiak ezagutzen.
- 2) Analfabeto edo gutxi alfabetatuak: ikasketa/irakurketa-maila baxua dutenez, testu landuak

## HAP Masterra 2010/2011 ikasturtea

zailak iruditzen zaizkie.

- 3) Haurrak: oraindik hizkuntza osoa barneratuta ez dutenez, kontzeptu batzuk ez dituzte ulertzen; kasu honetan, sinplifikazioa batez ere lexiko mailan izango litzateke.
- 4) Afasikoak: hizkuntza ulertzeko eta hitz egiteko gaitasuna galdu dutenez, egitura jakin batzuk ulertzeko gaitasuna galdu dute.
- 5) Gorrak: munduaren kontzeptualizazio ezberdina dutenez, hizkuntza arrunta ulertzea zaila egiten zaie.
- 6) Gaixotasun kognitiboak dituztenak: *Alzheimerraren* eta antzeko beste gaixotasunen ondorioz, ulertzeko gaitasuna galtzen dutelako.

Makinei dagokienez, HPko zenbait sistematan (analizatzaile sintaktikoak (*parser*), itzulpen automatikoa, galde-erantzun sistemak, laburpen automatikoak egiteko sistemak...) aurreprozesu moduan erabil daiteke sinplifikazioa edo beste aparatuek (sakelako telefonoak, PDAk, tabletak...) dituzten pantaila txikietan informazioa modu erosoan agertzeko ere baliagarria izan daiteke.

Honela, Ixa taldean garatu eta garatzen ari garen baliabideak kontuan hartuz, batez ere 5 puntuan azalduko ditugun Muga Gramatika (MuGa) eta Idazlanen Autoebaluaziorako Sistema (IAS), euskararako sinplifikazio-proposamen sintaktikoen lehen hurbilpena egiteko egitura sintaktiko konplexuen azterketa linguistikoa eta sinplifikatzeko proposamenak egingo ditugu lan honetan. Horretarako bi corpusen (EPEC eta Consumer) laginetatik aukeraturiko esaldi luzeak hartuko ditugu, esaldi luzeak euskararen tratamendu automatikoan zailtasunak ematen dituztelako.

Labakaren (2010) tesi-lanean egin ziren esperimenduek aditzera ematen dutenez, esaldi luzeek arazoak sortzen dituzte itzulpen automatikoan; 1. taulan adierazten den bezala, esaldiaren luzera handitzen den heinean, errore-tasa (HTER)<sup>2</sup> ere handitu egiten da. Esperimendu horiek gaztelaniatik euskararako itzulpenekin egin dira, *Matxin* (Mayor et al., 2011) itzultzailea erabiliz. Ondorioz, esaldi horiek sinplifikatzea lortuko bagenu, prozesu hori erraztuko litzateke.

1 taula: itzulpen automatikoaren errore-tasaren ebaluazioa esaldien luzeraren arabera

Luzera	0 < Luzera <= 5	0 < Luzera <= 10	10 < Luzera <= 20	20 < Luzera
Esaldi kopurua laginean	5	41	100	59
Errore-tasa (HTER)	17,65	28,57	32,54	49,16

2 HTER (Human-targeted Translation Error Rate) metrika mota bat da eta gutxiengo *edit* (aldaketa) kopurua kalkulatzeko du gaitasuna helburu *post-edit* erreferentzia lortzeko. (Labaka, 2010)

Lehenengo taulako adibideak hobeto ulertu ahal izateko (1) eta (2) adibideak izango ditugu aztergai. (1) adibidean euskaratik gaztelaniarako esaldi luze baten itzulpena dugu, eta (2) adibidean, berriz, esaldi labur batena. Itzulpen horiek Alacanteko Unibertsitateak egindako erderatu.eu<sup>3</sup> itzultzailearen bitartez egin dira. Itzultzaile honek Apertium<sup>4</sup> (kode irekiko itzulpen automatikoko plataforma) (Ramirez-Sanchez et al., 2006) teknologia erabiltzen du eta oraindik garatze-fasean badago ere, itzulpenean esanahi orokorrari buruzko ideia bat egiteko baliagarria da. *Matxin* gaztelania-euskara itzultzaile prototipotik hiztegia hartzen du, eta gure esperimentuan baliagarria da informazio linguistikoa integratua duelako. Kontuan izan (1) eta (2) adibideetan agertzen diren izartxo horiek itzultzaile automatikoak hitz horiek ez dituela ezagutzen esan nahi duela.

	<b>Jatorrizko esaldia</b>	<b>Itzulpena</b>
(1)	Emakumezko hezitzaileen artean beti izan da lehia historiko bat neskak eta mutilak modu berean tratatzeko, eta curriculum bikoitza aintzakotzat ez hartzeko, baina jarrera hori dela medio ukatu egin dugu nesken artean alderik dagoenik ere, eta horrela funtzionatu du neska eta emakume askoren kontra, ahalmen gutxikoak zirela edo maila egokia ematen ez zutela esaten zelarik.	Entre de los educadores de mujer siempre ha sido una competencia histórica chica y el chico para tratar en mismo modo, y *curriculum doble *aintzakotzat no para coger, pero esa actitud es , *medio hemos negado entre de las chicas lado está que también, y de ese modo ha funcionado de chica y muchas mujeres contra, el de pocas facultades eran , o el nivel apropiado no daban , decía @larik.
(2)	Garzellik oinordekoa izan nahi du Mercatonekoak Pratonevoson irabazi eta 25 segundora du elastiko arrosa.	*Garzellik El heredero ser quiere *Mercatonekoak *Pratonevoson ganar y a los 25 segundos tiene *elastiko la rosa.

(1) adibideko esaldiaren gaztelaniazko itzulpena irakurtzerakoan zaila da esaldia ulertu ahal izatea baldin eta euskarazko jatorrizkoa ezagutzen ez bada. (2) adibidearen itzulpena irakurtzerakoan, ordea, esanahia errazago antzeman daiteke, nahiz eta akatsak dauden .

Beraz, sarrera honen ondoren, gure proposamenaren aurrekariak izan diren beste prototipoak, sistemak eta produktuak ikusiko ditugu 2 atalean. 3 atalean, euskarazko corpusaren berri eman eta bertan aurkitutako eta ikerketa-lan honetan erabiliko ditugun fenomenoak azalduko ditugu. 4. atalean, berriz, lan honetan sinplifikatuko ditugun perpaus motak eta proposamenak azalduko ditugu. 5. atalean lagungarri izan daitezkeen tresnen azalpenak eman eta erabilgarritasuna aztertuko dugu. Eta, azkenik amaitzeko, 6 atalean, ondorioak eta etorkizuneko lana zehaztuko dugu.

3 <http://www.erdatatu.eu/index.eu.html>

4 <http://www.apertium.org/?lang=es>



## **2 Aurrekariak**

Testu-simplifikazioan orain arte egin diren eta azken urte hauetan garatu diren sistemen, prototipoen eta proposamenen azalpen laburrak egingo ditugu 2.1 azpiatalean. Era berean, tresna horiek egiteko erabili diren metodoak ere adieraziko ditugu, beti ere egileek zehazten badituzte. 2.2 azpiatalean, mota honetako produktu batzuk azalduko ditugu, baita erabiltzaileen araberrako sailkapena ere. Azkenik, 2.3 azpiatalean, sinplifikazioarekin zerikusia eta lotura duten beste bi gai adieraziko ditugu, bata laburpen automatikoak (*autosummarization*) eta bestea, irakurketa-mailaren ebaluazioa (*readability assessment*), horietan ere lagungarriak izan daitezkeen ideiak aurki ditzakegulako.

### **2.1 Sistemak, prototipoak eta proposamenak**

Testuak errazteko motibazioak eta metodo batzuen proposamenak Chandrasekar et al.ek (1996) azaltzen dituzte. Chandrasekar et al. en artikulua horretan, testu sinplifikazioaren helburuak eta abantailak aipatu ondoren, lan honen sarreran beste abantaila batzuekin batera azaldu ditugunak, sintaxia errazteko makinei zuzenduriko bi proposamen egiten dituzte: lehendabizikoa, *Finite State Grammar* (FSG) oinarrituz eta bigarrena, dependentzia-zuhaitzak erabiliz. Lehendabizikoan proposamenean, sinplifikazio-erregelak egiten dituzte *chunketan* oinarrituta, eta bigarrenean ere bai, baina dependentzia-zuhaitzen egiturak identifikatuta. Horiez gain, sistema eratzerakoan aurkitu zituzten arazoak eta eginbeharrak azaltzen dituzte: esaldiak non moztu, sinplifikatu beharreko egituren hurrenkera...

Sussexeko Unibertsitatean PSET: Practical Simplification of English Text<sup>5</sup> (1997-2000) izeneko proiektuan afasikoei sintaxia errazteko SYSTAR sistema garatu dute. Batez ere egunkarietako sintaxia erraztea da helburua, afasikoen interneten estilo sofistikatu eta hitz zailekin idatzirik dagoen informazioa erraz atzi dezaten. Proiektu horretako prototipoan, Carroll et al.ek (1999) testuaren analisia egin ondoren, anaforaren desanbiguazioa egitea proposatzen dute sintaxia erraztu aurretik. Behin sintaxia erraztuta, lexikoa ordezkatzeko eta azkenik, morfologia zuzenketa egiteko, analisisian erabilitako prozesua alderantziz aplikatzen dute.

Maxek (2000), berriz, Liben-Nowellek (2000) egindako erregela konpilatzailea (*rule*

5 <http://www.informatics.sussex.ac.uk/research/groups/nlp/projects/pset.php>

## ***HAP Masterra 2010/2011 ikasturtea***

*compiler*) eta sinplifikazio motorra (*simplification engine*) erabiliz, afasikoei zuzenduriko sintaxia errazteko sistema bat aurkezten digu eta sistema hori hizkuntzalari batek idatzitako arauetan oinarrituta dago. Ondorengo lanetan, Maxek (2005) ulertzeko arazoak dituzten pertsonentzat berridazketan oinarritutako sistema baten sarrera egiten du.

Sistematik arrakastatsuen, ordea, Siddharthanen arkitektura (2002) da, non eta lehendabiziko aldiz kohesioa mantentzeko garrantzia ematen zaion. Hiru mailako sistema proposatzen du sintaxia sinplifikatzeko:

- i) analisisia (*analysis*): kategoria jarri, egiturak identifikatu (*pattern matching*), zatiak pasa (*chunking*) eta analisi sintaktikoa egin (*parsing*).
- ii) eraldaketa (*transformation*): eskuz idatzitako erregelak eta erregela horiek aplikatzeko hurrenkera.
- iii) birsortzea (*regeneration*): estiloa, esaldien hurrenkera, anaforak... aztertu, hots, kohesioa mantendu eta egokitu.

Esaldia ezin bada sintaktikoki gehiago sinplifikatu, sinplifikazio lexikalaren unea dela dio, hau da, parafraseatu eta sinonimoak ordezkatu egiten dira.

PorSimples proiektuan, alegia, Brasilgo alfabetatze-tasa baxuaren ondorioz testuak ulerterrazagoak egiteko helburua duen proiektuan garatu duten sistemak (Candido et al., 2009) berrikuntza bat du. Testuak sintaktikoki erraztu aurretik, (hori Siddharthanen prototipoan oinarrituz egite dute) testuaren zailtasun-maila aztertzen duen atal batetik pasatzen dituzte testuak (*readability assessment stage*), hots, testuaren zailtasun-maila neurtzen dute testua sinplifikatu aurretik. Hortaz, behin zailtasun-maila hori neurtu dutela, testua nola egokitu aukera dezakete. Izan ere, sistema honek sinplifikazio naturala (fenomeno batzuk bakarrik sinplifikatu) edo sinplifikazio bortitza (ahal den guztia sinplifikatu) egiteko aukera ematen du, testuak izango duen erabiltzailearen alfabetatze-mailaren arabera. Sinplifikazio-maila hauek 4 atalean azalduko ditugu.

PorSimplesen sisteman oinarritzen den Al-Baseeten, arabiarerako proposamenean (Afnanek et al., 2011), laugarren maila bat gehitzen diote aurreko sistemari: diakritizazioa. Pauso hau beharrezkoa ikusten dute egileek, arabiar hizkuntzaren egitura ulertzen laguntzen dutelako diakritikoei. Honako hau da lau mailako prototipo horren osaera: testu konplexutasunaren ebaluazioa, sinplifikazio lexikala, sinplifikazio sintaktikoa eta diakritizazioa.

Sistema hauen guztien muina sinplifikatzeko erabiltzen dituzten arauetan dago. Arau horiek hizkuntzalari batek eskuz idatzitakoak izan daitezke edo Chandrasekar et al.ek (1997) proposatzen duten bezala, ikasketa automatikoaren bitartez erauzitakoak. Hau da, bi corpus izanik, bata arrunta eta bestea sinplifikatua eta aurrekoaren baliokidea, bertatik arauak erauztea proposatzen dute.

Petersenek (2007) adierazten duen bezala, corpusek ez dute baliokideak izan behar; egitura sinple eta arruntak lortzeko, nahikoa da alde batetik, esaldi sinpleak eta bestetik, esaldi konplexuago edo arruntak dituzten corpusak aztertzearekin. Honela, corpus horietatik egitura sinpleen eta egitura arrunten edo konplexuen zerrendaketa egitea lortuko da. Izan ere, Chandrasekar et al.ek (1996) Afnanek et al.ek (2011) dioten bezala, identifikatu behar direnak egiturak dira, egitura konplexuak eta egitura errazak zeintzuk diren jakin ahal izateko.

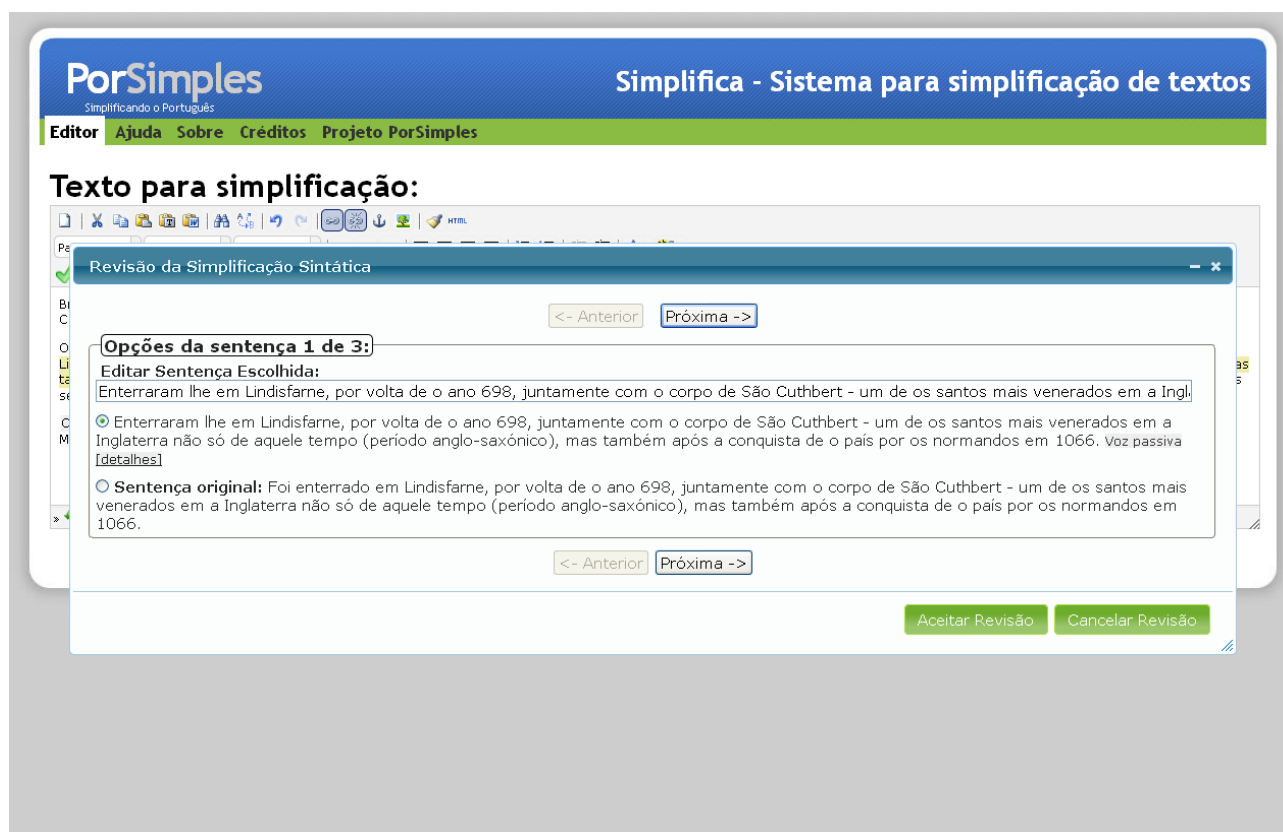
Aurreko paragrafoan esan bezala, testuak sinplifikatu ahal izateko, arauak idazteko egiturak ezagutu behar dira eta horiek automatikoki lortuak edo hizkuntzalari batek idatzirikoak izan daitezke. Halaber, automatikoki, ikasketa automatikoaren bitartez eskuratzen dituztela Chandrasekarek et al.ek (1996) adierazi dugu; eskuz egindako arauak eta egiturak zeintzuk diren jakin nahi badugu, gure aurrekari izan den portugoserako PorSimples proiektuan egindako azterketa bikaina dugu eskura (Specia et al., 2008). Bertan esaldi motaren arabera, esaldi bakoitzarekin zer egin adierazten dute adibideekin batera.

## ***2.2 Testu-simplifikazioa egiten duten produktuak***

Simplifikatzeko sistemez gain, jadanik hau integraturik duten testu-egokitzaileak edo testu-erraztatzaileak proposatzen dituzte Bursteinek (2009) eta PorSimples proiektukoek (2008). Testu-egokitzaileak idazlea idazten doan heinean esaldiak sinplifikatzeko laguntzaileak dira. Bertan aukeratutako esaldiaren sinplifikazio posible batzuk azaltzen zaizkio idazleari, sinplifikatu nahi duen mailaren arabera antolatuak, eta idazleak esaldi horiek aukeratu edo bazter ditzake. Mota honetako tresnak PorSimples proiektuan garaturiko SIMPLIFICA (1. irudia) testu-editorea (Candido et al., 2009; Scarton et al., 2010) eta Bursteinek (2009) azalduko *Text Adaptor*, webean oinarritutako sistema, dira. Hala ere, azken horrek sinplifikatu baino horretarako iradokizunak

## HAP Masterra 2010/2011 ikasturtea

ematen ditu. Era berean, PorSimples proiektukoek web nabigatzailean integratzen den *FACILITA Educacional* (Watanabe et al., 2009) izeneko testu-erraztzailea sortu dute. Honek testuak errazten ditu irakurleak Interneten nabigatzen duen bitartean. Laburpen automatikoak integratuak dituen *FACILITA* tresna ere eskuragarri jarri dute proba eta ebaluazio fasean.



1 irudia: SIMPLIFICA<sup>6</sup>: esaldi baten sinplifikazio sintaktikorako aukerak

Siddharthanek (2002) izan ezik, gainontzeko egileek sistemak egiterakoan erabiltzaile talde bat izan dute helburu. Hala nola, Carrollek (1999) eta Maxek (2005) gaixotasun disfunzionalen ondorioz (adibidez afasia), irakurtzerakoan ulertzeko arazoak dituztenei zuzendu diete beraien proposamena. Inuik (2003), berriz, gorrentzat egiten du proposamena, hauek irakurtzerakoan dituzten arazoei erreparatuz. Petersenek (2007) eta Bursteinek (2009), bestalde, atzerriko hizkuntza bat, kasu horietan ingelesa, ikasten ari direnentzat egin dute lana. PorSimples proiektukoek (2008-), aldiz, Brasilen analfabetismoarekin duten arazoa aintzat hartuz, alfabetatu gabe edo gutxi alfabetaturiko erabiltzaileentzat egiten dituzte tresnak, hauen gizarteratzean laguntzeko asmoz. Arabiarrentzat ere erabiltzaile talde hori da helburua. Azkenik, De Belder-ek eta

6 <http://www.nilc.icmc.usp.br/porsimples/simplifica/>

Moensek (2010) haurrentzat egiten dute hurbilpen bat.

Sarreran aipatu bezala, sistema hauek ere makinei begira eraiki daitezke. Chandrasekar et al.ek (1996) proposaturiko metodoa HPko zenbait sistemarentzat aurreprozesatzaile bat da, baina Klevanov et al.ek (2004) adierazten duten bezala, sinplifikatzaileak informazioa bilatzeko aplikazioetarako ere baliagarriak izan daitezke. Siddharthan et al.ek (2004) ere sinplifikazio sintaktikoa dokumentu anitzeko laburpen automatikoan (*multi-document summarization*) edukien aukeratzea hobetzeko erabiltzen dute.

### ***2.3 Sinplifikazioa beste aplikazioetan: laburpen automatikoak eta irakurketa-mailaren ebaluazioa***

Testu sinplifikatzaileak beste mota batzuetako aplikazioetan ere baliagarriak izan daitezke; izan ere, laburpen automatikoak egiten dituzten programa batzuek sinplifikatzaileak edo hauen antzekoak integraturik dituzte laburpenak errazago egiteko, adibidez, Jing-en (2000) *sentence reduction* sistema edo goian aipaturiko Siddharthanen et al.en (2004) proposamena. Horietan laburpenei zuzenduriko sintaxi sinplifikatzaile jakin baten proposamena egiten da eta hori erabiliz laburpenak errazago egiten direla ondorioztatzen da.

Irakurketa-mailaren ebaluazioari dagokionez, PorSimples proiektuan garaturiko sistemak (Candido et al., 2009) testuen zailtasun-maila ebaluatzen duen atal batetik pasatzen dira. Brasilgo Alfabetatze Funtzionalen Adierazle Nazionalak lau alfabetatze-maila bereizten ditu: alfabetatugabeak (*illiterate*), gutxi alfabetatuak (*rudimentary*), oinarrikoak (*basic*) eta aurreratuak (*advanced*) (Gasperin et al., 2009). Ondorioz, maila hauek kontuan hartuz doitzen dituzte PorSimplesekoek sinplifikazio-mailak (4. puntuan azalduko diren maila bortitza eta naturala). Beraz, irakurketa-mailaren ebaluazioak (*readability assessment*) ere gure gaiarekin lotura badauka.

Hain zuzen ere, Ixa taldean, irakurketa-mailaren ebaluazioarekin erlazionaturik lan batzuk egin dira. Hala, ikasleen idazlanak ebaluatzeko Idazlanen Autoebaluaziorako Sistema (IAS) (Castro-Castro et al., 2008) sortu da. 5. atalean azalduko dugun tresna horrek jartotasuna, zuzentasuna, konplexutasun gramatikala eta konplexutasun lexikala aztertzen ditu. Tresna honen gaztelania ebaluatzeko bertsioa Kuban, *Centro de Lingüística Aplicada* (CLA) hobetzen ari dira.

Bigarren atalari amaiera emateko, esan behar da, euskarari dagokionez, oraindik ez dela testuen sinplifikazioa automatikoki egiten duen sistemarik proposatu. Hala ere, informatikaren aldetik lan hau jorratzen baldin bada, atal honetan aipatu diren Siddharthanena (2002) eta PorSimples (2008) proiektuko lanak berrikustea eta aztertzea proposatzen dugu, Siddharthanena beste prototipoak osatzerakoan eragin handiena izan duena baita (testu sinplifikazioaren literaturan behin eta berriro aipatzen den lana da) eta PorSimplesena oparoena eta eskuragarriena delako. Azterketa linguistikoei dagokienez, PorSimpleskoena (Specia et al., 2008) da linguistikoki osatuena, egituren azalpenak eta operazio guztiak adibideekin hornitzen dituztelako. Siddharthanek (2002), aldiz, egituren azalpenak egiten ditu, baina ez ditu operazioak argitzen.

### **3 Euskarazko Corpusaren azterketa**

Esaldien sinplifikazioa gauzatzeko bi corpusen laginak erabili ditugu: alde batetik, *Euskararen Prozesamendurako Erreferentzia Corpuseko* (EPEC) (Aduriz et al., 2003) esaldirik luzeenak eta bestetik, *Consumer Corpusetik* (Alcázar, 2005) esaldietako batzuk hartu ditugu ausaz.

EPEC corpora euskara batuan idatzitako 300.000 hitzez osaturiko testu-bilduma da. Testu horiek *XX. mendeko euskararen corpus estatistikotik*<sup>7</sup> eta *Euskaldunon Egunkariatik*<sup>8</sup> hartu dira eta maila desberdineko informazio linguistikoarekin (morfologia eta sintaxia, egina; semantika eta pragmatika, prozesuan) etiketatu dira eskuz eta automatikoki (Aldezabal et al., 2009). Corpus honek euskararen prozesamendurako erreferentzia corpora izateko asmoa du, alegia, euskararen prozesamendurako hainbat tresna garatzeko eta ebaluatzeko erabiliko da. Corpus honetan hartzen diren esaldiaren mugak honako hauek dira (Aranzabe, 2008):

- puntutik edo esaldi hasieratik puntura bitartekoak
- puntutik edo esaldi hasieratik bi puntura bitartekoak
- puntutik edo esaldi hasieratik puntu eta koma bitartekoak
- puntutik edo esaldi hasieratik harridura-marka bitartekoak
- puntutik edo esaldi hasieratik galdera-marka bitartekoak

Corpus honetan, beraz, puntuazio-markak esaldiak banatzeko erabiltzen dira. Salbuespena komak

<sup>7</sup> <http://www.euskaracorpora.net/>

<sup>8</sup> <http://www.egunero.info> (2002an atzitu)

## ***HAP Masterra 2010/2011 ikasturtea***

dira esaldiak edo sintagmak elkartzeko baliatzen direlako.

Ikerketa-lan honetan baliatuko dugun EPECeko laginean 595 esaldi dauzkagu; bertako esaldi markatuak, hau da, menderagailu bat dutenak, eta sinplifikatu beharko direnak (Arrieta, 2010) 488 dira. Sinplifikazioari begira egin dugun kontaketa honetan, ezin da ahaztu esaldi batean perpaus mota bat baino gehiago ager daitekeela; beraz, lagin honetan mendeko perpaus gehiago dago. Esaldirik luzeena 138 hitzekoa da, eta laburrena, berriz, 14 hitzekoa.

Consumer corpora<sup>9</sup>, berriz, *Consumer* aldizkariaren 1998 eta 2009 bitarteko 131 alek osatzen dute eta horietan 2.590 artikulua biltzen dira. Corpus honen bereizgarritasuna da artikulua horiek 4 hizkuntzetan (gaztelaniaz, euskaraz, galegoz eta katalanez) idatziak daudela eta Corpus espezializatua dela, kontsumo arlokoa hain zuzen. Hizkuntza horien arteko esaldiak automatikoki lerrokatu dira (Leturia et al., 2010). Corpus honetako laginean 196 esaldi ditugu eta horietatik 173 dira markatuak. Lagin honetan esaldirik luzeenak 63 hitz ditu, eta laburrenak, berriz, 22 hitz.

Lagin horiek aztertuta, esaldi zail edo luzeetan sinplifikatu beharreko hainbat fenomeno sintaktiko aurkitu ditugu. Ezin da ahaztu, bestalde, laginetan zuzenak ez diren esaldiak ere badaudela. Hemen, honenbestez, esaldi luzeetan agertzen diren fenomenoak azalpen labur bat emango dugu. Hala, corpus horietan aurkitutako fenomenoak bi azpiataletan sailkatu dira. Lehendabizikoan (3.1 atala) sintaxi mailako fenomenoak aurkeztuko ditugu eta, bigarreanean (3.2 atala), berriz, estilo-mailakoak. Bi maila hauen (sintaxia eta estiloa) arteko mugak ez dira oso argiak kasu batzuetan, horren adierazle ditugu egitura sintagmatiko luzeak 3.1.4 azpiataleko adibideak. Ondorioz, egitura sintagmatiko luzeak bi mailetan ager daitezke, alde batetik sintaxi mailan aztertzen direlako, baina baita estiloan ere, izen-estiloa estilo bat delako.

Adibideak bistaratzeko moduari dagokionez, esaldiak corpusetan dauden bezala adieraziko ditugu, akats tipografikoak eta ortografikoak zuzendu gabe. Era berean, azpimarraturik nabarmendu nahi ditugun fenomenoak azalduko ditugu.

### ***3.1 Sintaxi mailako fenomenoak***

---

<sup>9</sup> <http://corpus.consumer.es/corpus/>

## ***HAP Masterra 2010/2011 ikasturtea***

Azpiatal honetan sintaxiarekin lotuta dauden fenomenoak aztertuko ditugu. Fenomeno hauek izango dira lehen hurbilpen honetan sinplifikazio-proposamenak egiteko baliatuko ditugunak. Izan ere, sarreran azaldu dugun bezala, lehen hurbilpen honetan sintaxia landuko dugu.

### **3.1.1 Aposizioak eta azalpenezko egiturak**

Azpimultzo honetan gidoien eta parentesien bidez ematen diren azalpen osagarriak (3) eta (4), nahiz aposizio arruntak (5) sartuko ditugu sinplifikatzerakoan antzeko tratamendua izango dutelako.

- (3) Batetik, hura izan delako -baita greziarrak zakuan sartzen baditugu ere- poeta klasikoa, bai garai klasikozaletan bai alde honetatik nahasiagoak izan diren sasoietan -eta honen lekuko onena Eliot-ek Londreseko delakoan egindako hitzaldia dateke-; bestetik, azken mende hauetan ere -antzinatetik bertatik zetorren tradizio sendo eta zabal bati eutsita- Virgilio bezala sentitu nahi izan delako askotan(...).
- (4) Hala ere, badirudi Sabino (Badajozetik fitxatuta), Moha (Barcelona B-tik), Sena (Meridatik), Lekunberri (Compostelatik), Armentano (Elxetik), Shustikov (Santanderko Racingetik), Aitor Ocio (Athleticek utzita) eta Nuno (Coruñaiko Deportivok utzita) ez direla aurtengo fitxaketa bakarrak izango, ez behintzat Javier Miranda presidenteak iragarritako izar baten fitxaketa gauzatzen bada azkenean.
- (5) Batzorde eragileko idazkari instituzionalak, Mikel Mintegik, prentsa ohar batean azaldu zuenez, "Udalbiltzako batzarrean hautatuak diren kideak kanporatzearen aldeko ez gara, baina, aldi berean, giza eskubideen errespetuaren exigentzia batzarrean bertan argitu behar dela uste dugu.

### **3.1.2 Erlatibozko perpausak**

Erlatibozko esaldiak hizkuntzalaritza klasikoan mendeko esaldien artean sailkatzen diren arren, bibliografian ikusi ditugun sailkapenetan (Siddharthan 2002 eta Specia et al. 2008) atal propio batean aurkitu ditugu. Hori dela eta, 4. atalean ikusiko dugun moduan perpaus hauek sinplifikatzerakoan ere ezberdin jokatuko da, euren bereizgarri nagusia aurrekari bat izatea baita.

(6) adibidean bi erlatibo arrunt daude. (7) adibidean, berriz, *zein* izenordaina duen erlatibo egitura.



- (6) XVIII, XIX eta XXgarren mendeetako euskal itzultzaileekin komun dituzten eta gorago aipatu diren asmoez gain, XIV, XV eta XVIgarren mendeetako haiek ere, esaterako, pobrea ikusten zuten beren latinaren edo grekoaren ondoan, eta batzuetan hizkuntzari berari leporatzen zioten errua, beste batzuetan -geroago jeneralean- ez hizkuntzari baizik hiztunen axolagabekeriari... Ez ziren ezer berririk esaten ari:
- (7) Esteban Casado-k 1876ko dokumentu garrantzitsu bat ezagutarazi zuen, zeinak begiztatzen baitu Italiako gobernuak eta Espainiako Koroak elkarren artean egindako tratatu bat, posible egiten zuena oin berriko eraikin bat Akademiarentzat jasotzea, gure herrialdeak elizaren eta ren gain egikaritutako patronatoa aitortu ondoren.

### **3.1.3 Perpaus markatuak**

Perpaus markatuak menderagailu bat duten perpausak dira eta mota askotakoak dira: denborazkoak, ondoriozkoak, baldintzazkoak.... Mota honetako perpausak ez dira beti isolaturik agertzen, hau da, (8) adibidean ikus daitekeen bezala, esaldi batean mendeko perpaus bat baino gehiago egon daitezke, edota (9) adibidean ikusten den moduan mendeko perpausaz osatutako perpaus elkartua izan daiteke.

Simplifikatzerakoan, oro har, esaldi hauek antzeko tratamendua jasoko dute.

- (8) Testamentua egina dagoenean: -seme-alabarik baldin bada: izatezko bikoteetan, zendu den seme-alaba baten ordezkartzan seme-alaba edo lobarik baldin bada, zenduaren bikote-kideak ondasun guztien herena heredatzeko eskubidea du, testamentuan berariaz horrela agertzen bada.
- (9) Genevako Konbentzioaren alderdiek onartu zuten hatsapenaren arabera, babes-eskaera izapidetzeko onartzen ez bada edo ezetza eman ez gero, eskatzaileak Espainiatik kanpora joan behar du, Espanian sartzeko edo egoteko baldintzak betetzen ez baditu behintzat, atzerritarrentzako sistema orokorra kontuan hartuz, edo, salbuespen gisa, gizalegezko edo herri-onurako arrazoiengatik herrialdean geratzeko baimena ematen ez bazaio.

### **3.1.4 Egitura sintagmatiko luzeak**

Nominalizazioaren eta izenlagun ugari erabiltzearen ondorio dira egitura hauek.. IVAPeko Estilo-

### ***HAP Masterra 2010/2011 ikasturtea***

liburuan ematen diren irizpideetan, esaldien luzera 20 hitzekoa izatea proposatzen da (Arakama et al., 2009); muga proposamen horretatik gertu dago (10), adibidez, (11) adibideak, aldiz, muga hori gainditu egiten du.

Egitura zailak dira, eta sinplifikatzeko estilo nominaletik aditz-estilora aldatzea izan daiteke konponbide bat, maiz aditz ez-jokatu batekin osatu direlako (10) eta (11) adibideetan azpimarratutako sintagmek islatzen duten moduan.

(10) Korsikako Lurralde Asanbleak uhartearen bilakaera instituzional eta politiko baten alde proposatutako plana gehiengo oso handiz onartu eta lau egunen buru, Chiracek argi utzi nahi izan du Errepublika frantziarreko batasunaren berme geratzen dela eta, halaber, desbideratze bati oztopo egin lekiola.

(11) Euskal Herriko Unibertsitatearen Errektoregoak, abuztuaren 25eko 11/83 Legearen 3. eta 18. artikuluek, UPV / EHUko Estatutuen 71. eta 75. artikuluek emandako esleipenak erabiliz, eta otsailaren 7ko 315/1964 Erret Dekretuan, abuztuaren 2ko 3/1984 legeari dagokionez, uztailaren 6ko 6/1984 Legeari dagokionez, uztailaren 6ko 6/1989 Euskal Funtzio Publikoari buruzko legearen 91. 92. eta 93. artikuluetan eta indarrean dauden gainerako lege-xedapenetan expreski agindutakoarekin bat etorritik, M Argentina Marcal Lopez andrea, 14.701.203 duena, UPV / EHUko Menpekoen Eskalako Behinbehineko Funtzionaria izendatzen du, aipaturiko plaza behinbehinekotasunez beteko duelarik, berori karrerako funtzionariaren bidez hornitu arte edo izendapena egiteko presazko edo premiazko arazoak desagertu areraino, eta zerbitzuak esakintzearen izaerak ez du funtzionari edo langile laboral finko izateko inolako lehentasun-eskubiderik ekarriko, administrazio-harremanaren moetak horretarako bide ematen duen neurrian Funtzionarien estatutu-erregimena aplikatuko zaiolarik, edonola ere, betetzen duen lanpostuan esleituriko eta beraren titulazio-taldeari dagozkion oinarrizko lansari osagarriak jasoko dituelarik.

### ***3.2 Estilo mailako fenomenoak***

Azpiatal honetan laginetan aurkituriko eta estiloarekin zerikusia duten fenomenoak azalduko ditugu. Fenomeno hauek estilo egokia erabiliz gero, ez lukete sinplifikazioaren beharrik izan behariko. Hala ere, lagineko esaldi hauek luzeak direnez, hauek sortzen dituzten arazoak jakitea beharrezkoa da, beste arlo batzuetan ere tratatuak izan daitezten, adibidez estilo zuzentzailea garatzerakoan.

### **3.2.1 Puntuazio-markak gaizki**

Corpusetan aurkitu dugun beste fenomeno bat da puntuazio-markak ez direla egoki erabili; ondorioz etiketatzerakoan esaldi hauek zail eta luze moduan sailkatu dira. Berez, oker dauden puntuazio-marka zuzendu ondoren, ez genituzke esaldiak zertan sinplifikatu, aurretik aipaturiko beste fenomenorik izan ezean.

Esate baterako, (12) adibidean, *bitartean* eta *Israelgo* hitzen artean puntua jarritz gero, esaldia laburtuko genuke. Puntu hori esaldiaren beste gune batzuetan ere jar daiteke, helburu bera lortuz. (13) adibidean ere antzekoa gertatzen da, *EAK* eta *Erantzukizunak* hitzen artean puntu bat jarrita esaldia ulerterrazagoa litzateke.

(12) Barakek palestinarren egoitzak bonbardatu ditu, bakeaz mintzatu nahi duela dioen bitartean Israelgo Gobernuak Simon Peres lehen ministro ohia eskatu dio bitartekari lana egiteko eta Jasser Arafatekin elkarrizketatzeko Bake ahaleginekin batera, Israelgo helikopteroek Gaza eta Zisjordaniako herriak bonbardatu dituzte Barakek etsi egin du Likud Gobernuan sartzeko ahaleginean; orain bere asmoa da Shas alderdiaren babesarekin irautea.

(13) Aznarrek bere ardurak ekiditeko jotzen du euskal abertzaleen kontra EAJ eta EAren ustez ETA legitimatzen dutela esanez Aznarrek egin zituen adierazpenak gogor kritikatu dituzte EAJk eta EAK *Erantzukizunak* besteei egotzi ordez, bere buruari begiratu eta dagozkion erantzukizunak hartzeko esan diote Aznarri Euskal abertzaletasuna "deslegitimatze" Espainiako gobernuburuak "gezurra" erabili duela azaldu dute EAK eta EAJk.

### **3.2.2 Polisindetonak**

Aurreko azpimultzoan gertatzen zen bezala, arazoak ez du sintaxi zailarekin zerikusirik; honako hauetan juntagailuen gehiegizko erabilera da arazoa. Beraz, estilo-zuzentzaile bat izanez gero edo estilo-liburuetan dauden gomendioak beteko bagenitu, arazo hau konponduko genuke.

Esaterako, (14) eta (15) esaldietan *eta* juntagailuaren bitartez esaldiak gehiegi luzatu dira, esaldia konplexuegia bilakaturik. (16) esaldian, ordea, 4 juntagailu daude, baina esaldi hori literatur generokoa denez, fenomeno hau bertan errekurtsio literario-estilistiko bat da, eta ez estilo akatsa. Beraz, fenomeno hau testu-generoaren arabera sinplifikagarria izango da.

## ***HAP Masterra 2010/2011 ikasturtea***

- (14) Hala, Abeille-Languedoc ontzi-zirgatzaila eta hiru helikoptero joan ziren hondoraketa gertatu egin zen lekurat eta zenbait ordu berantago radar batez eta robot batez hornitutako bi itsasontzi militar zein ere itsas gainerat agertu litekeen kutsadura bat biltzeko gaitasuna duen Alcyon ontzia bertaratu egin ziren
- (15) Italiak indultoak emango ditu kartzelak hustu eta Justizia krisiari aurre egiteko Neurriek delitu txikiak izango dituzte kontuan, eta ez dituzte ustelkeria, mafia eta biolentzia politikoa bere baitan hartuko Eskuineko oposizioak "segurtasun arrazoiak" argudiatu ditu amnistia edo indulto mota guztiak baztertzeko Italian 50.000 preso inguru daude egun, herrialde horretako kartezelek teoriar onartzen dutena ia halako bi.
- (16) Eta orain beste kafetegi aseptiko honetatik, -eguneko hirugarren kafea hartzera sartua-, mahaitxoan esertzearekin batera, gutuna bukatzera noala iragartzen dizuet, eta bukatutakoan, paperontzira bota ere egingo dudala, zeren -eta hau ez esateko idatzi nahi nizuen azken finean-katarro bat baino dexente serioagoa izaten den gaixotasun bat sumatu baitidate joan den astean ginekologoarenean,.

### **3.2.3 Estilo nahasketa**

Zehar-estiloa eta estilo zuzena, biak, erabiltzeak esaldi berean ere zailtzen du esaldiaren irakurketa (17) adibidean ikus daitekeen moduan. Baina esaldi luze hauetan sinplifikazioa egin baino, estilo bakarra erabiltzea izango litzateke konponbidea.

- (17) Garbi asko esan zuen hori Miguel Madariagak, Fundazioko presidenteak taldeko argazkia atera eta gero, "David Etxebarria Euskaltel-Euskadiko elastikoarekin (ONCEkoarekin esan zuen lehenik, eta barre algaratan lehertu ziren kazetariak eta txirrindulariak) aurkeztea da gaur elkartu gaituen arrazoa".

### **3.2.4 Hurrenkera logikoa eta hitz-hurrenkera kanonikoa**

Mendeko esaldi batean perpausen arteko hurrenkera logikoa jarraituz edota subjektu-objektu-aditz (SOV) hurrenkera kanonikoa jarraituz egiten bada, erabiliz gero, gure garunak errazago prozesatzen dituela esan ohi da (Erdozia, 2009). Horrela, bada, hurrenkera hori ez duten esaldiak dira (18), (19) eta (20) adibideak. Ikus daitekeen moduan, baldintzako perpausa den (18) adibidean baldintza

## ***HAP Masterra 2010/2011 ikasturtea***

ondorioaren atzean dago; eta (19) eta (20) adibideetan subjektua aditzaren ondoren dago (OVS) ordenatuko bagenu, ulermena erraztuko genuke.

- (18) Foru-aldundiek zerbitzuak beren kasa eskaini ahal izango dituzte, edo baita zerbitzu propioak dituzten udalekin hitzarmenak egin ere, aldundiaren zerbitzuen iharduera materialak kudeatzeko eremu edo toki zehatzetan, zerbitzuaren eraginkortasunerako egokiagoa baldin bada.
- (19) Hasiera batean aste honetan partidurik ez jokatzea aurreikusita zegoen arren, azken orduan ostiralean Holandara abiatu aurretik partido bat jokatu nahi izan du Alavesek.
- (20) Eroski taldearen irabaziei esker hainbat ekintza egiten du fundazioak: kontsumitzaileen aldeko ekintzak bultzatu, zabalpen kanpainak egin, aldizkari eta gidak argitaratu, bekak eman, hirugarren munduari laguntzen lan egiten duten erakundeekin elkarlanean aritu eta elikagaiak, jostailuak, testu-liburuak etab. biltzeko kanpainetan lagundu, besteak beste.

Orain arte aipatutako fenomenoek gain jarraian aipatzen diren lagin hauetako esaldietan arazoak ere aurkitu ditugu: Hauek dira horietako batzuk: sinplifikazioarekin edo erraztearekin ia harremanik ez dutenak, baina hizkuntzaren prozesamendu automatikoan arazoak sor ditzaketenak:

- Ortografia ez batua: (15) *dexente*; (18) *iharduera*
- Entitateak: (13) *EA, EAJ...* (17) *Euskaltel-Euskadi, David Etxeberria eta Miguel Madariaga pertsona eta erakunde izenak*
- Datak: (8) *XVIII, XIX eta XXgarren mendeak*; (9) *1876*
- Kopuruak: (16) *50.000 preso*
- Legeak (11) *abuztuaren 25eko 11/83 Legearen 3. eta 18. artikuluek*
- Karaktere arraroak...

Arazo horiek sinplifikazioarekin edo erraztearekin ia harremanik ez dute, baina hizkuntzaren prozesamendu automatikoan arazoak sor ditzaketenak dira.

Bestalde, sinplifikazioari begira, azterketa honetan aurkitu ditugun fenomenoak beste hizkuntzetan ere gertatzen direla esan behar da. PorSimples proiektuan (Specia et al., 2008), adibidez, 2. taulako fenomenoak aipatzen dituzte. Gure azterketarekin alderatuta, aditza boz pasiboan duten esaldiak izan ezik, gainontzeko guztiak aurkitu ditugu.

2. taula: PorSimples proiektuan sinplifikatzeko erabili duten eskema

<b>Egiturak</b>
Aposizioak
Erlatibozko perpausak
Mendeko perpausak
Esaldi koordinatuak
Aditz ez-jokatua duten perpausak
Boz pasiboa

Siddharthanek (2002) ere egitura horiek azaltzen ditu, sailkapena beste modu batean egiten badu ere.

## **4 Sinplifikazio-proposamenak**

Hirugarren atalean azaldutako EPEC eta Consumer corpusen laginen azterketan oinarrituz, sinplifikatu nahi ditugun perpaus motak azalduko ditugu hemen, perpaus mota bakoitzari dagokion sinplifikazio-proposamenarekin batera. Aipatu behar da, ikerketa-lan honetan ez ditugula sinplifikazioak perpaus guztientzat proposatuko. Portugeserako egindako azterketan ipintzen duten eskema jarraituz (2. taula) (Specia eta al., 2008), euskararako lehen proposamenak egingo ditugu. Esperimentu honetan, beraz, aposizioak, erlatibozko perpausak, eta perpaus markatuetatik denborazkoak aztertuko ditugu. Perpaus koordinatuak, aditz ez-jokatua duten perpausak eta boz pasiboa (gure laginetan aurkitu ez dugun fenomeno) hurrengo lanetan lantzeko utziko ditugu.

Sinplifikazio maila dela-eta, lan honetan sinplifikazioa naturala gauzatuko dugu aztergai ditugun mendeko perpaus jokatuekin. Hala, esaldi luzeak bi perpausetan banatuko ditugu, esaldi laburragoak sortzeko. Sinplifikazioa bortitza egin nahiko bagenu, aditz ez-jokatuak ere landu beharko genituzke, baina mementoz, hasteko, sinplifikazio naturala egiteko proposamenak egingo ditugu.

Sinplifikazio naturala eta bortitzaren kontzeptuak PorSimples proiektuan garaturiko lanetan ikus ditzakegu. Sinplifikazio naturala egiterakoan, norberak aukera dezake zein fenomenotan egin operazioak. Bortitzean, bestalde, eskuliburu batean zehazturiko operazioak egiten dira, ahal den

## ***HAP Masterra 2010/2011 ikasturtea***

guztia sinplifikatuz. Sinplifikazio-maila hauek irakurleen alfabetatze-mailaren arabera banatzen dituzte. Horrela, bada, sinplifikazio bortitza gaixotasun kognitiboak dituztenei eta irakaskuntzako (*ensino fundamental*) bigarren eta laugarren maila arteko irakurleei, hau da, alfabetatze-maila pobrea dutenei (*Rudimentary literacy level*), zuzenduta dago, eta naturala, berriz, irakaskuntzako 5. mailan daudenei, hots, oinarritzko alfabetatze-maila (*Basic literacy level*).

Gure adibideetan oro har sinplifikazio naturala egingo dugun arren, zenbaitetan esaldian agertzen diren fenomeno guztiak sinplifikatuko ditugu sinplifikazio bortitzaren adibideak (25b, 26b) ere ikusi ahal izateko.

Jarraian, 4.1, 4.2 eta 4.3 azpiataletan, aposizioak eta azalpenezko egiturak, erlatiboak perpausak eta denborazko perpausak sinplifikatzeko proposamenak azalduko ditugu, hurrenez hurren. Horiek azaltzeko erabili ditugun adibideei dagokienez, lehendabizi jatorrizko esaldia agertuko da zenbakiturik eta alfabetoko lehen letrarekin. Jatorrizko esaldi hau corpusean agertzen den bezala (akats tipografikoak eta ortografikoak zuzendu gabe), adierazko da. Jatorrizkoaren ondotik, moldaketak dituen esaldia etorriko da jatorrizkoaren zenbakia berarekin eta alfabetoko hurrengo letrarekin; jatorrizkoan ez bezala, hau zuzendu egingo dugu, zuzenketa-lana sinplifikazioan ere kontuan hartzen delako. Adibide bakoitzean, halaber, azpimarraturik, moldatu behar diren egiturak adierazi ditugu 3.1. eta 3.2 azpiataletan bezala, eta letra etzanez, moldaketen ondorioz esaldietan ageri diren aldaketarik garrantzitsuenak.

### **4.1 Aposizioak eta azalpenezko egiturak**

Hizkuntza erromaniko eta germaniko askotan ez bezala, euskaraz izen sintagma barneko aposizio arruntak, izen batekin batera azalpena edo kargua (hondarkia) azaltzen dutenak, ez dira koma artean idazten<sup>10</sup>, estilo akatsik ez badago behintzat. Horren adierazle ditugu (21a) eta (21b) adibideak:

(21a) *La secretaria general de Instituciones Penitenciarias, Mercedes Gallizo*

(21b) Mercedes Gallizo Espetxeetako Zuzendaritzako idazkari nagusiak (eitb.com, 2011)

Beraz, beste hizkuntza horiekin alderatuz, sintagma barneko aposizioak detektatzea

---

<sup>10</sup> [http://www.euskaltzaindia.net/dok/arauak/Araua\\_0109.pdf](http://www.euskaltzaindia.net/dok/arauak/Araua_0109.pdf)

## ***HAP Masterra 2010/2011 ikasturtea***

zailagoa da gure kasuan.

Dena dela, baditugu beste egitura batzuk aposizioen antzeko funtzioa dutenak, detektatzeko errazak direnak. Egitura horiek, tartekiak, alegia, azalpenak emateko erabiltzen dira (batzuetan perpausaren egituratik aparte) eta koma, gidoi eta parentesi artean idazten dira. Horiek sinplifikatzeko proposamenak aposizio arruntak sinplifikatzeko proposamenen antzekoak direnez, atal honetan landuko ditugu.

Aposizio arruntan kasuan, sinplifikazioa honela egingo dugu:

- i) Esaldia sintagmetan banatuko dugu. Horrela, (21b) adibidean ikus daitekeen bezala, aposizioa duen sintagma bi ataletan banatuko dugu: alde batetik, izena *Mercedes Gallizo*, eta bestetik, azalpena edo hondarkia *Espetxeetako Zuzendaritzako idazkari nagusiak*.
- ii) Izena jatorrizkoan zegoen moduan esaldi nagusian utziko genuke, azalpenik edo hondarkirik gabe eta puntu bat jarriko genuke.
- iii) Lehendabiziko esaldi hori amaitzerakoan, beste esaldi batean izena eta azalpena *izan* aditzarekin lotuko genituzke eta esaldi berri hori aurrekoaren ondotik idatziko genuke.
- iv) Esaldi berria osatzerakoan, ezin dugu ahaztu komunztadura ere moldatuko beharko dugula kasu batzuetan.

Jarraian, (22a) aposizioz beteriko esaldia dugu eta (22b) esaldi bera sinplifikatuta:

(22a) Pankarta eraman zuten, besteak beste, Jose Maria Aznar Espainiako presidentea, Javier Arenas PPko idazkari nagusiak, Manuel Chaves PSOEko lehendakariak, Gaspar Llamazares IUko koordinatzaile nagusiak, Adolfo Suarez eta Leopoldo Calvo Sotelo Espainiako Gobernuako presidente izandakoak, Candido Mendez UGTko buruak eta Jose Maria Fidalgo CCOOko buruak.

(22b) Pankarta eraman zuten, besteak beste, JM Aznarrek, Javier Arenasek, Manuel Chavesek, Gaspar Llamazaresek, Adolfo Suarezek eta Leopoldo Calvo Sotelok, Candido Mendezek eta Jose Maria Fidalgok. Jose Maria Aznar Espainiako presidentea *da*. Javier Arenas PPko idazkari nagusia *da*. Manuel Chaves PSOEko lehendakaria *da*. Gaspar Llamazares IUko koordinatzaile nagusia *da*. Adolfo Suarez eta Leopoldo Calvo Sotelo presidente izandakoak *dira*. Candido Mendez UGTko burua *da*. Jose Maria Fidalgo CCOOko burua *da*.



## ***HAP Masterra 2010/2011 ikasturtea***

Gizakientzat (22b) oso sinplea eta egitura aldetik errepikakorra dirudien arren, oso kontuan izan behar dugu euskararen tratamendu automatikorako baliagarria zaigula esaldi errazagoak lortzea.

Aurreko jarraibide horiek berak aplikatuko ditugu (23a) esaldian; honetan, ordea, aposizioak izan ordez, parentesi arteko azalpenak (tartekiak) agertzen dira. Zehatzago esanda hau egingo dugu:

- i) Tartekiak esalditik atera ditugu.
- ii) Tartekien barnean dagoena *izan* aditzarekin lotu dugu, perpaus nagusiaren ondoren, perpaus berri horiek gehituz.

(23a) Hala ere, badirudi Sabino (Badajozetik fitxatuta), Moha (Barcelona B-tik), Sena (Meridatik), Lekunberri (Compostelatik), Armentano (Elxetik), Shustikov (Santanderko Racingetik), Aitor Ocio (Athleticek utzita) eta Nuno (Coruñako Deportivok utzita) ez direla aurtengo fitxaketa bakarrak izango, ez behintzat Javier Miranda presidenteak iragarritako izar baten fitxaketa gauzatzen bada azkenean.

(23b) Hala ere, badirudi Sabino, Moha, Sena, Lekunberri, Armentano, Shustikov, Aitor Ocio eta Nuno ez direla aurtengo fitxaketa bakarrak izango, ez behintzat Javier Miranda presidenteak iragarritako izar baten fitxaketa gauzatzen bada azkenean. Sabino Badajozetik fitxatua *da*. Moha Barcelona Btik fitxatua *da*. Sena Meridatik fitxatua *da*. Lekunberri Compostelatik fitxatua *da*. Armentano Elxetik fitxatua *da*. Shustikov Santanderreko Racingetik fitxatua *da*. Aitor Ocio Athleticek utzita *da*. Nuno Coruñako Deportivok utzita *da*.

Egin duguna (23b) esaldian ikus daiteke. Hala ere, elipsiaren arazoarekin egiten dugu topo. Arazo hori konpontzeko, aurreko sintagmaren azalpenenean dagoen aditza berreskuratu dugu, (23b) kasuan *fitxatua*, eta azalpenik ez duen sintagmarako berrerabili dugu.

(24a) adibidean, berriz, parentesi artean dugun azalpena beste kasuetan baino luzeagoa denez, aurreko adibide horietan erabili ditugun irizpideak jarraitzeaz gain, izena genitibo kasuarekin osatu ondoren, esaldiaren ardatza (kontzeptu nagusia) ere gehitu diogu; izan ere, azalpena gehitzen zaien kontzeptuak izen horien azpikontzeptuak dira. Hau da, (24a)ko *lana*, *urtaro* eta *uzta* gertaera motak dira, eta horrelako kontzeptualizazioak *Wordneten* euskarazko bertsioaren (Pociello et al., 2010) bitartez lor genitzake, ezagutza base horretan kontzeptuak

## ***HAP Masterra 2010/2011 ikasturtea***

mailakaturik agertzen direlako. Ulermenaren aldetik (24b) arraroa gerta badaiteke ere, sintaxiaren aldetik zuzena eta sinplea da. Dena den, esan behar da horrelako esaldiak moldatzea zailagoa iruditzen zaigula.

(24a) Berbera gertatzen da, besteak beste, lana (bai lanera joateko, itzultzeko edo egiteko, erritmoa markatzeko erabili izan diren kantu eta doinuak), urtaro (urtaro aldaketak ospatzeko festak, besteak beste udaberri-ihauteriak, uda - San Juan, udazkena - San Martin, negua-eguberriak), uzta (lurra esnarazteko, eguraldi egokia lortu ahal izateko, eta bestelako ekintzak), gizarteko gertaera ezberdin, e.a.-ekin lotua dagoen musikarekin.

(24b) Berbera gertatzen da, besteak beste, lana, urtaro, uzta, gizarteko gertaera ezberdin, e.a.ekin lotua dagoen musikarekin. Laneko *gertaerak* bai lanera joateko, itzultzeko edo egiteko, erritmoa markatzeko erabili izan diren kantu eta doinuak *dira*. Urtaroko *gertaerak* urtaro aldaketak ospatzeko festak, besteak beste udaberri-inauteriak, uda - San Juan, udazkena - San Martin, negua-eguberriak *dira*. Uztako *gertaerak* lurra esnarazteko, eguraldi egokia lortu ahal izateko, eta bestelako ekintzak *dira*.

Beste adibide honetan, (25a)n, alegia, aditz jokatua ageri da tartekian. Kasu honetan ere, (25a) aurreko esaldietan bezala jokatu dugu, baina tartekiko aditz hori aditz nagusia den kasuetan, ez diogu aditzik gehituko. Zehazki hau da egingo duguna:

- i) Esaldia sintagmatan banatu dugu.
- ii) Parentesi artean dagoena esalditik atera dugu eta bestea parentesirik gabe utzi.
- iii) Esaldi berria sortzeko, aurreko esaldiko izenak berreskuratu ditugu.
- iv) Izen bakoitza bere parentesi barruan zegoenarekin batu dugu.

Hori guztia eginda lortuko dugun esaldia da (25b).

(25a) lagin guztietan baliaturiko koipea landarezkoa den arren, hiru jatorrikoa da: soja olioia (aurkituriko hiruretan osasuntsuena, batik bat gantz poliintsaturatuez osatua dagoelako) sei laginetan; koltza olioia (gantz monointsaturatutzat jotzen delarik, ez da sojakoa bezain osasuntsua baina egokia da) froxá laginean; palmondoko olioia (hiruretan apalena, osasun sailkapenari dagokionez, bere gantzaren %90 saturatua baitu) elmar laginean

(25b) Lagin guztietan baliaturiko koipea landarezkoa den arren, hiru jatorrikoa da: soja olioia sei laginetan; koltza olioia froxá laginean; palmondoko olioia elmar laginean. *Soja olioia* aurkituriko hiruretan osasuntsuena *da*, batik bat gantz poliintsaturatuez osatua dagoelako. *Koltza olioia* gantz monointsaturatutzat jotzen delarik, ez da sojakoa bezain osasuntsua baina egokia da. *Palmondoko*

*olioa* hiruretan apalena *da*, osasun sailkapenari dagokionez, bere gantzaren %90 saturatua baitu

Hurrengo urratsetan sakonago aztertu beharko dugu zer egin aditz jokatua duten esaldi horiekin. Hala ere, esaldi horiek denborazkoak edo erlatibozkoak izanez gero, atal horietan aipaturiko operazioak egin ditzakegu, 4.2 eta 4.3 ataletan azalduko dugun bezala.

## **4.2 Erlatibozko perpausak**

Erlatibozko perpausak funtzio adjektibala dute eta gainerako mendeko perpaus modifikatzaileak ez balaza, izen kategoriako hitzaren mende daude. Perpaus hauen ezaugarri nagusia da bi perpausetan agertzen den izen sintagma bat dutela (ardatza), sintagma hori perpaus batean isilduta dagoen arren (Euskaltzaindia, 1999). Esaldian ardatzarekiko duten kokalekua ikusirik, erlatibozko perpausa laguntzen dion izenaren ezkerrean (26 eta 27 adibideak) ager daiteke edota izenaren eskuinean (28, 29 eta 30 adibideak). Hauek perpaus erlatibo arrunten barnean sartzen dira (Euskaltzaindia, 1999) eta atal honetan hauek eta *zein* izenordaina dutenak tratatuko ditugu, erlatibozko beste perpaus motak (bait- erlatiboak, korrelazioz moldaturikoak, aurrekaririk gabeko galde izenordaina dutenak) dauden arren.

Jarraian, (26) esaldian, azpimarraturik dagoen perpaus erlatiboan, ardatza (*Ollanta Moises Humala*) esaldi nagusiaren subjektua da. Honela, bada, konponbidea hauxe da:

- i) Esaldia perpausetan banatu dugu (nagusia eta mendekoa).
- ii) Mendeko perpausaren aditzari menderagailua kendu diogu.
- iii) Aditza aditz nagusi bihurtu dugu.
- iv) Puntua jarri dugu perpausaren amaieran.
- v) Ardatza errepikatuz esaldi nagusiaren subjektua bihurtu dugu.

Pauso horiek guztiak eginda lortu da (26b) adibidea. Esaldi honetan (26a) aposizioak ere daudenez, sinplifikazio bortitza egin nahi bada, horiek 4.1 puntuan proposatutako irizpideekin molda ditzakegu: Honela, sinplifika daitezkeen fenomeno guztiak sinplifikatu ondoren, (26c) lortuko dugu.

(26a) JOAN den igandean geroztik Alberto Fujimori Peruko presidentearen aurka altxamendu militar bat

## ***HAP Masterra 2010/2011 ikasturtea***

gizatzen ari den Ollanta Moises Humala teniente koronelak ez du uste bakarrik dagoenik (...)

(26b) Joan den igandeaz geroztik Alberto Fujimori Peruko presidentearen aurka altxamendu militar bat gidatzen ari *da Ollanta Moises Humala teniente koronela. Ollanta Moises Humala teniente koronelak ez du uste bakarrik dagoenik (...)*

(26c) Ollanta Moises Humala teniente koronela *da*. Alberto Fujimori Peruko presidentea *da*. Joan den igandeaz geroztik Ollanta Moises Humala Alberto Fujimoriren aurka altxamendu militar bat gidatzen *da. Ollanta Moises Humallak ez du uste bakarrik dagoenik.*

Hala ere, 4.4 azpiatalean aztertuko dugu proposamen-erregela horien hurrenkerak nola aplikatu.

Ardatza esaldi nagusiaren objektua den kasuetan (27a), antzeko prozesua egin behar dugu. Ardatza bigarren esaldiaren subjektu izan beharrean, objektua izango da. Aztergai dugun (27a) adibideko bigarren perpaus erlatiboa (“*gero “ ” (...)* *plazaratu zituenak*”) (30) adibidearekin zehatzago azalduko dugun proposamenaren bitartez sinplifikatu dugu. (27c) adibidean, berriz, 4.3 azpiataletan denborazko perpausarekin adieraziko ditugun moldaketak ere aplikatu ditugu, sinplifikazio bortitzagoa eginez.

(27a) Mitxelenak fonetika eta morfosintaxia alorretan egin zitezkeen galderak zehaztu zituen bitartean, aditzaren barrutian egin zitezkeenak tajutu zituen Yrizarrek, gero " " [ orain I 155-170 ] txostenean plazaratu zituenak

(27b) Mitxelenak galderak zehaztu zituen. *Galdera (horiek)* fonetika eta morfosintaxi arloan egin zitezkeen bitartean, aditzaren barrutian egin zitezkeenak taxutu zituen Yrizarrek. Gero “ ” *horiek* [orain I 155-170 ] txostenean plazaratu zituen.

(27c) Mitxelenak galderak zehatu zituen. *Galdera (horiek)* fonetika eta morfosintaxi arloan egin zitezkeen. Bitartean, aditzaren barrutian egin zitezkeenak taxutu zituen Yrizarrek. Gero “ ” *horiek* [orain I 155-170 ] txostenean plazaratu zituen.

Perpaus erlatiboak izenaren eskuinera agertzen direnean (28a), lehenik eta behin begiratu behar dugu ea erlatibozko perpaus horrek *zein* izenordaina duen, horrek lagunduko baitu perpausa identifikatzen. Kasu askotan, ordea, ez dugu hori aurkituko eta komen erabileraren bitartez identifikatu beharko ditugu (30a eta 31a). Horrela, bada, (28a) adibidearen eta antzeko egituren sinplifikazioa egiteko pausuak hauek dira:

- i) Esaldi nagusia jarraian idatzi, koma artekoa kenduta.

## ***HAP Masterra 2010/2011 ikasturtea***

- ii) Esaldi nagusiko puntuaren ondoren, esaldi nagusiko ardatzari erakuslea (dagokion kasu-markarekin) edo determinatzailea erantsi.
- iii) Mendeko perpausari menderagailua kendu.
- iv) Aditza aditz nagusi bihurtu ondoren, amaieran puntua jarri eta perpaus nagusiaren ondotik idatzi.

Aurreko pausoak eginak dituen esaldia dugu (28b).

(28a) 1873ko urriaren 6ko dekretu gehigarri batek, zeina Errepublikako Gobernu-Presidente Emilio Castelar-ek eta Estatuko Ministro, Jose de Carvajal-ek izenpetu baitzuten, bi pentsio gehitzen zituen merituzko plaza banarentzat, arkitektura eta sakongrabortuko alorretan.

(28b) 1873ko urriaren 6ko dekretu gehigarri batek bi pentsio gehitzen zituen merituzko plaza banarentzat, arkitektura eta sakongrabortuko alorretan. *Dekretu hori/Dekretua* Errepublikako Gobernu-Presidente Emilio Castelar-ek eta Estatuko Ministro, Jose de Carvajal-ek izenpetu *zuten*.

Aurreko operazio hori bera egingo dugu *zein* izenordaina deklinatuta duten perpausetan. Hemen ere izen-ardatzari erakuslea edo determinatzailea gehituko diogu (29b) adibidean ikusten den moduan.

(29a) dantzari baten esfortzua ikaragarri handia da, eta bere ibilbidea oso luzea, arteari emandako edozein pertsonarena bezalakoa, zeinaren sakrifizioa eta ikasketak etengabeak eta oso zorrotzak diren, baina ez da horregatik mitifikatu behar .

(29b) Dantzari baten esfortzua ikaragarri handia da, eta bere ibilbidea oso luzea, arteari emandako edozein pertsonarena bezalakoa. *Dantzariaren/ dantzari horren* sakrifizioa eta ikasketak etengabeak eta oso zorrotzak *dira*, baina ez da horregatik mitifikatu behar.

Erlatibozko perpausa izen-ardatzaren eskuinean *zein* izenordaina elidituta agertzen den kasuetan (30a), honela egingo dugu:

- i) Perpaus nagusia osatu,
- ii) Mendeko perpausa osatzeko, ardatza dagokion funtzioan jarri
- iii) Aditzari menderagailua kendu
- iv) Ezin dugu ahaztu esaldia moldatzerakoan, kasu marka dagokion kasuan jarri behar direla.

## ***HAP Masterra 2010/2011 ikasturtea***

Pauso hauek jarraitu ondoren, (30a) adibidearen sinplifikazioa gauzatu dugu eta (30b) da horren emaitza:

(30a) alabaina, internetez kanpo, bi paziente mota bereizi behar ditugu: pasiboak, sendagilearengana era paternalean doazenak, medikuaren eskuetan bere sendatze-prozesu ia osoa uzten dutenak, alegia; bestetik, aktiboak ditugu, fakultatibo bat baino gehiagoren iritzia bilatu eta, areago, erreklamazioak egiten dituztenak.

(30b) Alabaina, internetez kanpo, bi paziente mota bereizi behar ditugu: pasiboak, bestetik, aktiboak ditugu. *Pasiboak* sendagilearengana era paternalean *doaz*. *Pasiboek* medikuaren eskuetan bere sendatze-prozesu ia osoa uzten *dute*, alegia. *Aktiboek* fakultatibo bat baino gehiagoren iritzia bilatu eta, areago, erreklamazioak egiten *dituzte*.

Erlatibozko perpausak izenaren ezkerrean nahiz eskuinean agertzen diren adibideak ere aurkitu ditugu, esaterako (31a) Horrelakoetan, (31b)n azaltzen den moduan, azpiatal honetan aipaturiko proposamenak aplikatuko ditugu. Hemen izan dezakegun arazoa hurrenkera da; teoriarik ez dago arazorik bata bestearen aurretik jartzeko, baina esaldi originalaren hurrenkera mantenduz banaturiko lehendabiziko esaldia ezkerrekoa izango da eta bigarrena, berriz, eskuinaldekoa.

(31a) (...) hor daude Samaniego bera, Iriarte eta gaur ezezagunak egiten zaizkigun beste alegilari asko, denak La Fontainerengandik esku beteka hartu zutenak

(31b) (...) hor daude Samaniego bera, Iriarte eta beste alegilari asko. *Alegilari(ak) (batzuk)* gaur ezezagunak egiten zaizkigu. *Denak* La Fontainerengandik esku beteka hartu *zuten*.

### ***4.3 Denborazko perpausak***

Denborazko perpausak adberbioen funtzioak izaten dituzte eta esaldi nagusiarekiko denbora-harremana adierazten dute. Horretarako, mendeko perpaus hauen menderagailuari kasu-markak eta postposizio konplexuak gehitzen zaizkie, eta marka horiek izango dira perpaus hauek sinplifikatzeko erabiliko ditugun gakoak. Adierazi behar da azpimultzo hau oso zabala dela, baina, oro har, sinplifikazio proposamenak talde guztirako balio duenez, adibide adierazgarri batzuk besterik ez ditugu azalduko.

## ***HAP Masterra 2010/2011 ikasturtea***

Denborazko perpausak postposizio konplexua duenean (32a) honako hau egingo dugu:

- i) Esaldia perpausetan banatu
- ii) Mendeko perpausari menderagailua kendu
- iii) Mendeko perpausari aditza aditz nagusi bihurtu,
- iv) Postposizioa beste esaldian (jatorrizkoan nagusia zena) adberbio kategoriarekin jarri.

Bestalde, ulergarritasunaren aldetik ere errazagoa izan dadin, mendeko perpausa, jatorrizko esaldian hasieran ez badago, lehendabiziko perpausa izatea proposatzen dugu.

Pauso horiek (32a) esaldiari aplikatuz dizkiogu (32b) lortzeko:

(32a) ospitalean egongo den bitartean, pazientea bere ebakuntzatik suspertuko da, eta transplanteak zuzen funtzionatzeko egin behar duena azaltzen zaio, izan ere, bere bizitzako gainerako denboran ohitura zorrotz batzuei jarraitu beharko die eta medikamentua hartu .

(32b) Ospitalean egongo *da. Bitartean*, pazientea bere ebakuntzatik suspertuko da. (...)

Azpimultzo honetan badira menderagailuarekin batera beste osagai bat duten eta adberbio kontsideratu ezin diren beste egitura batzuk ere: *-en bakoitzean*, *-en guztietan*... Horren adibide dugu esaterako (33a)n jasotako esaldia. Hauekin aurrekoekin bezala, honako hau egingo dugu:

- i) Esaldiak zatituko ditugu mendeko perpausari menderagailua kenduz,
- ii) Aditza aditz nagusi bihurtuko dugu
- iii) Adberbio bat jarri ordez, sintagma bat jarriko dugu.

Hau da, lehendabizikoa *memento/aldi bakoitzean* sintagmarekin ordezkatzeari proposatzen dugu, eta bigarrena, *memento/aldi horietan guztietan* sintagmarekin.

(33a) Kristaudiko xakelari onena ezin da xakejokalari onena baino askoz gehiago izan; baina an trebea izateak arrakastarako gaitasuna inplikatzeko du garrantzi handiagoko beste eginkizun batzuetan, adimena adimenaren kontra borrokatzen den guztietan.

(33b) (...) baina hain trebea izateak arrakastarako gaitasuna inplikatzeko du garrantzi handiagoko beste eginkizun batzuetan. (*Memento*) *horietan guztietan* adimenaren kontra borrokatzen *da*.

Denborazko esaldiak postposizio konplexurik ez badu, hau da, menderagailuari

## ***HAP Masterra 2010/2011 ikasturtea***

atxikita bakarrik kasu-marka badu (34a) eta (35a), gure proposamena honako hau da, eta horein emaitzak (34b) eta (35b):

- i) Esaldia banatu
- ii) Aditzari menderagailua eta kasu-marka kendu
- iii) Aditza nagusi bihurtu eta esaldi nagusia sortu
- iv) Hurrengo esaldian, *ordu*-ri aurreko menderagailuak zuen kasu-marka jarri, hau da, *orduan*, *ordutik*... adberbio bat bezala.

(34a) erabakia hartu behar izan zuenean, ez zuen inolako zalantzarik izan don Polikarpo Gogorzak, polizia edo zainzailea izan nahi zuela aitortu baitzuen behin eta berriro, (...)

(34b) Erabakia hartu behar izan zuen. *Orduan*, ez zuen inolako zalantzarik izan don Polikarpo Gogorzak, polizia edo zainzailea izan nahi zuela aitortu baitzuen behin eta berriro, (...)

(35a) ondorioen arabera, azken hamabi urteetan kilimanjaro mendiko izotz-azalaren herena gutxienez urtu egin da, eta 1912an lehen aldiz kartografiatu zenetik, gutxi gorabehera izotz-kaparen %82 galdu du

(35b) Ondorioen arabera, azken hamabi urteetan Kilimanjaro mendiko izotz-azalaren herena gutxienez urtu egin da, eta 1912an lehen aldiz kartografiatu *zen*. *Ordutik* gutxi gorabehera izotz-kaparen % 82 galdu du.

Kasu batzuetan (36a) mendeko perpausaren eta perpaus nagusiaren arteko hurrenkera aldatzea mesedegarria gerta daiteke esaldiaren ulergarritasuna errazteko. Perpausen hurrenkera bakarrik aldatzen den kasuei, hau da, operaziorik egiten ez denean, erraztea deituko diegu. (36a) adibidean, esaldia moldatzeko, bi prozesuak gauzatu ditugu: erraztea eta sinplifikazioa. Adibide honetan gertatzen den bezala, juntagailu batek (adibide honetan *baina*) moldatu behar den perpausa hurrengoarekin lotzen badu, denborazkoa aurrera eramatean egitura arraroak sortzea saihesten da. (36a) esaldia gehiago sinplifikatu nahi izanez gero, *Vladimir Putin Errusiako presidentek* aposizioa ere molda dezakegu.

(36a) Kolokan jarri zen Errusiak eta Frantziak iraganean izan zuten harreman ona Txetxeniako gerra piztu zenean, baina Vladimir Putin Errusiako presidentek herenegun Parisen, Errusia eta Europako Batzordearen arteko bileran, konponbide politikoaren alde egindako adierazpenek berreskuratu egin dute antzinako harremana eta elkarlana eman du fruitu.

(36b) Txetxeniako gerra piztu *zen*. *Orduan*, kolokan jarri zen Errusiak eta Frantziak iraganean izan zuten



## ***HAP Masterra 2010/2011 ikasturtea***

harreman ona, baina Vladimir Putin Errusiako presidentea herenegun Parisen, Errusia eta Europako Batzordearen arteko bileran, konponbide politikoaren alde egindako adierazpenek berreskuratu egin dute antzinako harremana eta elkarlana eman du fruitu.

Denborazko perpausetarako egin ditugun proposamen guztiak orokorrean ikusirik, lehendabiziko hiru pausoak errepikatzen direla ondorioztatu dezakegu. Hiru pauso horiek lan honetan tratatuko ez ditugun beste egiturentzako ere baliagarriak dira, beraz, hemendik aurrera ere kontuan hartu beharrekoak, 4.5 azpiatalean laburbilduko dugun bezala.

### ***4.4 Sinplifikazio erregelen hurrenkera***

Simplifikazio-erregelen hurrenkerari dagokionez, memento proposamen multzo txiki bat besterik aurkeztu ez dugunez, erregela horien hurrenkera nola gauzatuez da oso garrantzitsua, baina etorkizunean zehatuko beharko litzatekeen puntu nagusi bat izan behar du, sistematikoki jokatzeko ezinbestekoa izango dena.

Dena dela, sinplifikazio-erregelak hurrenkera desberdina gauzatuta lortu ditugun emaitzak aztertuko ditugu jarraian.

(37a) adibidean erlatibozko esaldi bat dugu eta ondoren bi parentesi arteko azalpenezko egitura. Lehendabiziko moldaketan erlatibozkoa moldatu dugu eta bigarrean, aldiz, azalpenezko egiturak. Azken hauek sinplifikatzean gizakioi egitura arraroak iruditzen zaizkigunak izan daitezke, baina ezin dugu ahaztu, gure helburua euskararen tratamendu automatikoan esaldi luzeak laburragoak bihurtzea dela, betiere zuzentasunari eta jatorrizko esanahiri eutsiz. Problematika hau (26a) esaldian ikusi dugu, baina tartekia izan ordez, aposizio arrunt batekin. Bertan (26c) moldaketan aposizioa jarri dugu lehenik, baina (37b) adibidean egin dugun bezala, erlatiboaren sinplifikazioa aurretik jar daiteke. (37c)n, ordea, aposizioen moldaketak jarri ditugu erlatibozko perpausaren moldaketaren aurretik.

(37a) zurgaturiko zikinkeria biltzen duen zorroak (paperezkoa , denetan ere) 1,8 litrotik (taurus eta panasonic) 3 litro (bosch) bitarteko edukiera du .

(37b) Zorroak xurgaturiko zikinkeria biltzen *du*. Zorroak 1,8 litrotik 3 litro bitarteko edukiera du. Zorroa paperezkoa denetan ere *da*. 1,8 litrotik zorroak Taurus eta Panasonic *dira*. 3 litrotik zorroa Bosch *da*.

## **HAP Masterra 2010/2011 ikasturtea**

(37c) Zorroak 1,8 litrotik 3 litro bitarteko edukiera du. Zorroa paperezkoa denetan ere *da*. 1,8 litrotik zorroak Taurus eta Panasonic *dira*. 3 litrotik zorroa Bosch *da*. Zorroak xurgaturiko zikinkeria biltzen *du*.

Erlatibozko arrunta eta denborazkoa badaude mendeko perpaus batean (37a) erlatibozkoa lehendabiziko posizioan izatea proposatuko dugu. Hala ere, (38a) adibideko esaldi horrekin, beste bi arazo aurkitzen ditugu: hitz hurrenkera ez kanonikoa eta ezezkotasuna. Beraz, (38b)n hurrenkera aldaketa ere egin beharko genuke, eta ezezkotasunari dagokionez, hurrengo lanetan tratatzeko beste arazoen zakuan sartu beharko genuke, askoz zailagoa baita ezezko esaldi bat prozesatzea baiezko bat baino.

(38a) Askatasun eta eskubideak bermatzen ez dituzten erregimenak egoteko baldintzak jartzen ez diren bitartean, logikoa da pertsona taldeak izatea alde egitera behartuta egongo direnak; (...)

(38b) *Erregimenek* askatasun eta eskubideak *ez dituzte bermatzen*. Erregimenak egoteko baldintza *ez dira* jartzen. *Bitartean* logikoa da pertsona taldeak izatea alde egitera behartuta egongo direnak; (...)

### **4.5 Sinplifikatzeko irizpideak, laburbilduz**

Laugarren atalean aipatu ditugun adibide horietan guztietan egin behar dira. Nahiz eta esaldiak tipologikoki ezberdinak izan, pauso batzuk esaldi guztietan egin behar dira. Hau beste hizkuntzetan egindako lanetan ere gertatzen da (PorSimples, 2008; Chandrasekar et al., 1996; eta Siddharthan, 2003). Pausoak, oro har, hauek dira:

- i) Esaldiak analizatu eta mendekoak badira, banatu. Horretarako esaldien mugak jakin behar dira.
- ii) Esaldi horiek sailkatu behar ditugu, nola sinplifikatu behar diren jakiteko. Lan honetan, pauso hauek eskuz egin ditugu, baina behin egitura guztiak sailkatuta izanez gero, sinplifikazioari begirako sailkapena automatikoki egitea lortuko genuke.
- iii) Sinplifikatzea izango litzateke, goiko proposamenak aplikatuz. Normalena izango da esaldi batean operazio bat baino gehiago egin behar izatea. Horretarako hurrenkera zehaztu beharko dugu.
- iv) Berreraikitako esaldien zuzentasuna kontrolatu beharko da (komunztadura, kasu-markak...), ezin baitugu esaldi akastunak sortu.

Pauso horiek sintaxia sinplifikatzearen laburpena izango lirateke; hala ere, testuak erraztea sintaxia sinplifikatzea baino gehiago da. Beraz, hurrengo pausoa lexikoa sinplifikatzea izango litzateke, baina hori ez dugu lan honetan aztertuko. Hemen esaldi jakin batzuk banan-banan tratatzen ari garenez, ez dugu testu-mailako kohesioaren moldaketari buruz hitz egin, baina ezin da ahaztu hori ere testu sinplifikazioaren atal bat dela.

## **5 Baliabideak eta tresnak**

Orain arte lan honetan euskarazko corpus laginetan aurkitu duguna (3.atala) eta perpaus mota jakin batzuekin egin dezakeguna (4.atala) azaldu dugu. Lan hori guk egin dugun bezala eskuz egin daiteke, baina horrek corpus handi batekin luze joko luke. Horrela, bada, lan hori errazteko, baliagarriak izan daitezkeen tresnak erabil ditzakegu. Ixa taldean dauzkagun tresnen eta baliabideen artean, bi dira egokienak lehen urrats batean: Muga Gramatika (MuGa) eta Idazlanak Autoebaluatzeako Sistema (IAS). Hala ere, lan honetarako morfologian, sintaxian, semantikan eta pragmatikan egin diren lan guztiak kontuan hartzea beharrezkoa da, testu bat analizatzeko eta horren sinplifikazioa sortzeko, maila horietan analizatzea beharrezkoa baita.

Jarraian baliabide horien berri emango dugu modu laburrean.

### **5.1 Muga Gramatika**

Muga Gramatika (Aduriz et al., 2006) ezagutza linguistikoan oinarrituta osatu den baliabidea da. Izenak berak adierazten duen bezala, testu batean esaldien arteko mugak ipintzen ditu. Esaterako, (39) eta (40) esaldietan denborazko perpausen ondoren ezarri du muga esaldi bakoitza osatzen duten bi perpausen artean marra bertikala ( | ) txertatuta.

(39) Hemen dagoenetik | , oso alai dabil.

(40) Hurbiltzen hasi zirenean | , Kurt asaldatzen hasi zen.

MuGa gramatika *Constraint Grammar* formalismoan (Karlsson et al., 1995) oinarrituta idatzi da, eta gramatika bera aplikatzeko CG-2 *parser*a (Tapanainen, 1996) erabili da.

## HAP Masterra 2010/2011 ikasturtea

Bestalde, ezagutza linguistikoan oinarritutako MuGa gramatikaz gain, bada Ixa taldean estatistikan oinarrituta garatu den beste muga gramatika bat (Arrieta, 2010). Estatistikan oinarritutako gramatika honek perpausen arteko mugak identifikatzen ditu puntuazio-markak baliatuta. Horrela, testu batean puntuazio-marka gogorrak (puntu, bi puntu, hiru puntu, puntu eta koma, harridura-marka, galdera-marka) badaude, puntuazio-marka gogor horien ondotik eta hurrengo esaldiaren aurretik bi etiketa jartzen dira; esaterako, esaldi baten amaierako punturen ondotik *S*) etiketa, esaldi amaiera etiketa, jartzen da eta hurrengo hitzaren aurretik, (*S* etiketa, esaldi hasiera etiketa. Halaber, bi etiketa hauek ipintzen dira puntuazio-marka koma baldin bada eta hurrengo hitza juntagailua bada.

Jarraian ageri den 2. irudian ikus daiteke estatistikan oinarritutako muga gramatika baliatuta (41) esaldiarekin sortzen duen analisia:

(41) familiak duen adierak eta gizartearen funtsezko erakundetzat hartua izateak indarrean diraute, familia horren estrukturak aldatu diren eta elkarbizitza-pleguak aniztu diren arren.

familiak familia IZE ARR - - B-NP B 0 \* (S(S(\*  
duen ukan ADT PNT - - B-VP O 1 \* \*  
adierak adiera IZE ARR - - O O 0 \* \*S)  
eta eta LOT JNT - - O O 0 \* \*  
gizartearen gizarte IZE ARR - - O O 0 \* \*  
funtsezko funtsezko ADJ ARR - - O O 0 \* \*  
erakundetzat erakunde IZE ARR - - B-NP O 0 \* \*  
hartua hartu ADI SIN - - B-VP O 0 \* \*  
izateak izan ADI SIN - - B-VP O 0 \* \*  
indarrean indar IZE ARR - - B-NP O 0 \* \*  
diraute iraun ADT PNT - - B-VP O 0 \* \*  
, , PUNT\_KOMA PUNT\_KOMA - - O O 0 \* \*  
familia familia IZE ARR - - B-NP O 0 \* (S\*  
horren hori DET ERKARR - - I-NP O 0 \* \*  
estrukturak estruktura IZE ARR - - I-NP O 0 \* \*  
aldatu aldatu ADI SIN - - B-VP O 0 \* \*  
diren izan ADL ZHG - - I-VP O 1 \* \*S)S)  
eta eta LOT JNT - - O O 0 \* \*  
elkarbizitza-pleguak elkarbizitza-plegu IZE ARR - - B-NP O 0 \* (S\*  
aniztu aniztu ADI SIN - - B-VP O 0 \* \*  
diren izan ADL MOS - - I-VP O 0 \* \*  
arren arren LOT MEN - KONT O O 0 \* \*  
. . PUNT\_PUNT PUNT\_PUNT - - O E 0 \* \*S)S)

2.irudia: MuGak gramatika estatistikoaren adibidea

Deskribatu ditugun bi gramatika hauek Ixa taldeko analisi katea erabiltzen dute. Hau da, MORPHEUS analizatzaile morfologikoa (Alegria, 1995), HABIL hitz anitzeko unitate lexikoak identifikatzeko tresna (Ezeiza 2003), EIHERA, entitateen izenak mugatzeko eta sailkatzeko tresna

(Alegria et al. 2003), eta EUSTAGGER lematizatzaila (Ezeiza et al. 2003) erabili dira.

Beraz, baliabide hauek guztiz baliagarriak dira sinplifikazioa gauzatzeko, sinplifikatu beharreko perpausak detektatzen eta banatzen laguntzen dutelako. 2. irudiko analisitik, batez ere, ( $S$  eta  $S$ ) etiketak dira kontuan hartu beharrekoak.

## **5.2 Idazlanen Autoebaluaziorako Sistema**

Esku artean dugun azterketarako baliagarria den beste tresna bat Idazlanen Autoebaluaziorako Sistema (IAS) (Castro-Castro et al., 2008). Sistema hau ikasleen idazlanak automatikoki ebaluatzen dituen prototipo bat da. Hala, testu idatzi bat emanda, testu horren zuzenketa egiten du 3. irudiko ezkerreko zutabearen ikus daitezkeen parametroen arabera..

Sistema honek testuak hiru irizpide orokorren arabera sailkatzen ditu: zuzentasuna, konplexutasuna eta jariotasuna. Sinplifikaziorako bereziki azkeneko biak dira interesgarri, uste dugulako testuak zuzenak izango direla, baina laginetan aurkitu ditugun akatsak kontuan izanik, ez dago gaizki, testuen zuzentasuna berrestea.

Gure lanerako baliagarria dela irizten diogu da tresna horrek testuen konplexutasun maila aztertzen duelako. Honela, behin testuak lantzen hasten garenean, testua konplexua bada, testu hori sinplifikatu beharrekoa dela automatikoki jakingo genuke esaldi luzeak izeneko parametroa kontuan hartuta, hau da, esaldi konplexuak badaude, esaldi horiek markatzen ditu 3. irudiko *esaldi luzeak* parametroaren bitartez. Behin esaldi luzeak identifikatuak ditugula horiek sinplifikatuko ditugu, ikerketa-lan honetan adierazitako irizpideak baliatuta. Hortaz, sistema honek, esan bezala, konplexutasunaren arabera testuak sailkatzeko balioko digu eta sinplifikazioaren aurreprozesatzaile bat izango da, PorSimplezen irakurketa-mailaren ebaluazioa (Aluisio et al., 2010) egiten duen atalaren (*readability assessment stage*) antzeko funtzioa duena.

**ixax**

Irudiak Asteazkena, 2011(e)ko uztailaren 27a

Kaixo olatz!!!

Izenburua:

**Kategoria Gramatikalak:**

- Izena: 21
- Izenondoa
- Aditzondoa
- Aditza: 14
- Preposizio
- Lotura Elementua: 7
- Determinatzailea: 15
- Interjekzio
- Izenordaina: 3
- Zenbatzaile

**Erregistrora Itzuli**

**Idatzi**

**Jatorrizko Testua**

**Ebaluatnuk orrira itzuli**

**Hitz kopurua: 101**

**Lema kopurua: 56**

**Esaldi Laburrak: 10**

**Esaldi luzeak: 2**

**Aberastasun lexikala: Ez**

**Erreduantzia testuan: Bai**

**Monotonia testuan: Ez**

**Loturazko Elementuen Konkurrentzia: Ez**

**Akats Ortografikoak: Bai**

**Sinonimoak eta familiakoak: Bai**

**Adierazleak**

Gehlen gustatzen zaidan eguna San Fermnesaren lenengo eguna da. Eguna hobereena parranda egiteko delako. Ni eta nire lagunak batzuk beti bederatzietan gelditzen gara. Jamon arraultzekin jaten dugu, ardoa batzuk edaten dugu eta gero karajilo, kopa bat eta puroa bat. Tira, orain guk prestatuta gaude. Autobusa batean geltokia batera joaten gara. Autobusa bat hartzen dugu baino lehen ostiak batzuk behar du bestela sartzea ezineskoa da. Iruñean udaletxeko enperantza joaten gara. Botilaren xanpaina bat erosten dugu eta enperantza sartzen gara. Txupinazoa bat eta gero peñara joaten gara dantzazera eta abestzera. Arratsalde, seletan gutxi gora behera edozein gauzak jaten behar du. Eta horrela gaueraino.

Eginda

3 irudia: Ikasle batek egindako idazlanaren ebaluazioa IAS sistemaren bitartez

## 6 Ondorioak eta etorkizuneko lana

Ikerketa-lan honetan euskararen tratamendu automatikorako baliagarria izango den testuen sinplifikazioa nola egin aztertu dugu, esaldi konplexuek itzulpen automatikoan arazoak sortzen dituztelako. Horretarako, egitura sintaktiko konplexuen azterketarekin hasi gara, eta lan hau egingarria dela frogatzeko asmoarekin aposizioak, erlatibozko perpausak eta denborazko perpausak aztertzeari ekin diogu.

Hori gauzatzeko beste hizkuntzetan egin diren sistemen, proposamenen eta produktuen deskribapena egin dugu. Horietatik Siddharthanen arkitektura (2002) eta PorSimples proiektuan (2008) egindako lanak eta produktuak dira nabarmendu beharrekoak beste lanetan izan duten eraginagatik eta eskuragarritasunagatik. Eraginari dagokionez, Siddharthanena azpimarratu nahi dugu PorSimplesen arkitekturaren oinarri delako eta bi horiek era berean AI-Basetena. Eskuragarritasunari dagokionez, PorSimples proiektukoena da baliagarriena sortu duten material

## ***HAP Masterra 2010/2011 ikasturtea***

guztia wiki<sup>11</sup> batean jarri dutelako.

Euskarazko testuen sinplifikazio sintaktikoa gauzatzeko *EPEC* eta *Consumer* corpusen laginak aztertu ditugu; zehatzago esanda aposizioak, erlatiboak, erlatiboak eta denborazko egiturak izan ditugu aztergai. Sinplifikazioaren lan hori errazteko Ixa taldeko bi baliabide aurkeztu ditugu: MuGak eta IAS. Halaber, lan hori egiten ari ginela aurkitu ditugun arazo batzuk (ezezko esaldiak, elipsiak, moldatzeko zailak diren esaldiak, erregelen hurrenkera e.a.) ere aipatu ditugu.

Adibideekin batera azaldu ditugun sinplifikazioa proposamenak syntaxira mugatu ditugu, baina ezin dugu alde batera utzi testuen sinplifikazioan beste bi atal daudela: sinplifikazio lexikala eta testu egokitzapena.

Etorkizunari begira, lehenik eta behin bibliografia aztertzen jarraitu beharko genuke; testu sinplifikazioa une honetan oihartzun handiko gaia denez, txosten hau idatzi bitartean lan berriak argitaratu direlako. Euskarari dagokionez, egitura konplexu guztien azterketarekin jarraitu beharko genuke, lan hau hasiera besterik ez delako. Hemen ere alde batera utzi ditugun aditz jokatu gabeek sortzen dituzten egiturak ere kontuan hartu beharko ditugu azterketa egiterakoan, besteak beste.

Corpus-laginetan aurkituriko akatsak ikusirik, informatikaren aldetik sinplifikazioa tratatzen bada, 2. atalean sistema horren oinarri izan daitezkeen aurrekariak azterturik, zuzenketa maila bat gehitzea proposatzen dugu, testu originalak beti zuzenak ez direlako eta sinplifikatu ondoren sortutako testuak zuzenak izan behar direlako.

Bestalde, 5. atalean aztertu diren tresnak gure helburuetara egokituz edo hobetuz gero, hau da, mendeko perpausak errazago identifikatzea, testuak parametro jakin batzuen arabera analizatzea, e.a. lortu ahal izanez gero, sinplifikazio-lana eta zehaztasuna erraztuko genituzke. Euskararen prozesamendurako garatu diren zenbait tresna informatikoren berrerabilpena aurreikusi eta aztertu badugu ere, definituko diren egitura konplexuak detektatzeko bakoitzak zertan eta nola lagunduko duen zehatzago aztertuko dugu euren baliagarritasuna neurtu ahal izateko. Honez gain,

<sup>11</sup> <http://cavelas.icmc.usp.br/wiki/index.php/Principal>

## ***HAP Masterra 2010/2011 ikasturtea***

testu errazen eta testu zailen corpusak sortuz gero, ikasketa automatikoaren bitartez azterketa linguistikoa osatzeko erregelak lortzeko aukera izango genuke.

Gainera, hemen egindako sinplifikazio-proposamenak makinei begira zehaztu eta formalizatu beharko genituzke, sinplifikazioak automatikoki egin ahal izateko. Hala, testu bateko egitura konplexuen markaketa gauzatzeko baliatuko den ezagutza linguistikoan oinarritutako gramatika ere definituko beharko genuke ondoren eskuz markatuta dagoen beste lagin batekin alderatzeko eta bere estaldura eta doitasuna neurtzeko.

Amaitzeko, testu sinplifikazioak dituen onurak eta erabilerak azpimarratu nahi ditugu. Esaldi laburrago eta sinpleagoak erabiliz beste aplikazioen (itzulpen automatikoa, galde-erantzun sistemak) lana erraztu ez ezik, gizakiontzat ere lagungarriak dira. Beraz, egindakoa begiratzuz eta egin beharrekoa aurreikusiz, lana hasi besterik ez dela egin esan dezakegu.

## **Bibliografia**

Afnan A., AI-Subaihin, Rend S., AI-Khalifa. 2011. *AI-Baseet: A proposed Simplification Authoring Tool for the Arabic Language. 2011 International Conference on Communications and Information Technology (ICCIT)*, Aqaba

Alcázar, A. (2005). Towards linguistically searchable text. In *Proceedings of BIDE Summer School of Linguistics*.

Aluísio S.M., Specia L., Gasperin C., Scarton C. 2010. Readability Assessment for Text Simplification. *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–9, Los Angeles, California.

Aristegieta X. 2009. Euskarazko testuen komunikagarritasun-problema larriak. *Senez 37*. Online: <http://www.eizie.org/Argitalpenak/Senez/20091026/aristegieta>

Burstein J. 2009. Opportunities for Natural Language Processing Research in Education. In the *Proceedings of CICLing*, pp. 6-27.

Candido Jr. A., Maziero, E.G., Gasperin, C., Pardo, T.A.S., Specia L., Aluísio S.M. 2009 Supporting the Adaptation of Texts for Poor Literacy Readers: a Text Simplification Editor for Brazilian Portuguese. *Proceedings of the NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications*, pp 34–42, Boulder, Colorado.



## ***HAP Masterra 2010/2011 ikasturtea***

Candido Jr. A., Oliveira M., Aluísio S.M. 2009. Simplifica: um Sistema Web de Autoria de Textos Simplificados. *In the Proceedings of WEBMEDIA 2009 – Simpósio Brasileiro de Sistemas Multimídia e Web*. Fortaleza - CE. v. 2. pp. 55-58.

Carroll J., Minnen G., Pearce D., Canning Y., Devlin S., and Tait J. 1999. Simplifying English text for language impaired readers. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Bergen, Norway.

Caseli H.M., Pereira T.F., Specia L., Pardo T.A.S., Gasperin C., Aluísio S.M. 2009. Building a Brazilian Portuguese parallel corpus of original and simplified texts. In Alexander Gelbukh (ed), *Advances in Computational Linguistics, Research in Computer Science*, vol 41, pp. 59-70.

Chandrasekar R., Srinivas B. 1996. Automatic Induction of Rules for Text Simplification. PhD thesis, Institute for Research in Cognitive Science, University of Pennsylvania.

Chandrasekar R., Srinivas B. 1997 Automatic induction of rules for text simplification. *Knowledge-Based Systems, 10*. pp.183.190.

Chandrasekar R., Doran C., Srinivas, B. 1996. Motivations and Methods for Text Simplification. In *Proceedings of COLING96, poster paper*, Copenhagen, Denmark.

De Belder J., Moens M.F. 2010. *Text Simplification for Children. SIGIR: Workshop on Accessible Search Systems*. Geneva, Switzerland

Erdozia K. 2009. Burmuineko aktibitate elektrikoa hizkuntz gaitasuna neurtzeko [Brain's electric activity measuring the knowledge of language]. *Ekaia: Euskal Herriko Unibertsitateko Zientzia eta Teknologia Aldizkaria*.

Gasperin C., Specia L., Pereira T., Aluísio S. 2009 Learning When to Simplify Sentences for Natural Text Simplification. In: *CSBC 2009 - XXIX Congresso da Sociedade Brasileira de Computação, Bento Gonçalves*. pp.809-818.

Inui K., Fujita A., Takahashi T., Iid, R., Iwakura T. 2003. Text Simplification for Reading Assistance: A Project Note. In the Proceedings of the Second International Workshop on Paraphrasing. pp. 9 -16.

Jing H. 2000. *Sentence Reduction for Automatic Text Summarization. ANLC '00 Proceedings of the sixth conference on Applied natural language processing*.

Karlsson F., Voutilainen A., Heikkilä J., Anttila A. 1995. *Constraint Grammar: A Language independent System for Parsing Unrestricted Text*. Mouton de Gruyter. Berlin.

Klebanov B., Knight K., Marcu D. 2004. Text Simplification for Information-Seeking Applications. *On the Move to Meaningful Internet Systems*. LNCS, V. 3290, pp. 735-747.

Leturia I., Martinez E. 2010. Consumer Corpora. Online: [http://www.euskaltzaindia.net/dok/plazaberri/2010/otsaila/igor\\_edurne.pdf](http://www.euskaltzaindia.net/dok/plazaberri/2010/otsaila/igor_edurne.pdf)

## ***HAP Masterra 2010/2011 ikasturtea***

- Liben-Nowell, D. 2000. *Syntactic Simplification*. MPhil thesis. University of Cambridge.
- Max A. 2000. *Syntactic Simplification. An application to text for aphasic readers*. MPhil in Computer Speech and Language Processing. University of Cambridge.
- Max A. 2005. Simplification interactive pour la production de textes adaptés aux personnes souffrant de troubles de la compréhension. In the *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*.
- Petersen S.E., Ostendorf M. 2007. Text Simplification for Language Learners: A Corpus Analysis. In the *Proceedings of the Speech and Language Technology for Education Workshop*, pp. 69-72.
- Ramírez-Sánchez G., Sánchez-Martínez F., Ortiz-Rojas S., Pérez-Ortiz J.A. Forcada M.L. 2006. Opentrad Apertium open-source machine translation system: an opportunity for bussines and research. *Proceedings of the Twenty-Eighth International Conference on Translating and the Computer*. London.
- Scarton C., Oliveira M., Candido A. Jr., Gasperin C, and Aluísio S.M. 2010a. SIMPLIFICA: a tool for authoring simplified texts in Brazilian Portuguese guided by readability assessments. *Proceedings of the NAACL HLT 2010: Demonstration Session*, pp. 41–44, Los Angeles, California.
- Scarton C., Oliveira M., Candido A. Jr., Gasperin C, and Aluísio S.M. 2010b SIMPLIFICA: an authoring system targeting simplified texts in Brazilian Portuguese. In the *Proceedings of PROPOR 2010, 9th International Conference on Computational Processing of the Portuguese Language, Extended Activities Proceedings*, 1 CD-ROM v1.
- Siddharthan A. 2002. An Architecture for a Text Simplification System. In the *Proceedings of the Language Engineering Conference*, pp. 64-71.
- Siddharthan A. 2003. *Syntactic Simplification and Text Cohesion*. PhD Thesis. University of Cambridge.
- Siddharthan A., Nenkova A., McKeown, K. 2004. Syntactic simplification for improving content selection in multi-document summarization. *COLING '04 Proceedings of the 20th international conference on Computational Linguistics*.
- Specia L., Aluisio S.M., Pardo T.A.S. 2008. Manual de Simplificação Sintática para o Português. Technical Report NILC-TR-08-06, 27 p.
- Tapanainen P. 1996. The Constraint Grammar Parser CG-2. *Publications of the Department of General Linguistics*, University of Helsinki, 1–48. Yliopistopaino. Helsinki.
- Watanabe W.M., Fortes R. P. M., Pardo T. A. S., Aluisio S. M. 2009. Facilita: auxílio à leitura de textos disponíveis na Web. In the *Proceedings of WEBMEDIA 2009 – Simpósio Brasileiro de Sistemas Multimídia e Web*. Fortaleza - CE. v. 2. pp. 27 - 30.

## **HAP Masterra 2010/2011 ikasturtea**

Zabala I., Odriozola J.C. 1992. *Hitz-ordena, galdegaia eta komaren erabilera. Idazkera tekniko 1*. Euskal Herriko Unibertistatea. Bilbao.

## **Gramatikak, estilo-liburuak eta hizkuntza errazteko gomendioak**

Arakama J.M., Arrieta A., Lozano J., Robles J., Urrutia R.M. 2009. *IVAPeko Estilo-liburua*. Herri Ardularitzaren Euskal Erakundea.

Berria. 2006. *Estilo Liburua*. Berria. Andoain.

Durangoko berritzegunea. 2003. *Idatzizko testuak idazten*. Online: [http://www.aulaintercultural.org/IMG/pdf/IDATZIZKO\\_TESTUAK\\_ERRAZTEN.pdf](http://www.aulaintercultural.org/IMG/pdf/IDATZIZKO_TESTUAK_ERRAZTEN.pdf)

Egunkaria. 2001. *Estilo Liburua*. Euskaldunon Egunkaria. Andoain.

Euskaltzaindia. 109. Araua. Komunztadura Aposizioetan. Online: [http://www.euskaltzaindia.net/dok/arauak/Araua\\_0109.pdf](http://www.euskaltzaindia.net/dok/arauak/Araua_0109.pdf)

Euskaltzaindia. Gramatika Batzordea. 1999. *Euskal Gramatika. Lehen Urratsak V (Mendeko Perpausak 1)*. Euskaltzaindia.

Euskaltzaindia. Gramatika Batzordea. 2005. *Euskal Gramatika. Lehen Urratsak VI (Mendeko Perpausak 2)*. Euskaltzaindia.

Plain Language.org. 2011. Federal Plain Language Guidelines. Online: <http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/FederalPLGuidelines.pdf>

Zubimendi J. M., Esnal P. 1993. *Idazkera-Liburua*. Eusko Jaurlaritzaren Argitalpen Zerbitzu Ofiziala. Gasteiz.

## **Ixa taldeko tresnak eta baliabideak**

Aduriz I., Aranzabe M., Arriola J., Atutxa A., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Oronoz M., Soroa A., Urizar R. 2003. Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing. *Proceedings of the Corpus Linguistics 2003 conference, 16 (1)*. Ed. Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery. Pp 10-11. Lancaster University (UK)

Aduriz I., Arrieta B., Arriola J., Díaz de Ilarraza A., Izagirre E., Ondarra, A. 2006. Muga Gramatikaren Optimizazioa. *UPV/EHU/LSI/TR 9-2006*

Alegria I. 1995. *Euskal morfologiaren tratamendu automatikorako tresnak*. Doktoretza-tesia.

## ***HAP Masterra 2010/2011 ikasturtea***

Informatika Fakultatea. UPV-EHU. Donostia.

Aldezabal I., Aranzabe M., Arriola J., Díaz de Ilarraza A. 2009. Syntactic annotation in the Reference Corpus for the Processing of Basque (EPEC): Theoretical and practical issues. *Corpus Linguistics and Linguistic Theory* 5-2 ,pp. 241-269. Mouton de Gruyter. Berlin-New York.

Alegria I., Ezeiza N., Fernandez I., Urizar R. 2003. Named Entity Recognition and Classification for texts in Basque. *II Jornadas de Tratamiento y Recuperación de Información, JOTRI, Madrid*.

Aranzabe, M. J. 2008. *Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala*. Doktoretza-tesia. Euskal Filologia Saila. Donostia.

Arrieta B. 2010. *Azaleko sintaxiaren tratamendua ikasketa automatikoko tekniken bidez: euskarako kateen eta perpausen identifikazioa eta bere erabilera koma-zuzentzaile batean*. Doktoretza-tesia. Informatika Fakultatea (EHU). Donostia.

Castro-Castro D., Lannes-Losada R., Maritxalar M., Niebla I., Pérez-Marqués C., Alamo-Suarez N.C., Pons-Porrata, A. 2008. A Multilingual Application for Automated Essay Scoring Lecture Notes in Advances in Artificial Intelligence - *LNAI 5290 - IBERAMIA ISBN 3-540-99308-8 Springer New York pp. 243-25*

Ezeiza N., Alegria I., Aranzabe M., Ezeiza A., Urizar R. 2003 Robustez y flexibilidad de un lematizador/etiquetador *Actas-I del VIII Simposio Internacional de Comunicación Social. pp. 528-532, 2*

Ezeiza N., Lopez de Ipiña K., Graña M., Hernández M., Zulueta E., Ezeiza A., Tovar C. 2003. Selection of Lexical Units for Continuous Speech Recognition of Basque. *LNCS: Progress in Pattern Recognition, Speech and Image Analysis Publisher: Springer Berlin/Heidelberg pp. 245-250*

Labaka G. 2010. *EUSMT: Incorporating Linguistic Information into SMT for a Morphologically Rich Language. Its use in SMT-RBMT-EBMT hybridation*. Doktoretza-tesia. Lengoaia eta Sistema Informatikoak Saila (UPV-EHU). Donostia.

Mayor A., Alegria, I., Díaz de Ilarraza A., Labaka G., Lersundi M., Sarasola K. 2011. Matxin, an open-source rule-based machine translation system for Basque. *Machine Translation Journal*.

Pociello E., Agirre E., Aldezabal I. 2010. Methodology and construction of the Basque WordNet. In: *Language Resources and Evaluation. Springer. Netherlands*.