

Morfologia konputazionala

Euskararen morfologiaren deskribapena

Iñaki Alegria

Miriam Urkia

Udako Euskal Unibertsitatea
Bilbo, 2002

Babesleak



Kolaboratzaileak



© Iñaki Alegria, Miriam Urkia

© Udako Euskal Unibertsitatea

ISBN: 84-8438-034-3

Lege-gordailua: BI-2679-02

Inprimategia: RGM Bilbo

Azalaren diseinua: Iñigo Ordoizgoiti

Banatzzaileak: UEU. Erribera 14, 1. D BILBO / 94-6790546
argitalpenak@ueu.org / www.ueu.org

Zabaltzen. Igerabide, 88 DONOSTIA

Hitzaurrea

Inprentaren sorkuntzak hizkuntzaren tratamendua eta zabalkuntza irauli zituen XV. mendean, baina konputagailuak ez du iraultza txikiagoa ekarri XX. mendearen bukaeran eta XXI.aren hasieran.

Hasteko, gero eta gehiago erabiltzen ditugu konputagailuak eta konputagailu-programak geure eguneroko jardunean, eta programa horietako askok testua nola edo hala tratatu edo prozesatu egiten dute.

Bestalde, konputagailuekiko komunikazioa hizkuntza arruntaren bitartez, eta ez lengoia formal baten bidez, egitea gero eta normalago izango da. Gizarte eleanitzak hizkuntza batetik bestera egin behar izaten dituen joan-etorriak leuntzeko ere, aparteko lagun dugu konputagailua.

Gainera, telekomunikazioetan gertatutako aurrerapen izugarriak eragin duen Internet fenomenoak, areagotu egin du hizkuntzaren tratamendu automatikoaren beharra; interesatzen zaigun informazioa ondo selekzionatzeko, esaterako, tratamendu linguistiko lagungarria ezinbestekoa baita.

Hizkuntzaren industria oso bat sortzen ari da, konputagailuaz baliatuz hizkuntza prozesatzea helburu duena. Jakintza arlo berri honetan ezinbestekoa da informatikari eta hizkuntzalarien arteko lankidetzak.

Testuinguru honetan, ezinbestekoak dira teknologia hauek ezagutuko dituzten informatikari eta hizkuntzalariak. Udako Euskal Unibertsitateak 2001. urtean ikusi zuen une egokia zela, unibertsitate desberdinetako irakasleak bilduz, hurrengo ikasturteetan heziketa-behar hori bete ahal izateko. Horrela sortu zen UEUren Hizkuntza-Teknologiak (*Hiztek*) graduondokoa, geroago arrakasta handia lortu duena (<http://www.ueu.org/titulazioak/Hiztek.htm>).

Hainbat unibertsitateko irakasleak bilduz, programa bat antolatzeke masa kritikoa lortu daitekeela frogatu dugu. UEU saiatu da beste unibertsitateen parte hartzea zabaltzen, tituluen kalitatea eta etorkizuna hobetzeko asmoarekin. Helburu horrekin talde hauek bildu ditugu Hizkuntza-Teknologiak masterraren inguruan: EHUKo Ixa Taldea, EHUKo Aholab taldea, EHUKo Euskal Filologia saileko irakasleak, Deustuko Unibertsitatea, Mondragon Unibertsitatea eta Baionako Iker Taldea.

Titulazio honetako irakasleek ahalegin handia egin dute material didaktikoaren sorkuntzan. Modulu bakoitzaren hasieran, modulu horri dagokion material idatzia jasotzen dute ikasleek. Material hori ikusita, plan bat egin genuen material hori epe ertainean argitaratzeko. Eta liburu hau da plan horren lehenengo fruitua.

Morfologia izan da, linguistikaren barruan, automatizatzeko garaian arrakastatsuen gertatu den arloetako bat. Horretan oinarrituta, hainbat aplikazio erabiltzen dugu egunero: testu-zuzentzaileak, leman oinarritutako Interneteko bilatzaileak, datu-base dokumentalen atzipen adimenduna, etab.

Tratamendu morfologikoa sintaxiarekin edo semantikarekin alderatuta sinplea da, baina, hala eta guztiz ere, nahiko konplexutasun du automatizatzeko garaian adimen artifizialaren barruan kokatu ahal izateko.

Liburu honetan hizkuntzaren morfologia deskribatzeko landu diren formalismoen aurkezpena egiten da batetik; eta euskararen deskribapen morfologiko osoa bestetik.

Egileak, Iñaki Alegria eta Miriam Urkia doktoreak, aitzindariak izan dira euskararen morfologia konputazionalen, ikuspuntu informatikotik lehena eta linguistikotik bigarrena. Eta biak dira geroago sortu diren hainbat tresna informatikoren *gurasoak*, *Xuxen* zuzentzaile ortografikoa, *Morfeus* analizatzaile morfologikoa eta *Euslem* lematizatzailea. Azken tresna horiek UZEIk landu duen "XX. mendeko euskararen corpus estatistikoa"ren (www.euskaracorpora.net/XXmendea) azken urteak lematizatzeko erabili dira, baita zientzia.net (www.zientzia.net) eta Egunkariako hemerrotekan (www.egunkaria.com/hemeroteka) guneeetan bilaketak egiteko ere. Tresna horien oinarri teknikoak azaltzen dira liburu honetan.

Kepa Sarasola (*Hiztek* programaren arduraduna)

Aurkibidea

1	Sarrera.....	9
1.1	Motibazioa	9
1.2	Helburuak	9
1.3	Egitura.....	10
1.4	Ariketak eta zalantzak.....	10
2	Morfologia konputazionala.....	11
2.1	Sarrera.....	11
2.2	Oinarrizko kontzeptuak	12
2.3	Aplikazioak.....	14
2.4	Morfologiarako eredu konputazionalen ezaugarriak	15
2.5	Garapen historikoa eta zenbait sistema.....	18
2.6	Sailkapena	24
2.7	Laburpena	25
3	Bi mailatako morfologia	27
3.1	Sarrera.....	27
3.2	Lexikoa	27
3.3	Erregela fonologikoak	31
3.4	Hobekuntza: lexiko-itzultzaileak	41
3.5	Sistema erabilgarriak	45
3.6	Laburpena	47
4	LinguistX eta adibideak	49
4.1	Sarrera.....	49
4.2	<i>twolc</i> programa	49
4.3	<i>lexc</i> programa.....	59
4.4	Probak	64
4.5	Adibide oso bat	65
4.6	Metodologia.....	68
4.7	Laburpena	70
5	Euskararen morfologiaren deskribapena	71
5.1	Sarrera.....	71
5.2	Sistema lexikoa	73
5.3	Erregelak.....	100
5.4	Emaitzak	109
5.5	Hobekuntza-proposamena: informazio egituratuaren beharra.....	116
6	Estaldura osorako urratsak.....	129
6.1	Sarrera.....	129
6.2	Aldaeren tratamendua	131
6.3	Lexikorik gabeko analisisa	134
7	Ariketak	137
8	Glosategia	139
9	Bibliografia	141
9.1	Oinarrizko bibliografia	141
9.2	Bestelako bibliografia.....	141
10	Aurkibide alfabetikoa	147

1 Sarrera

1.1 Motibazioa

Ingelesaren flexio-morfologia sinplea dela-eta, ordenadorez egindako analisi/sintesi morfologikoari ez zitzaion kasu handiegirik egiten. Programa eta ezagumendu linguistikoa nahasten zuten sistema primitiboak ziren ohikoak duela urte gutxi arte. Azken urteetan, aldiz, arlo honetan egindako lanak ugaritu egin dira hainbat arrazoi direla medio: beste hizkuntzetarako sistema automatikoen garapena, corpusetan oinarritutako analisi linguistikorako morfologiak eskaintzen duen abantaila, Interneteko informazioa indexatzeko eta hobeto bilatzeko eskaintzen duten aukera, ...

Euskara bezalako flexio handiko hizkuntzetan, tratamendu morfologikoa eta lematizazioa funtsezkoak dira tresna linguistikoak egin nahi direnean. Horrela, hainbat hizkuntzatan egiaztapen ortografikoa hitz-zerrendetan oinarrituta egiten den arren, euskaraz modu horretan egitea (ia) ezinezkoa da, lema batetik sor daitekeen forma kopurua izugarri handia baita.

Adibidez, IXA taldean garatu dugun analizatzaile morfologikoa, *Morfeus* izenekoa, funtsezko tresna izan da hainbat aplikazio komertzial garatu ahal izateko: Xuxen zuzentzaile ortografikoa, Elhuyar hiztegia *on-line* kontsultatzeko sistema (kontsultatu aurretik lematizatzea ezinbestekoa da), *Galn* Internet/Intranet bilatzailea eta UZEIren *Sinonimoen Hiztegia*, *MS-Word*-en integratua, besteak beste.

1.2 Helburuak

Liburu honen helburuak hauek dira:

- Morfologiaren oinarrizko kontzeptuak eta fenomenoak azpimarratzea, ikuspuntu konputazionalari dagokionez.
- Morfologian oinarritutako aplikazio informatikoak azaltzea.
- Morfologia konputazionalki tratatzeko formalismoen aurkezpena egitea.
- Bi mailatako morfologiaren funtsa ezagutzea eta praktikatzea.
- Euskararen deskribapen morfologikoa aurkeztea, bi mailatako morfologiaren ereduaren arabera.

1.3 Egitura

Bi atal nagusi daude liburuan: hizkuntzaren morfologia deskribatzeko, eta sistema automatiko batean integratzeko, landu diren formalismoen aurkezpena batetik; eta euskararen deskribapen morfologiko osoa zehazten duen atala bestetik.

Euskararen deskribapen morfologikoa bosgarren kapituluaren egiten da. Lehen atal nagusiak, berriz, hiru kapitulu hartzen ditu: formalismo desberdinen azterketa egiten duen bigarren kapituluak, bi mailatako morfologia deskribatzen duen hirugarrena eta *LinguistX* izeneko tresnaren zehaztasunak azaltzen dituen laugarren kapituluak. Aipatutako tresna, zoritxarrez, ez da software librea, baina bibliografian agertzen den *Finite State Morphology: Xerox Tools and Techniques* (Beesley & Karttunen, 2002) izeneko liburuarekin batera banatzen da.

Bi atal nagusi horiek osatzeko deskribapen morfologikoen estaldura handitzeko hainbat proposamen azaltzen dira seigarren kapituluaren.

Ariketa batzuk proposatu ondoren, glosategiak, bibliografiak eta aurkibide alfabetikoak osatzen dute liburua.

1.4 Ariketak eta zalantzak

Alderdi teorikoa eta praktikoa konbinatzen saiatuko gara liburuan. Azaldutakoa finkatzeko eta gogoratzeko laguntza izan nahi dute bukaerako ariketek, glosategiak eta aurkibide alfabetikoak.

Liburuaren inguruko zalantzak, proposamenak edota oharra bidera daitezke i.alegria@euskalerrria.org helbidearen bitartez.

2 Morfologia konputazionala

2.1 Sarrera

Morfologiarako eredu konputazional desberdinen aurkezpena egitea eta horien barruan bi mailatako morfologiarena kokatzea eta sakontzea da kapitulu honen xede nagusia. Hala ere, formalismo morfologikoak aztertu baino lehen, hauek ezagutzen lagunduko diguten zenbait funtsezko ezaugarri aurkezten dira hasteko. Ezaugarri horietan oinarrituz sailkapen bat proposatzen da, klasifikazio honen barruan literaturan agertzen diren zenbait sistema kokatuz. Sistema hauetako batzuen nondik norakoak ere azaltzen dira.

Dena den, hasi aurretik ikus dezagun zein den gure helburua: hizkuntza baterako analizatzaile morfologiko bat eraikitzea. Zer egin beharko luke tresna edo programa horrek? Minimoa, morfemen banaketa litzateke.

2.1 adibidea: banaketa morfologikoa

“*handien*” sakatzen baldin badugu, emaitza hau lortzea:

handi+en

Aurreko horri banaketa morfologikoa esaten zaio, baina batzuetan ez da nahiko. Hori baino gehiago eska daiteke, morfema bakoitzari dagokion informazioa jartzea ez baitirudi hain zaila denik.

2.2 adibidea: analisi morfologikoa

“*handien*” sakatuta hau litzateke emaitza oraingoan:

handi [ADJ] + [GRA] [SUP]

handi [ADJ] + [GEN] [MP]

horrela irakurriko genuke: *handienaren lema handi da, baina bi analisi posible ditu: graduatzaile superlatiboa edo genitibo mugatu plurala.*

Programari, analizatzaile deitu beharrean, prozesadore morfologikoa deitzen zaio batzuetan, analisisa egiteaz gain sorkuntza egiteko ere gai delako.

2.3 adibidea: sorkuntza morfologikoa

Sarrera hau emanda:

handi [ADJ] + [GRA] [SUP]

Emaitzak *handien* izan beharko luke.

Dena den, sistema batzuetan, *Unix/Linuxeko ispell* programa hedatuan esaterako, *stemming* izeneko prozesu sasilinguistikoa burutzen da, bere bitartez, hitz bat eman eta

dagokion lema edo erroa lortuz. Sasilinguistikoa esatean hau da adierazi nahi dena: lema lortzeko prozesua ez ohi da oso zehatza, eta erroreak edo desbiderapenak onartzen dira, konputazionalki oso metodo sinpleak erabiltzen baitira: lema eta atzizkien kateaketa zein salbuespenen zerrenda bat gutxi gorabehera.

Zenbait autorek *stemming* eta lematizazioa nahasten badute ere, lematizazioa harantzago doan prozesua da, desanbiguazioa ezinbesteko urratsa du-eta.

Morfologia konputazionala fonologia konputazionalarekin oso lotua dago, eta aztertuko diren formalismo asko bietarako sortu dira. Horrela, fonologiako analizatzaileak fonemetatik lexikorako bihurketa egingo luke, eta sorkuntzak morfema-katetik fonemetara. Tartean egin liteke sistema bat hitzen eta fonemen arteko jauzia emateko. Hala eta guztiz ere, adibideetan testu-hitzetan agertzen diren fenomenoekin arituko gara batez ere, ikuspuntu morfologikoa areagotuz.

2.2 Oinarrizko kontzeptuak

Morfologia teorikoan sakontzeko batere asmorik gabe, ondoren azalduko diren eredu konputazionalak alderatu eta sailkatu ahal izateko, beharrezkoa da morfologiaren kontzeptu orokorrak gainbegiratzea.

Gai honetarako funtsezkoa den liburuan (Sproat, 1992:17) egiten diren galderak izan daitezke kontzeptu horiek azaltzeko iturburu:

“In particular, I shall discuss the following issues:

What sort of things can morphology mark in different languages?

How are words built up from smaller meaningful units-morphemes? ...

What are the constraints on the order of morphemes within words?

Do phonological rules complicate the problem of morphological analysis?”

Galdera hauei erantzutean, konputazio-ereduei lotutako morfologiaren kontzeptu garrantzitsuenak ondorioztatzen dira:

- Funtzioei begira, hiru kontzeptu¹ azaltzen dira nagusiki: *flexio-morfologia*, *eratorpen-morfologia* eta *elkarketa*. Flexioa sintaxiak eskatua da, erregularra da kategoriaren arabera, eta ez du funtzio sintaktikoa aldatzen. Eratorpena, aldiz, ez da, oro har, sintaxiak eragindakoa, ez da erregularra eta kategoria gramatikaren aldaketa ekarri ohi du. Elkarketa lema bat baino gehiago biltzean sortzen da eta,

maiz testu-hitzaren muga gainditzen duenez, bere tratamendua aurreko biena baino korapilatsuagoa da. Eratorpena eta elkarketa askotan saihesten dira morfologia konputazionalan, duten erregularitasun-faltagatik ez baita erraza erabakitzea konbinatu beharreko lemak/morfemak zeintzuk diren².

- Osagaien arteko loturak bideratzeko murriztapenak, askotan *morfotaktika* eta *hitzaren gramatika* deritzana. Inguruko morfemen funtzioa izatea da arruntena, gehienetan *lemei aurrizkiak* eta *atzizkiak* eransten zaizkielako, nahiz eta hizkuntzen arabera erregularitasun- eta hurbiltasun-gradu oso desberdinak aurkitu, baita *artizkiak* ere.

Morfemen *kateatze sinplea* arruntena bada ere, badaude formalismo ahaltsuagoak eskatzen dituzten fenomenoak, *erro-patroietan* oinarritutako fenomenoak, adibidez, hizkuntza semitikoetan gertatzen direnak. Konplexutasunaren aldetik tartean egongo liratekeen beste fenomenoak ere aipa daitezke: artizkien bidezko lotura, morfema-bikoizketa, etab.

Inguruko hizkuntzetan ere, kateatze hutsa ez-osoa dela azaltzeko *urruneko mendekotasuna* izeneko fenomeno dugu, morfema ez-jarraituen arteko baldintzak gerta baitaitezke. Adibidez, ingelesez *en*, *joy* eta *able* morfemak elkarren segidan joan badaitezke ere, *joy* eta *able* bakarrik ezin dira lotu (**joyable* ezin da erabili ingelesez). Alegia, *en*-ek *able*-en erabilera baldintzatzen du, edo, beste era batera esanda, *en* eta *able*-ren artean urruneko mendekotasuna dago. Euskaraz ere antzera gertatzen da *bait-* aurrizkia, aditz jokatua eta *-lako* atzizkiaren artean. Aurrizkiaren atzean aditz jokatua joan daiteke (*baikara*), eta aditz jokatuaren atzetik atzizkia (*garelako*), baina inola ere ez hirurak batera (**baikarelako*).

- *Aldaketa fonologikoak*: morfemak biltzean gertatzen diren aldaketak dira. Fonologiak zuzenean eragindakoak izan daitezke batzuetan (suomieraz, adibidez), ortografiak berak ere eragin ditzake (ingelesez, esaterako), eta morfologikoak ere aurki ditzakegu (euskaraz, zenbait kasutan); horrexegatik, *morfofonologikoak* ere deituko dira txostenean, bibliografian aurreko izenez gain *morfografemika* ere agertu arren. Aldaketa hauen kopurua eta horiek gertatzeko

¹ Morfologia teorikoan banaketa bitarikoa bada ere: flexio-morfologia vs. morfologia lexikoa. Hain zuzen, azken honetan biltzen dira eratorpena eta hitz-elkarketa.

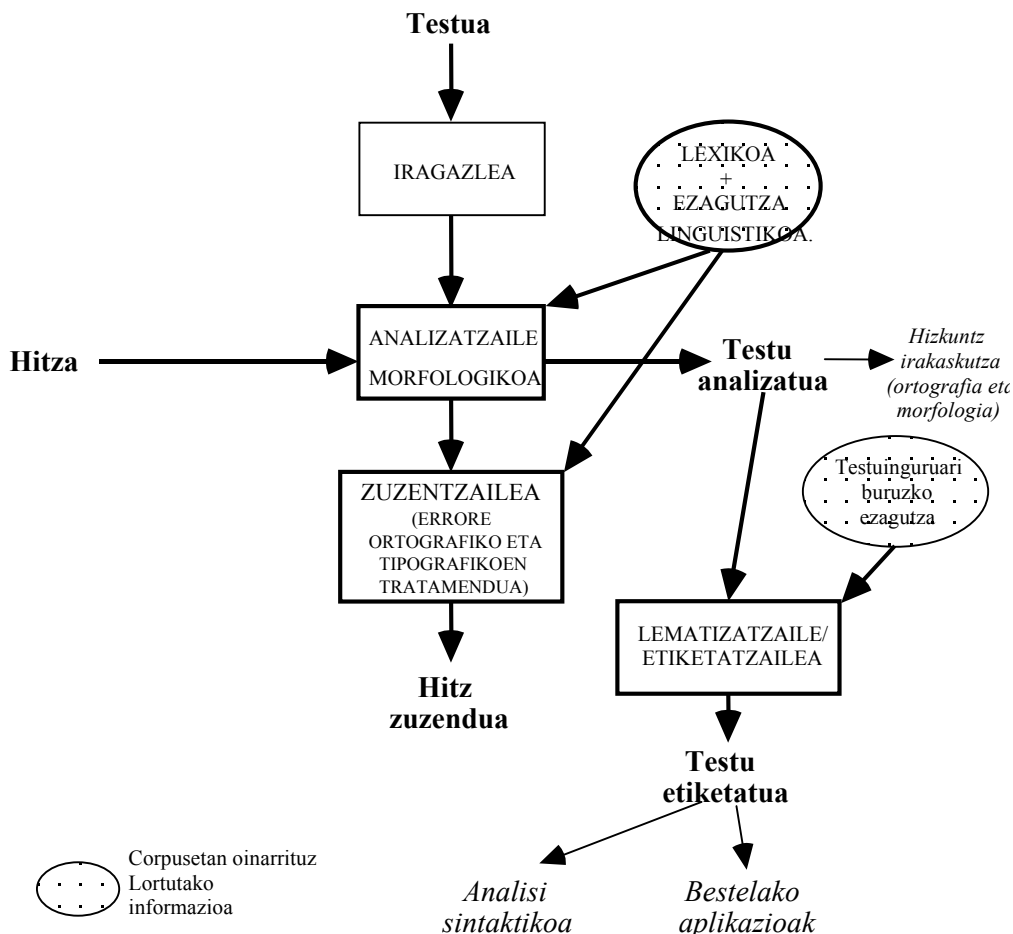
² Hizkuntzalaritza teorikoan ere, garai batean morfologia lexikoa lexikoaren ataltzat hartua izan zen.

baldintzak arras desberdinak dira hizkuntza desberdinen arabera. 3.3.8 atalean hizkuntza desberdinetan agertzen diren hainbat fenomeno azaltzen dira.

Fenomeno honen aurrean, analisi morfologikoa egiterakoan, sistema batzuetan *alomorfoak* erabiltzen dira, hau da, morfema bera adierazteko forma bat baino gehiago zehazten dira lexikoan. Aldaketa fonologiko konplexuaren adibide gisa, hizkuntza batzuetan gertatzen den bokal-harmoniaren fenomenoak dugu: puntu batean gertatzen den bokal baten aldaketak ondoko edo aurreko bokal guztien aldaketa ere eragin dezake. Kasu gehienetan, baina, aldaketa horiek lokalak dira, morfema-mugan gertatzen dira-eta.

2.3 Aplikazioak

Hizkuntza-teknologiako sistemetan, morfologia, sintaxia eta semantika dira garrantzi handieneko lanbideak; zuzenean aplikazio komertzializagarriak ez izan arren, tresna komertzial gehienek funtsa baitira. Morfologia da lehen urratsa, ondoren sintaxia, eta semantika azkena.



1. irudia.- Analizatzaile morfologiko baten zenbait erabilera

Morfologiaren eta sintaxiaren artean *tagging* edo *etiketatz*e izeneko prozesua burutu ohi da, forma bakoitzari dagokion informazio guztia esleituz; ondoren, testuinguruaren informazioan oinarrituta, hitz bakoitzari analisi egoki bakarra utziko zaio. Analisisien artean bat hautatzeko prozedurari *desanbiguazio* morfologikoa esaten zaio.

Morfologian oinarritutako produktuak aipatzean, hauek azpimarra daitezke:

- zuzentzaile ortografikoak: nahiz eta merkatuko gehienak hitz-zerrendetan oinarritu, flexio handiko hizkuntzetan irtenbide hau ez da bideragarria. Halakoetan, hitz bat onartzeko ala baztertzeko, hitzaren analisi morfologikoa buru daiteke, eta analisirik ez duenean desegokitza jo. Gain-analisisa (ikus ondoko atala) ekiditea ezinbesteko da aplikazio honetarako.
- datu-base dokumentaletan edo Internet/Intranet bilatzaileetan erabiltzen diren lematizatzaileak. Bilaketetan lema erabiltzeak sistema erosoago eta ahaltsuago egiten du.
- hiztegiak eta corpusak kontsultatzeko ere analisi zein sorkuntza morfologikoa oso praktikoa da: hiztegian lema baino ez daudenez, bilatu aurretik analisi morfologikoa behar baita. Hiztegia kontsultatu ondoren sorkuntza morfologikoa egiten bada, ordainarekin hasierako formaren baliokidea lortuko da (sinonimoen hiztegia, esaterako).
- bestelako aplikazioak. Morfologia ez da nahiko izango beste aplikazio batzuetarako, baina bai ezinbesteko: hizketaren analisisa eta sintesia, lengoia naturalaren bidezko interfazeak, testuen sorrera eta ulermena, ordenagailu bidezko hizkuntza-irakaskuntza, itzultzaile automatikoak edo erdiautomatikoak, etab.

1. irudian zenbait aplikazioen eskema ikus daiteke.

2.4 Morfologiarako eredu konputazionalen ezaugarriak

Ingelesaren flexio-morfologia sinplea³ dela-eta, tradizioan ez zaio kasu handirik egin ordenagailuz landutako analisi/sintesi morfologikoari (Winograd, 1983). Programa eta ezagutza linguistikoa nahasten zituzten sistema primitiboak ohikoak ziren duela gutxi

arte. Azken urteetan, aldiz, arlo honetan egindako lanak ugaritu egin dira, beste hizkuntzetarako sistema automatikoen garapenak indarra hartu baitu eta, horrez gain, corpusetan oinarritutako analisiak erraztasun handiak ekarri baitituzte.

Gaur egun prozesadore⁴ morfologiko ugari aurki daiteke bibliografian, zein bere ikuspuntu eta ezaugarriekin. Haien arteko konparaketa egin ahal izateko irizpide batzuk zehaztu behar dira aurretik; horretarako, aurreko atalean azaldutako galderetatik abiatuko gara, ondorio hauetara iristeko:

- Formalismo edo ereduaren *deskribapen-ahalmena*, hau da, zein fenomeno adieraz edo analiza daiteke eredu hori erabiliz. Flexioa, eratorpena zein hitz-elkarketa adierazteko gai izatea da egokiena. Era berean, irizpide honen barruan, *analisi* eta *sorkuntza* burutzeko gaitasuna edo bietako bat bakarrik burutzekoa bereiziko dugu.
- Morfologiari ekiteko modua. Teoria linguistikoek eraginda, bai eta hizkuntzaren egiturak zein sistema eraikitze konputazio-ikuspegiak ere, bi eredu bereizten dira: lexikoan oinarrituak batetik, eta paradigmak oinarrituak bestetik. Lehena da ohikoena eta aztertuko ditugun ereduak ildo horretatik doaz; hala ere, zenbait lanetan (Calder, 1989; Anick & Artemieff, 1992) hiztegia bigarren mailan geratzen da, eta paradigma da eraketaren funtsa. Azkenik *stemmer*-ak aipatu behar dira, tresna hauetan atzizkiak baino ez baitira lantzen eta prozesuaren funtsa hitzen bukaeran atzizkiak bilatzea baita.
- Morfotaktika ebazteko modua. Aurretik aipatu den bezala, morfemen arteko lotura posibleak zehazteko moduarekin dago lotua. Morfemetan oinarritutako sistemetan bi prozesaketa-mota agertzen dira nagusiki: *egoera finituetako morfotaktika* deituko duguna eta *baterakuntza-mekanismoetan* oinarritutakoak. Lehenengoetan morfemen arteko erlazioak grafo-eran ikus daitezke, nodoak morfemak eta arkuak onartutako kateatzeak direla. Baterakuntza-mekanismoek syntaxian erabili ohi diren ezaugarrietan oinarritutako gramatikak aintzakotzat hartzen dituzte, eta, ondorioz, malguagoak dira, tratamendu morfologikoa —edo morfosintaktikoa— errazten dute, baina konplexuagoak dira konputazioaren

³ Ingelesaren morfologia sinpletzat hartu izan bada ere, hau guztiz zalantzazkoa da. Sproat-ek (1992:152-53) azpimarratzen du, morfologia konplexua duten hizkuntzetan (suomiera edo turkiera, esaterako) konplexutasuna luzeraren sinonimotzat hartu izan dela, erregularitasuna eta aldaketen kasuistika kontuan hartu gabe.

⁴ Analisi edo/eta sintesi morfologikoa burutzen duen programari prozesadore morfologiko deituko diogu.

ikuspuntutik, eta, beraz, motelagoak abiaduraren aldetik. Azken hauek, eredu paradigmaticotik egindako hurbilpenak dira askotan, objektuei zuzendutako ereduetan ohiko diren herentzia bezalako kontzeptuak erabiltzen dituztenak (de Smedt, 1984; Calder, 1989; Zajac, 1997).

- Aldaketa morfofonologikoak adierazteko modua. Bestelakoak badaude ere, bibliografian bi metodo gailentzen dira: duela urte batzuk ohikoak ziren programa bidezko metodo *ad-hoc*ak eta gaur egun arrakastatsu bihurtu den bi mailatako formalismoa.
- Lexikoan gordetzen diren osagai-motak. Ohikoa da morfemak gordetzea, sistema batzuetan erroak gordetzen ez badira ere; baina batzuetan hitz-zatiak gordetzen dira, aldaketa morfofonologikoak adierazteko modurik ez dagoelako edo eraginkortasunari eusteko. Adibidez, gaztelaniaz *hacer* aditzerako *hac*, *hag*, *hic*, *hiz* eta *hech* alomorfoak diren hitz-zatiak erabiltzen dira sistema batzuetan. Beste aukerak ere badira, silabekin lan egiten dutena, adibidez (Cahill, 1990). Irizpide honen barruan kokatzen dira lexikoan batzuetan agertzen diren bi fenomeno:
 - alomorfoen erabilera, hau da, morfema bera adierazteko lexikoan forma bat baino gehiago erabiltzea (adib. *hau* eta *hon*).
 - morfemen desitxuratzeta, morfema bere forma ezagunean edo kanonikoan ez gordetzea, hain zuzen. Aipatuko diren diakritikoak dira honen adibide. Eta euskararen kasuan, esaterako, *r* gogorra/biguna markatzeko erabil liteke: *ur ura*-ren lema eta *huR hurra*-ren lema).

Beste irizpideak *estaldura*, *doitasuna*, *gainsorkuntza* eta *eraginkortasuna* lirateke. Dena den, ez dira aintzakotzat hartu sistemen sailkapenerako, batzuetan inplementazioaren araberakoak direlako, ez formalismoaren ezaugarri. Hala eta guztiz ere, haien definizioak ematea argigarri izan daiteke:

- *Estaldura* (*coverage* ingelesez) deskribapen morfologiko baten zabalera neurtzeko erabiltzen den ezaugarria da, eta *doitasuna* (*precision*) kalitatea neurtzeko. Zenbat eta deskribapen zabalagoa, orduan eta estaldura handiagokoa izango da, hitz gehiago analizatzeko edo sortzeko gai izango baita; hau da, hizkuntzaren atal zabalagoa *estaliko* du. Hitz guztiak analizatzen badira, baina analisi batzuk ez badira zuzenak doitasun falta egongo da. Estaldura neurtzeko *corpus* handi bat hartu ohi da eta analizatutako hitzen ehunekoa kalkulatzeko da.

Doitasuna kalkulatzeko, analizatutako hitzak baino ez dira aztertzen, eta ondo analizatutakoen portzentaia lortzen da. Morfologian doitasuna %99 baino handiagoa lortu ohi da, baina estaldura hain handia izatea zaila da, eragozpen batzuk baitaude tartean: aldaera dialektalak, maileguak, pertsona- zein toki-izenak, laburdurak, etab.

- *Gainsorkuntza* da ekidin beharreko ezaugarri bat (analisiari bakarrik begiratzen zaionean gain-analisisa ere esaten zaio). Gainsorkuntza gertatzen da hizkuntzan existitzen ez diren hitzak sortu edo analizatzen direnean. Adibidez, euskaraz *egin* eta *ten* morfemak kateatzean *egiten* sortu behar da, baina sistema batean gerta liteke *egiten* zein **eginten* onartzea (analizatzea edota sortzea) eta kasu horretan gainsorkuntza gertatuko litzateke. Deskribapen morfologiko sinpleetan gainsorkuntza eragin ohi da, baina deskribapen zehatzetan hori ekiditea da helburu nagusietako bat, batez ere sorkuntza egiteko erabiliko bada. Dena den, morfologia lexikoan, hau da, eratorpenean eta elkarketan, oso zaila da gainsorkuntza ekiditea, ez baita erregularra. Horrela, euskaraz aditzen nominalizaziorako eratorpen atzizkiak asko dira (*-keta*, *-zio*, *-pen*, *-dura*, ...), baina ez dira guztiak aditz guztiekin erabiltzen eta ez dago arrazoi linguistiko argirik, ez sintaktikorik ez semantikorik, kasu batzuetan *-keta* edo besteetan *-pen* erabiltzeko, erabilera ez bada.
- *Eraginkortasuna* abiaduraren sinonimotzat hartu ohi da, eta *analizatutako hitzen kopurua segundoko* da ohiko neurria. Hasierako sistemek hitz gutxi batzuk analizatzen zituzten segundoko, baina gaur egungoak denbora-tarte horretan milaka hitz analizatzera irits daitezke.

2.5 Garapen historikoa eta zenbait sistema

Atal honetan bibliografiako zenbait prozesadore morfologiko aurkeztuko ditugu. Aurkezpenean ez da osoa, adibide adierazgarri batzuk besterik ez baitira azaltzen; hala ere, sistema bakoitzarekin berarengandik gertu dauden beste batzuen bibliografiako erreferentzia ematen da. Aurkezpenean jarraitzen den ordena kronologikoa da (ordezkaria hautatzeko garaian behintzat, nahiz eta aldamenean antzeko adibide berriagoak zehaztu), alde batetik kontzeptuen bilakaeraz konturatzeko, eta bestetik ezaugarrien arabera aurkeztea nahiko konplexua gerta zitekeelako. Dena den, programa

eta informazio linguistikoa nahasten dituzten sistemak baztertu ditugu hasieratik, eskala handiko sistema bat egin nahi denean oso metodologia okerra delako.

2.5.1 DECOMP

Analizatzaile hau lehenengo analizatzaile morfologikoetako bat da, eta, ondoko bertsioak izan baditu ere, 1960.eko hamarkadaren erdialdean garatu zen MITn⁵, *MITalk* izeneko proiektuaren barruan (Allen eta beste, 1987). Hitz-zerrendak biltegiratzeko zeuden arazoek zein sistemaren hedadura nahikoak bere garrantzia eduki bazuten ere, analisi morfologikoa burutzeko izan zuten arrazoi nagusia ingelesez morfologiaren eta hizketaren artean dagoen lotura izan zen. DECOMPen funtsezko ezaugarriak hauek dira:

- Flexioa, eratorpena eta hitzaren mailako elkarketa hartzen ditu kontuan. Analisirako tresna da eta ez du sorkuntzarako aplikaziorik.
- Egoera finituetako morfotaktika erabiltzen du, morfema-motetan oinarritua. Morfotaktika definitzeko erregela sinple batzuk erabiltzen ditu.
- Aldaketa morfofonologikoak oso erregela sinpleen bidez deskribatzen ditu. Aldaketa hauek morfemen artean gertatzera mugatuta daude, eta oso aldaketa sinpleak adieraz daitezke. Morfema baten azken letraren aldaketa, ezabaketa edo sorrera besterik ez da kontuan hartzen sistema honen erregeletan.

Morfemetan banatzeko erabiltzen den algoritmoak eskuinetik ezkerrera tratatzen du hitza, errekursiboa da, eta anbiguitasuna ekiditeko morfotaktikari dagozkion egoera-aldaketei pisu bat esleitzen die, analisi-eredu batzuk beste batzuei gailentzearen eta analisisia azkartzearen. Horrela *scarcity*-ren eratorpen gisako analisisia “*scarce+ity*” lortuko da eta ez “*scar+cite+y*” elkarketa. Eraitzen aldetik, analisisetan %95eko doitasuna lortzen dela diote, baina badirudi neurri hori beste moduluen lana kontuan hartuz egiten dela.

Sistema hau aspaldikoa da, baina urteekin hobetu egin dute. Oso ezaugarri interesgarriak ditu: morfotaktikaren tratamendu dotorea, desanbiguazio-mekanismoa eta eraginkortasuna. Eragozpenak ere leporatu behar zaizkio: analisirako baino ez balio izatea —bere aplikaziorako nahikoa bada ere— eta aldaketa morfofonologikoen ahalmen eskasa —ingelesaren tratamendurako nahikoa izan arren—. Azken arrazoi

⁵ MIT: Massachusetts Institute of Technology.

hauek direla-eta, ez da morfologiarako eredu orokorra, eta ez du jarraitzaile askorik izan.

Espainierarako MARS (Mey, 1987) izeneko analizatzaile morfologikoak antz handia du DECOMP sistemarekin, ezaugarri guztiak, analisia egin ahala burututako desanbiguazioa barne, pareka baitaitezke: analisirako bakarrik balio izatea, egoera finituetako morfotaktika, aldaketa morfofonologikoak oso erregela sinpleen bidez —nahiz eta arlo honetan DECOMPeekin desberdintasunak izan—, eta lexikoan morfemei dagozkien erregelei buruzko informazioa ere gordetzea. Lexikoan alomorfoak erabiltzen dira beren aldaketari dagokion erregela morfofonologikoa orokorra ez denean. MARS (*Morphological Analysis for Retrieval Support*) datuak berreskuratzen laguntzeko sistema baten barruan erabiltzen da.

2.5.2 ATEF

ATEF itzulpen automatikorako ingurune baten barruan dagoen analizatzaile morfologikoa da. Ingurune hau GETA⁶ laborategian erabiltzen da (GETA, 1982), eta 1970.eko hamarkadaren bukaeran garatu zen. Ingurune horren barruan harreman estua du ROBRA izeneko analizatzaile/sortzaile sintaktikoarekin. Hizkuntza askotarako erabili izan da, frantsesa, alemana, errusiera eta Asiako ekialdeko zenbait hizkuntzatarako analizatzaileak eraiki baitira. IXA taldeak prototipo bat burutu zuen euskararako (Arregi & Urkia, 1989). ATEFen osagaiak hauek dira:

- Aldagaiak: analisi morfologikoaren emaitza den informazio morfologikoa jasotzen duten aldagai sinbolikoak.
- Hiztegiak: morfemak biltzen dituzten azpilexikoak. Gehienez zazpi dira, erregeletatik kudea daitezke, eta bertan datuok azaltzen dira: hitz-zatia, hau da, aldatzen ez den morfemaren zatirik luzeena, dagokion formatua eta unitate lexikoa —erro komuna duten hitz-zatiak biltzeko erabilia— eta gainerako informazio morfologikoa.
- Formatuak: hiztegiko unitate-multzo bati dagokion informazioa biltzen duen eredu. Ohikoa da atzizki berdinak hartzen dituzten lexikoko unitateei formatu bera egokitzea.

⁶ GETA: Groupe d'Etudes pour la Traduction Automatique (Grenoble, Frantzia).

- Gramatika (erregelak): erregelen multzoa, hiztegietan aurkitutako hitz-zatiei dagozkien formatuen arabera aktibatzen direnak eta, zenbait ekintza buru daitezen eragiten dutenak. Berauetan bestelako baldintzak ere zehatz daitezke: aldagaien gaineko eragiketak, hiztegien irekitzea edo ixtea, eta testu-aldaketa.

Programak etengabe bilatzen ditu hitz-zatiak hiztegietan, eta aurkitutakoei dagokien informazioa aldagaiei esleitzeaz gain, berauen formatuen arabera aplikatzen ditu erregelak.

Aipatutako osagaiekin, morfotaktikaren zein tratamendu morfosintaktikoaren deskripzioa erraza eta malgua da, salbuespenak modu dotorean adieraztea bideratuz. Aldaketa morfofonologikoen tratamendua, berriz, oso ahula da, horretarako morfotaktika helburu duten erregelak erabili behar baitira. Hori dela eta, aldaketa morfofonologiko sinpleak adierazteko ere, zenbait zeharkako bide erabili behar dira beti.

Horrez gain, beste bi eragozpen ditu sistema honek:

- Programa ezagumendu linguistikotik bereiz badago ere, gramatikaren idazketa ez da erazagutzailea, metahizkuntza agintzaile batetik gertu dagoen zerbait baizik.
- Programa ez da eskuragarria eta bere zehaztasunak ez dira ezagunak; gainera, garaiko IBM *mainframe*-etan baino ezin da erabili.

Martí-k (1987) espainierarako proposatutako AM analizatzaileak, lematizatzaile baten parte denak, zenbait ezaugarri du komunean aurrekoarekin.

2.5.3 KIMMO

Aurretik ikusitako ereduak, morfotaktikaren aldetik nahiko ahaltsuak baziren ere, aldaketa morfofonologikoetarako pobreak ziren. Horren arrazoia aplikazio-hizkuntzen ezaugarrietan aurki daiteke, normalean oso flexio urri eta aldaketa erregular gutxi duten ingelesa bezalako hizkuntzetarako egiten baitziren prozesadore morfologikoak.

Koskenniemi (1983) bere tesian eredu berri bat proposatu zuen, *bi mailatako morfologia* deitutakoa, oso arrakastatsua izan dena: analisi zein sintesirako aldaketa morfofonologikoak adierazteko formalismo ahaltsu, orokor eta eraginkorra⁷ da.

⁷ Ezaugarri hauen gainean ñabardurak egingo dira geroago.

Suomierarako prestatu bazuen ere, berehala etorri zen KIMMO⁸ izeneko ingeleserako bertsioa, Karttunen-ek (1983) egina. Aldaketa morfofonologikoak adierazteko egoera finituetako automata itzultzaileetan konpilatzen diren bi mailatako erregela paraleloak erabiltzen dira. Formalismo hau da euskararako oinarritzko tresnak diseinatzerakoan IXA taldeak aukeratu duena ere.

Hala ere, KIMMO ez zen aldaketa morfofonologikoak deskribatzeko erregela orokorrak diseinatu zituen lehena izan. Beste batzuen artean, aldaketa morfofonologikoak adierazteko Kaplan-ek eta Kay-k (1981) automatatan konpilatzen ziren erregela sekuentzialak erabiltzea proposatu zuten —Koskenniemirengan eragin handia izan zuena—, nahiz tarteko egoerekin arazoak gertatzen ziren. Ondoren *keçi* izeneko prozesadore morfologikoa egiterakoan Hankamer-ek (1986) *sortu eta egiaztatu* filosofiarekin zebiltzan erregela sekuentzialak ere proposatu zituen.

Bi mailatako morfologiaren ezaugarriak zehatz-mehatz hurrengo atalean aztertuko ditugun arren, formalismo guztien artean sailkapen bat egiteko ezinbestekoa da ezaugarri hauek laburtzea:

- Analisi zein sintesirako baliagarria da. Kateatze mailako fenomenoak bakarrik deskriba daitezke, baina, bi mailatako ideia n mailalara zabalduz, beste fenomeno konplexuagoak —hizkuntza semitikoak, adibidez— ebatz daitezke (Kay, 1987; Beesley, 1990; Kiraz, 1994).
- Azpilexikoetan oinarritutako morfotaktika. Morfema bakoitzari bere ondotik etor daitezkeen morfemen multzoa definitzen duen *jarraitze-klasea* egokitzen zaio. Hori dela eta, morfemen kateatze-sekuentzia bateko i -garren morfemak $(i+1)$ -garrena baino ezin du baldintzatu, urruneko mendekotasuna deituriko fenomeno deskribaezina bihurtuz. Beraz, mekanismoa oso sinplea da, baina batzuetan ez da nahiko esanguratsua, eta horrexegatik proposatu dira aldaketak arlo honetan.
- Aldaketa morfofonologikoak. Gorago aipatu den bezala, sistemaren atalik berritzaileena da hau, egoera finituetako automaten ideia modu arrakastatsuan erabiltzen duelako xede honetarako.

⁸ KIMMO izenarekin Koskenniemik proposatutako bi mailatako morfologian oinarritutako prozesadore morfologikoa guztiak izendatuko ditugu.

- Lexikoan morfemak gordetzen dira eta, beharrezkoa ez bada ere, alomorfoak erabiltzea ez du baztertzen Koskenniemi. Morfema desitxuratu, baina erregelen aplikazioa kontrolatzen duten *diakritikoak* (berak *hautapen-markak* deitzen dituenak) erabili ohi dira.

Formalismo honen arrakasta izugarria izan da. Cahill-ek (1989) horrela zioen:

“The field of computational morphology was revolutionized by the work of Kimmo Koskenniemi, whose two-level model of morphonology has been used for the description of several languages, including English, French, Finnish and Japanese.”

Literaturan gehiago aurkitzea erraza bada ere, hona eredu honetaz ari diren erreferentzia garrantzitsu batzuk:

- Hobekuntzak: Karttunen eta beste, 1987; Kay, 1987; Ritchie eta beste, 1987; Bear, 1988; Trost, 1990; Karttunen eta beste, 1992; Karttunen, 1993.
- Inplementazioak (aldaketa handirik proposatu gabe): Karttunen & Wittenburg, 1983; Karlsson, 1992; Clemenceau & Roche, 1993; Oflazer, 1994; Kim eta beste, 1994; Kiraz, 1994; Alegria eta beste, 1996; Sánchez León, 1997.
- Banaketa libreko tresnak: PC-KIMMO (Antworth, 1990; Karp eta beste, 1992).

Gaur egun, bi enpresa gutxienez formalismo hau erabiltzen ari dira produktu komertzialak lortzeko: *Xerox/Inxight* eta *Lingsoft*. Azken hau, 1994an alemana analizatzen sistema onena hautatzeko *Morpholympics* izeneko lehiaketan izan zen irabazle.

2.5.4 Samba

Samba Nuevo Mexico Estatuko unibertsitatean garatutako tresna da, morfologiari aurre egiteko baterakuntza eta egoera finituetarako erregelak modu erazagutzailean konbinatzen dituen (Zajac, 1998). Horrela, bi mailatako morfologian baino morfotaktikaren egitura konplexuagoak defini daitezke.

Morfotaktikaren eredua zeharo aldatzea proposaturik zegoen lehendik, bi mailatako morfologiaren jatorrizkoa oso pobrea delako. Proposamen ezagunenak Bear-ek (1986), Trost-ek (1990, 1994), eta Alvey-n (Ritchie eta beste, 1992) azaldutakoak dira. Hiruretan baterakuntza-mekanismoak proposatzen dira. Bear PATR formalismoan

oinarritzen den bitartean, Ritchie-ren taldean GPSG⁹ aukeratu zuten. Alvey-ren kasuan, baterakuntzak morfotaktika burutzea baino helburu zabalagoa du, eratorpenak sortutako kategoria-aldaketa zein elkarketen eta informazio morfosintaktikoaren tratamendua bideratzen baitu¹⁰ —aipatutako liburuaren (Ritchie eta beste, 1992) hirugarren, laugarren eta bosgarren kapituluak oso gomendagarriak dira—.

Hala eta guztiz ere, Zajac-en proposamena harantzago doa, morfologia zehazteko orduan mota desberdinetako deskribapenak egin daitezkeelako, bakoitza bere aldetik edo modu konbinatuan, aurretik Calder-ek (1988) eta beste batzuk proposatutako eredu paradigmatikoa ere integratuz:

- Hizkien konbinaketa espresio erregularrak erabiliz. Hizkiak ezaugarri-egituren bitartez aberastu daitezke propietate morfologikoak zehaztu ahal izateko.
- Morfologia paradigmatikoa erregela-taulak erabiliz, eta herentzia-hierarkian antolatuta.
- Hizkien eransketa (*agglutination*) aurrizki, atzizki eta erregelen bitartez.
- Morfotaktika konplexua erregelen espresio erregularren bitartez.

Erregelak egoera finituetatako itzultzaileetan konpilatzen dira, baina 3. atalean azalduko diren itzultzaileekin desberdintasun azpimarragarri bat dago: horietan bi mailetan zehazten diren elementuak karaktereak diren bitartean, honetan karaktereak eta ezaugarri-egiturak konbinatzen dira.

Sistema konplexu honen bitartez, hizkuntza askoren deskribapena bideratzeko plataforma bat garatu nahi izan dute, sistema oraindik garatze bidean bada ere.

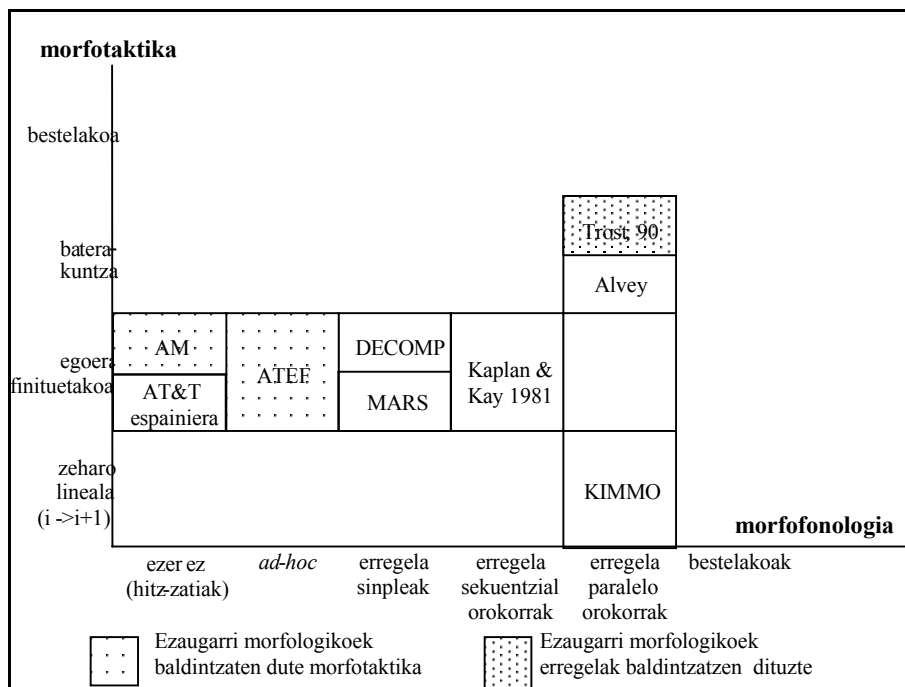
2.6 Sailkapena

Aurreko adibideak aztertu ondoren, 2. irudian ikus daitekeen sailkapena proposa dezakegu.

Sailkapen honetan sartzeko, sistemek ezaugarri hauek bete behar dituzte: morfemez osatutako lexikotan oinarriturik egon eta kateatze-mailako fenomeno morfologikoak deskribatu. Halako sistemetarako eskema orokorra da morfotaktika eta morf fonologia

⁹ Formalismo hauek ez ditugu azalduko, sintaxiaren gaitzat hartu izan baitira, eta gure aplikazioan ez baitira erabili. Hala ere, honetan sakontzeko liburu hau gomendatzen dugu: Shieber S.M. *An introduction to unification-based approaches to grammar*. CSLI Lecture Notes 4. Chicago U. Press. 1986.

bereiztea, eta kasu berriak aztertu ahala egunera liteke irudia. Beste batzuek aipatzen duten bezala (Ritchie eta beste, 1992), beste ereduakin konparazio zehatzak egitea oso zaila da eta, horrexegatik, horiek iruditik kanpo gelditzen dira.



2. irudia.- Azaldutako prozesadore morfoloģikoen sailkapena

2.7 Laburpena

Morfologia konputazionalaren oinarriak eta aplikazioak aztertu ondoren, ohiko sistemen ezaugarri garrantzitsuenak deskribatu dira kapitulu honetan. Horretan oinarrituta, garapen historikoa eta zenbait sistemaren oinarritzko propietateak azpimarratu dira eta sailkapen bat proposatu.

Dena den, azken urteetan arrakasta handiena izan duen proposamena bi mailatako morfoloģiarena izan da, zalantzarik gabe, KIMMO izenarekin deskribatu dena. Horretan oinarritu dira bibliografiako sistema gehienak, batez ere morfoloģia aberatsa duten hizkuntzak aztergai izan dituztenak. Horregatik, proposamen hori aztertuko da hurrengo kapitulan.

¹⁰ Tratamendu honi *morfosintaktikoa* deituko diogu. Ez da harritzekoa kasu honetan, halako tratamendua burutzean, *hitzaren gramatika* terminoa erabiltzea eta ez morfofaktika.

3 Bi mailatako morfologia

3.1 Sarrera

1983an Koskenniemi bi mailatako morfologiaren eredu konputazionala definitu zuen, aurreko sailkapenean KIMMO izenarekin aipatu duguna. Egoera finituetako morfologiari bultzada handia eman zion eredu honek harrera bikaina jaso du ondoko urteetan, besteak beste, dituen ezaugarri hauengatik:

- Eredu orokorra da, kateatze-mailako morfologian behinik behin eta, beraz, gure inguruko edozein hizkuntzari aplikatu dakiogke.
- Ezagutza linguistikoa eta algoritmoa bereizi egiten ditu eta, ondorioz, programa berak edozein hizkuntzatarako balio dezake.
- Baliagarria da hitzen analisi morfologikorako zein sorkuntzarako.
- Analizatu edo sortuko den hitzaren *azaleko maila* eta hiztegian (lexiko-sisteman) errepresentatzen den *lexikoko maila* —sakonekoa ere esaten zaio— argi eta garbi bereizten ditu. Hau dela eta, ez dago aldaketa morfofonologikoengatik sortutako morfema baten forma desberdinak (alomorfoak) gorde beharrik.
- Fonologia sortzaileko berridazketa-erregelak erabili beharrean erregela paraleloak erabiltzen ditu, kontzeptualki zein konputazionalki errazago bihurtuz.
- Aurreko ezaugarriak kontuan hartuz, esan daiteke sistemaren konputazio-komplexutasuna ez dela altua eta, ondorioz, makina txikietan sistema errealak ezartzea bideratzen duela.

Ondoko pasarteetan formalismo honen osagai garrantzitsuenak diren sistema lexikoa, morfotaktika ere definitzen duena, eta erregela morfofonologikoak sakonean azaldu ondoren, sistemari egindako hobekuntzak zerrendatuko dira eta, bukatzeko, garapenak egiteko orduan erabil daitezkeen sistemak azalduko dira.

3.2 Lexikoa

Lexiko-sistemak morfema-multzoa eta morfotaktika definitzen ditu. Funtsezko hiru elementuk osatzen dute lexiko-sistema: lexiko-sarrerek, azpilexikoeak eta jarraitze-klaseak.

3.2.1 Sarrerak

Lexikoan sarrera bakoitzak hiru eremu izan ohi ditu:

- Lexiko-adierazpena, alfabeto lexikoko karaktere-sekuentzia bat da. Karaktere hauek azaleko karaktereak, *morfofonemak* eta *hautapen-markak* izan daitezke. Erregelen bitartez karaktere arruntei zein morfofonemei azaleko beste karaktere bat edo karaktere hutsa egoki lekizkiekeen bitartean, erregelen aplikazioa kontrolatzeko diren hautapen-markei karaktere hutsa besterik ez zaie egokituko. Bibliografian hautapen-markei diakritiko deitzen zaie morfema desitxuratu egiten dutelako, baina beharrezkoak dira erregelen aplikazioa murrizteko.
- Dagokion *jarraitze-klasea*. Geroxeago azalduko den bezala, morfemen arteko sekuentzia posibleak, morfotaktika ere deitu duguna, erregulatzen dituzte jarraitze-klaseek.
- Sarrerari dagokion *informazio morfologikoa*, analisiaren emaitza gisa agertuko den informazioa, alegia.

Erabiltzen den programaren arabera informazio hau adierazteko sintaxia alda daiteke, baina, edozein kasutan, zehaztu beharreko informazioa da.

Sarrerak azpilexikoetan multzokatzen dira, morfemen artean egon daitezkeen kateamenduen arabera. **Azpilexikoak**, ezaugarri berdineko elementu lexikoak (morfemak) biltzeko erabiltzen dira, betiere morfemekiko kateatzeari dagokionez. Hori dela eta, morfotaktikak behartzen du azpilexikoen eraketa, linguistikoki multzo artifizial samarrak diren arren. Horrela, deklinabide-atzizki guztiak, adibidez, ezin dira azpilexiko berean egon, batzuk lema-motaren arabera aukeran daudelako.

Azpilexiko guztien definizioek egitura bera dute: identifikatzailea den izena, dagozkion ezaugarriekin zenbait kasutan, eta sarrera-multzoa.

Aldaketa morfofonologikoak burutzeko erregela-multzoa dagoenez, lexikoan alomorforik definitzeko beharrik ez dago. Hala ere, Koskenniemi bi kasu berezitan onartzen ditu:

- azala eta lexikoaren artean aldaketa handiak edota ez-erregularrak gertatzen direnean. Horien artean daude berak erabiltzen dituen *aukera-patroiak*¹¹ izeneko azpilexiko txikiak. Adibidez: *ona:hobea* edo *good:best*.
- jarraitze-klaseak eratzeko orduan. Azpilexiko txiki asko sortzea lexiko-sarrerak errepikatzea baino egokiago jotzen badu ere, gehiegizko sakabanaketa ekiditeko horrelakorik egin daiteke.

3.2.2 Morfotaktika

Jarraitze-klasea azpilexiko multzo bat da, morfotaktikaren aldetik unitate bat dena, eta paradigma baten osagaiekin identifika daitekeena. Identifikatzaile batez ezagutzen da. Esan bezala, jarraitze-klase bat egokitzen zaio lexikoan morfema bakoitzari, eta jarraitze-klasean biltzen diren osagaiak dira definitutako sarreraren ondoren ager daitezkeen morfema guztiak eta bakarrak. Beraz, jarraitze-klaseak hitz batean ager daitezkeen morfemen arteko konbinaketa posibleak definitzeko mekanismoaren oinarri dira.

Morfotaktika-sistema honen deskribapen-ahalmena, esan bezala, txikia da oso, eta honegatik, kasu batzuetan beharrezkoa izango ez litzatekeen zenbait deskripzio-bikoizketa gertatu ohi da, aipatutako morfemen arteko urruneko mendekotasunaren kasuan, esaterako. Hori dela eta, aldaketak proposatu dira arlo honetan.

Jarraitze-klaseen eta azpilexikoen arteko bereizketa ez da funtsezkoa, lexiko-sistema beste modu honetan ikus baitaiteke: estekatutako azpilexiko-multzoa. Horrela, zenbait sistematan jarraitze-klase bat definitzeko azpilexikoz osatutako azpilexiko berri bat zehazten da. Ikuspegi honetatik, morfotaktika definitzen duen grafoa ondo eratzeko elementuak besterik ez dira jarraitze-klaseak eta azpilexikoak.

3.2.3 Adibideak

Sinplifikazio bat eginez, euskarazko adjektiboek eta izen arruntek deklinabide-atzizki berberak hartuko dituzte¹², baina lehenek baino ezin dute gradu-flexiorik hartu. 3. irudian ikus daiteke lexiko sinplifikatu honen eraketa.

¹¹ *aukera-patroiak*. Azpilexiko txikiokin egiten dena hau da: aldaketa morfofonologiko ez-erregularretarako gertatzen diren aldaketak jarraitze-klase desberdinetan antolatzea, eta jarraitze-klase berbera behar zuten morfemei desberdinak esleitzea (morfofonologiaren zenbait arazo morfotaktika bihurtuz).

¹² Noski, gradua hiztegi-sarrerari lotzen zaio zuzenean; deklinabide-kasua, aldiz, mugatasunetik pasa ondoren.

LEXIKOA izenak*ize1* izen_jk ...*ize2* izen_jk ...

...

LEXIKOA adjektiboak*adj1* adj_jk ...*adj2* adj_jk ...

...

LEXIKOA gradua*grad1* izen_jk ...*grad2* izen_jk ...

...

LEXIKOA deklinab*dek1* # ...*dek2* # ...

...

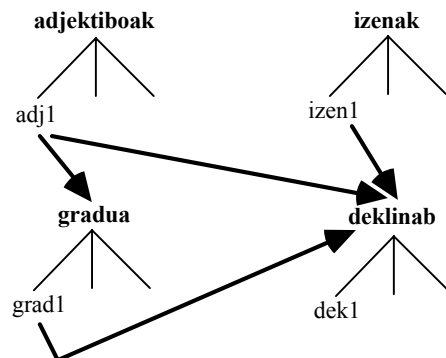
JARRAITZE-KLASEA izen_jk

deklinab

JARRAITZE-KLASEA adj_jk

deklinab

gradua

**3. irudia.- Lexiko-sistemaren erazagutzaren eta egituraren adibidea**

Azpilexikoen eta jarraitze-klaseen bidezko deskribapenetik dagokion grafoa lor daitekeela ere ikus daiteke 3. irudian. Alderantzizko bidea ere egin daiteke, alegia, grafotik abiatuta deskribapena lortzea.

3.1 adibidea: oinarri-oinarrizko lexiko bat euskaraz

lexc programaren sintaxiari jarraituz, ondoren azaltzen den deskribapena egin beharko litzateke (ikus 4. kapitulua, programaren nondik norakoetan sakontzeko). Kontuan hartu behar da hemen ez dela azpilexikoen eta jarraitze-klaseen definizioen artean bereizketarik egiten, jarraitze-klase bat definitzeko morfema hutsez osaturiko hainbat azpilexiko zehaztu besterik ez delako egiten.

Sintaxiaren aldetik bi lerro-mota bereizten dira: batetik *LEXICON* hasiera dutenak, azpilexikoak edo jarraitze-klaseak definitzeko, eta, bestetik, gainontzekoak, osagaiak adierazteko; azken hauek ‘,’ bereizgarriarekin bukatzen dira.

‘+’ karakterea jarri ohi da aurrizkien bukaeran eta atzizkien hasieran, erregela fonologikoetan morfema-mugaren marka gisa erabili ahal izateko.

```
LEXICON Root
  IZENAK;
  ADJEKTIBOAK;

LEXICON IZEN_JK
  DEKLINAB;

LEXICON ADJ_JK
  GRADUA;
  DEKLINAB;

LEXICON IZENAK
ate      IZEN_JK;
etxe     IZEN_JK;
ur       IZEN_JK;
...

LEXICON ADJEKTIBOAK
gorri    ADJ_JK;
...

LEXICON DEKLINAB
      #;
+a      #;
+ak     #;
+aren   #;
...

LEXICON GRADUA
+ago     DEKLINAB;
+en      DEKLINAB;
```

Deklinabidea berariaz sinplifikaturik ageri da, zenbait kasutan, genitiboan adibidez, beste deklinabide-kasuak erants daitezke-eta. Mugatasuna eta kasua ere batean bildu ditugu hemen.

Morfotaktika aplikatzeak eragindako irregulartasunak —forma defektiboak, etab.— kontrolatzeko jarraitze-klase bereziak sortu beharko dira eta erro/lema horiei esleitu. Adibidez, ingelesezko izenen jarraitze-klase erregularretik at geratuko dira plural berezia duten hainbat izen (*octopus*, *bus*, ...), erregela korapilatsu anitz idaztea nahi ez bada behintzat.

Adibideetako deskribapenarekin, analisiaren emaitza gisa morfema-banaketa lortuko genuke analisi morfologikoa baino gehiago. Koskenniemiren ohiko programetan, informazio morfologikoa estekaturik dago fitxategi batean. Lexiko-itzultzaileei esker informazio morfologikoa grafoaren barruan koka daiteke, 3.4 pasartean azalduko den moduan.

3.3 Erregela fonologikoak

Morfemak eranstea edo elkartzea gertatzen diren aldaketa fonologikoak, ortografikoak edo morfologikoak —morfofonologikoak ere deitzen dira—

deskribatzeko erabiltzen dira erregela hauek. Mota desberdinetakoak izan daitezke eta kontu handiz erabili beharko dira, oso mekanismo ahalsua da-eta.

3.3.1 Sarrera

Koskenniemiaren ekarpenik handiena, zalantzarik gabe, aldaketa morfofonologikoak deskribatzeko sortu zuen bi mailatako erregela-sistema izan zen,. Hori argi ikusten da bibliografian, nahiz, zenbait autore, hori kontuan harturik, eredu honen izena zalantzan jartzera iritsi diren, bere ordeztu bi mailatako fonologia proposatuz. Gorago aipatu dugun R. Sproat-en liburuan (1992: 92-93) aipamen hau egiten da:

“...The bulk of the machinery is designed to handle phonological rules, and in the original version of that system the actual morphology done is particularly interesting. Indeed, the term two-level morphology, used by Koskeniemi to describe the framework, is really a misnomer: the “two-level” part of the model describes the method for implementing phonological rules and has nothing to do with morphology per se.”

Bi mailatako erregelek errepresentazio lexikoaren eta azalekoaren artean parekatzea kontrolatzen dute. Erregelak egoera finituetako automata itzultzaile (FST)¹³ paralelo bihurtzen dira, eta karaktere-bikoteak onartuko dira baldin eta automata guztietan onartzen badira. Bi errepresentazioen artean, lexikokoaren eta azalekoaren artean hain zuzen, ez dago tarteko egoerarik, eta hauxe da fonologia sortzailearekiko desberdintasun nagusia. Beraz, hitzen analisisa azaleko formari dagozkion lexikoko errepresentazio onargarriak aurkitzean datza. Alderantziz gertatzen da sorkuntzan, lexikoko errepresentazio ezagunetik abiatu, eta berari dagozkion azaleko errepresentazioak bilatzen baitira.

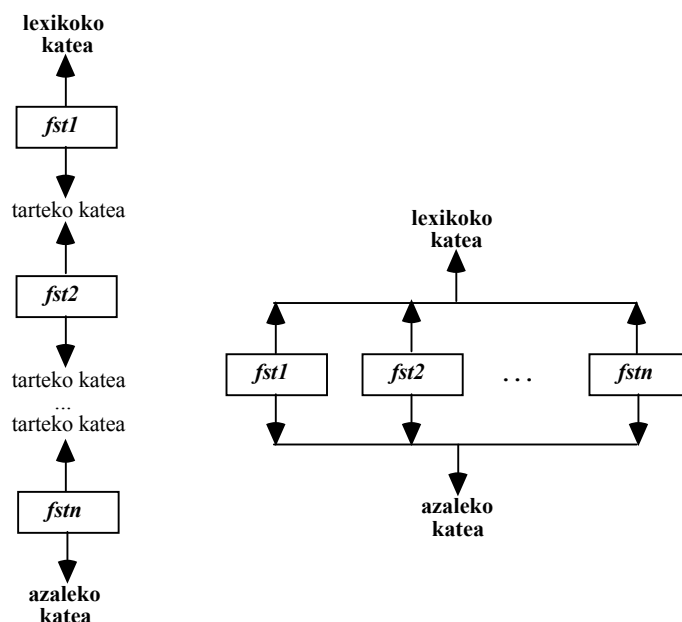
Aipatu den bezala, Koskenniemi proposatutakoaren (morfo)fonologia-eredu honen aurretik Kaplan-ek eta Kay-k (1981) berridazketa-erregeletan oinarritutako beste eredu bat proposatu zuten, erregela sekuentzialena, hain zuzen ere. Beren ereduan ere, erregelak egoera finituetako itzultzaileetan konpilatzen ziren¹⁴, eta, ondorioz, analisi zein sorkuntzarako balio zuen.

¹³ Egoera finituetako itzultzaile (*Finite State Transducer* - FST) eta egoera finituetako automata (*Finite State Automaton* - FSA) baten artean dagoen desberdintasuna hau da: FSAren alfabetoko osagaiak sinbolo sinpleak diren bitartean, FSTarenak bikoteak dira. *Transducer* ingelesezko hitzaren ordaina euskaraz *automata* itzultzaile edo *itzultzaile* izango da testuan, idazle batzuk *transduktore* ematen duten arren.

¹⁴ Ideia hau Jhonson-en (1972) ekarpena da.

Kaplan eta Kay-ren eredu hau eta Koskenniemiarena alderatuz —eratorpen-fonologia eta erazagutze-fonologia izenekin bereizten ditu Karttunenek (1992)—, desberdintasunok aurki daitezke:

- Kaplan eta Kay-ren ereduak ikuspegi sortzailea du; beraz, aldaketa morfofonologikoen arazoa urrats sinpleen bidez adieraz daiteke, eredu teoriko sinplea baita.
- Tarteko egoerak sortzen ditu, eta, beraz, sekuentziak garrantzi handia du. Ondorioz, testuinguruaren espezifikazioa konplexua bihur daiteke praktikan, ordenaren arabera aurreko erregelen emaitzak kontuan hartu behar direlako.
- Analisi zein sorkuntzarako baliagarria bada ere, sorkuntzaren kasuan prozesua ez-determinista gerta daiteke.



4. irudia.- Egoera finituetako itzultzaile ordenatu eta paraleloen arteko konparaketa

Hala ere, Kaplan-en arabera (Kaplan, 1988; Kaplan & Kay, 1994), bi eredueta karaktere-kateen arteko erlazio erregularrak adierazten dira, eta erlazio hau itxia da konposaketarekiko, baina ez ebaketarekiko. Izan ere, bi mailatako morfologiaren kasuan ebaketa ere itxia da. Honetan oinarriturik, Karttunen-ek bi ereduak erabiltzen dituzten lexiko-itzultzaileak proposatzen ditu, kapitulu honen amaieran azalduko direnak.

4. irudian konputazioaren aldetik bi ereduen artean dagoen desberdintasuna azaltzen da: alegia, bi sistemetan, erregelak egoera finituetako itzultzaile bihurtu arren, bi mailatako morfologian itzultzaile baino gehiago bikote-kontrolatzaileak dira.

Bi eredu ezaugarriez eta formalizazioaz sakontzeko oso egokiak dira Karttunen-en (1991) eta Kaplan & Kay-ren (1994) artikuluak.

3.3.2 Erregelen osagaiak

Esan bezala, bi mailatako morfologiaren ezaugarriarik garrantzitsuen lexiko-maila eta azalekoa parekatzen duten erregela-multzoan datza. Erregela hauen sintaxia zehaztu baino lehen, zenbait kontzeptu defini ditzagun:

- *bi mailatako konfigurazioa*: lexikoko eta azaleko karaktereak banan-banan sinkronizatzen dituen karaktere-kate pare. Adibidez:

e g i N + t e n (lexikoa)

e g i 0 0 t e n (azala)

Konbentzioz, pareetan beti lexiko/azala ordena mantenduko da. Zeroak karaktere hutsa adierazten du —beste lanetan ϵ sinboloarekin adierazten dena—. N morfofonema edo diakritikoa da, gal daitekeen n karakterea adierazteko.

- *azaleko alfabetoa*: analizatzeko formetan ager daitezkeen karaktere guztiak.
- *lexiko-alfabetoa*: lexikoan ager daitezkeen karaktere guztiak. Azaleko karaktereez gain, morfofonemak eta hautapen-markak (diakritikoak) daude.
- *karaktere-bikote konkretua*: lexikoko eta azaleko karaktere banaz osatutako bikotea. Øa bi mailetan ager daiteke, azalean morfofonema zein hautapen-markekin parekatzeko, eta orokorrean azaleko karaktereren baten desagertzea edo sorrera adierazteko.
- *karaktere-multzoa*: ezaugarri komunak dituzten karaktereez osatutako multzoa, identifikatzaile batez ezagutzen dena. Horrela V edo $Bokal$ bokalen multzoa izan ohi da, eta C edo $Konts$ kontsonanteena. Konbentzioz, = edo ? karaktereak alfabetoko edozein karaktere edo Ø adierazten du.
- *karaktere-bikote abstraktua*: lexiko-mailan eta azalekoan karaktere konkretu bat eduki beharrean, karaktere-multzoa duen bikotea. Maila batean karaktere konkretua eta bestean multzoa dutenei bikote erdiabstraktu deritze.

3.3.3 Erregelen formatua

Erregelen hasierako sintaxiak aldaketa batzuk izan ditu (Koskeniemi, 1985; Ritchie eta beste, 1992, Karttunen, 1993) deskribapen-ahalmena irabazi eta konpiladoreak inplementatu ahal izateko, eta automatetarako itzulpena eskuz egin behar ez izatea

ahalbideratzen duena gainera. Aipatutako azkena erabiliko da hemen, konpiladore hori balia dezakegu-eta.

Erregelen formatua eta osagaiak hauek dira:

formatua cp op lc _ rc

Korrespondentzia (cp), karaktere-bikote bat da. Karaktere hauek konkretu nahiz abstraktuak izan daitezke, eta azken hauek erregelen orokortzea ahalbidetzen dute. Gehienetan bikote konkretuak dira.

Eragilea (op), testuinguruaren eta korrespondentzian adierazitako bikotearen artean zer-nolako erlazioa dagoen finkatzen duena da. Lau eratakoa izan daiteke: *testuinguru-murritzapena* ($\Rightarrow$), *azalekoaren derrigortzea* ($\Leftarrow$), aurreko biak batera ($\Leftrightarrow$) edo *debeku-ezarpena* ($/\Leftarrow$). Azaleko derrigortzearena berridazketa-erregelatan erabiltzen denaren baliokidea izango litzateke. Azken mota, debekuarena hain zuzen, Bear-ek (1986) proposatu zuen, salbuespenen tratamendua modu errazagoan adierazi ahal izateko.

Bakoitzaren esanahia hau da:

op	adibidea	interpretazioa
$\Rightarrow$	$l:a \Rightarrow lc_rc$	l lexikoko karakterea azalean a bihurtzen da baldin testuingurua lc_rc bada.
$\Leftarrow$	$l:a \Leftarrow lc_rc$	lc_rc testuinguruan l beti gauzatzen da a bezala.
$\Leftrightarrow$	$l:a \Leftrightarrow lc_rc$	l karakterea a bezala gauzatzen da lc_rc testuinguruan eta ez beste inoiz.
$/\Leftarrow$	$l:a /\Leftarrow lc_rc$	l ez da inoiz a bezala gauzatzen lc_rc testuinguruan.

Gehien erabiltzen dena eragile konposatua ($\Leftrightarrow$ motakoa) da, baina aldaketa baten gauzatzea aukeran denean, testuinguruaren murritzapena ($\Rightarrow$ motakoa) ere erabili ohi da.

Testuingurua (lc_rc), korrespondentzia zein kasutan gertatzen den mugatzen duena, aurreko eta ondoko karaktereen arabera. _ karaktereak ezkerreko edo aurreko testuingurua (lc) eta eskuineko edo ondoko testuingurua (rc) banatzen ditu, bietako bat hutsa izan badaiteke ere. _ karaktereak, beraz, aldaketaren posizioa adierazten du.

Ezkerreko zein eskuineko testuinguruetan karaktere-bikoteen segidak adierazten dira; hala ere, bikotearen bi osagaiak berdinak direnean, karaktere bakarra zehaztea nahikoa da. Bi puntuak eta karaktere bakar bat zehazten bada, orduan karaktere horrekin era daitezkeen bikote posible guztiak erreferentziazten dira. # sinboloak hitzaren muga adierazten du; beraz, ezkerreko testuinguruan hitz-hasiera eta eskuinekoan bukaera adieraziko du. Testuingurua zehazteko modua programa/konpiladorearen arabera izango da, baina espresio erregularretan oinarritu ohi dira. 4. kapituluaz azalduko da erabiliko dugun sintaxia.

Bessley eta Karttunen-en (2002) liburutik hartutako adibidea eranstean da 4. irudian, erregela-motaren arabera debekatzen diren pareak zeintzuk diren adierazteko.

erregela	karaktere-kateak (goian lexikoa eta behean azala)			
$a:b \Rightarrow l_r$	lar	lar	lbr	xay
	lbr	lar	lbr	xby
$a:b \Leftarrow l_r$	lar	lar	lbr	xay
	lbr	lar	lbr	xby
$a:b \Leftrightarrow l_r$	lar	lar	lbr	xay
	lbr	lar	lbr	xby
$a:b / \Leftarrow l_r$	lar	lar	lbr	xay
	lbr	lar	lbr	xby

5. irudia.- Debekatutako bikoteak erregela-motaren arabera

3.3.4 Erregelatik automatara

Koskenniemik proposatutako eredu honen funtsezko ezaugarria egoera finituetakoa izatea da, horrek eraginkortasuna areagotzen du-eta. Ereduaren atal konplexuena erregela-multzoa da, baina, ahalmena handia izan arren, grafo edo automata¹⁵ bihur daitezke; beraz, egoera finituetakoa ere bada atal hau.

Erregelatik egoera finituetako automatara edo itzultzailera¹⁶ iragan ahal izateko konpiladoreak daude, baina, interesgarria da eskuz nola egiten den jakitea, erregela-mota desberdinen esanahia hobeto ulertu ahal izateko.

¹⁵ Automatak grafo bihur daitezke batere arazorik gabe.

¹⁶ Automata, itzultzaile eta transduktorea sinonimo gisa erabiltzen dira, nahiz eta itzultzaileek edo transduktoreek ezaugarri konkretuak (etiketa bikoitzak) dituzten automatak izan.

$l:i$ bikote bat bada, $a:b$ ezkerreko testuingurua eta $c:d$ eskuinekoa, ikus ditzagun lau erregela-motak eta dagozkien egoera finituetako itzultzaileak. Ondoko konbentzioak hartu beharko ditugu kontuan: bikoteetan lehen karakterea (goikoa automata marraztean) lexikokoa da eta bigarrena (behekoa) azalekoa; automataren 0 egoerak ezinezko bidea adierazten du, egoeraren ondoko bi puntu sinboloak bukaera-egoera eta puntu bakarrak ez-bukaerakoa. Demagun karaktere bakarreko testuingurua dugula lexikoan zein azalean, luzeagoa denean orokortzea oso erraza baita.

- testuinguru-murriztapena: $l:i \Rightarrow a:b _ c:d$

aldaketa gertatzeko testuingurua bete behar denez, ezkerreko testuingurua egiaztatu arte $l:i$ bikotea ez da onartzen, eta ezkerreko testuingurua egiaztatu ondoren, eskuineko testuingurua egiaztatu arte gainontzeko bikoteak debekatzen dira eta egoera ez-bukaerakoa zehaztuko da. 0 zenbakiak automataren porrota adierazten du, eta egoera ondoko '.' karaktereak egoera ez-bukaerakoa. Automataren ondoren dagokion grafoa erakusten da, automaten eta grafoen arteko parekotasuna azaltzeko.

```

a 1 c =
b i d =
1: 2 0 1 1
2: 2 3 1 1
3. 0 0 1 0

```

- azalekoaren derrigortzea: $l:i \Leftarrow a:b _ c:d$

ezkerreko testuingurua egiaztatuz doa eta lexikoko l karakterea duen $l:i$ ez den beste edozein bikote debekatzen da eskuineko testuingurua egiaztatzen denean.

```

a 1 1 c =
b i = d =
1: 2 1 1 1 1
2: 2 1 3 1 1
3: 2 1 1 0 1

```

- aurreko bion konposaketa: $l:i \Leftrightarrow a:b _ c:d$

```

a 1 1 c =
b i = d =
1: 2 0 1 1 1
2: 2 3 4 1 1
3. 0 0 0 1 0
4: 2 0 1 0 1

```

- debeku-ezarpena: $l:i / \leq a:b _ c:d$
testuingurua egiaztatuz doa eta $l:i$ bikotea debekatzen da testuinguru horretan.

```

a l c =
b i d =
1: 2 1 1 1
2: 2 3 1 1
3: 2 1 0 1

```

Gaian gehiago sakontzeko, Antworth-en (1990) PC-KIMMO liburuaren hirugarren kapitulua kontsulta daiteke.

3.3.5 Erabilera

Erregela konposatua ($\leq$ motakoa) da erabiliena, ia aldaketa guztiak ezinbestekoak baitira hizkuntza gehienetan.

Hautazko aldaketa bat deskribatu nahi denean testuinguru-murritzapena ($\Rightarrow$ motakoa) erabiltzen da, baina kontuan hartu behar da horrek, aldaketa benetan hautazkoa ez bada, gainsorkuntza ekar dezakeela, zehazten den aldaketaz gain, beste aukera ere, aldaketarik gabekoa hain zuzen, onartzen duelako.

Debeku-ezarpena ($/\leq$ motakoa) salbuespen-kasu konkrituak deskribatzeko erabili ohi da, alegia, deskribatuko den aldaketa bat testuinguru jakin batean eraginik gabe utzi nahi denean.

Azalekoaren derrigortzea ($\leq$ motakoa) hizkuntza gutxi batzuetan besterik ez da erabiltzen.

3.3.6 Konpiladorea

Eskuzko konpilazioa lagungarri da erregelen sintaxia ondo ulertzeko; hala ere, zaila ez bada ere, erregela-kopurua eta testuinguruen konplexutasuna handitu ahala, lana neketsu bihurtzen da, eta akatsak aurkitu eta zuzentzeko oso lan nekagarria burutu behar da. Gainera, eskuzko konpilazioaren ondoren gerta liteke erregelen esanahia eta automatena bat ez etortzea. Hau ekiditeko zenbait konpiladore sortu dira, haien artean Koskenniemiarena berarena (Koskenniemi, 1985), Ritchie-ren taldearena (Ritchie eta beste, 1992), Xerox-en (Beesley & Karttunen, 2002) *lexc* zein *twolc*, eta *multext* proiektuko *mmorph* (ikus 3.5 pasartea). Gainera, erregelen arteko gatazkak detekta daitezke eta, kasu batzuetan, automatikoki konpondu ere bai.

3.3.7 Adibideak

Erregela sinple batzuk ikusiko ditugu atal honetan, fenomeno morfologiko ezagunen deskribapena —sinplifikatua hainbat kasutan— bi mailatako erregelen bidez nola egiten den azaltzeko.

3.2 adibidea: *ber/bir* aldaketa euskaraz

lexikoko *ber* morfema *bir* bihurtzen da azalean (testuan), kontsonantez hasten diren aditz-errien aurrean (*ber+findu:birfindu*). Bokal baten aurrean *r* letra bikoiztu behar dela ahaztuko dugu momentuz.

```
e:i <=> # b _ r +:0 Konts ;
        ! ber+findu:birfindu
```

e/i (*e* lexikoan, *i* azalean) aldaketa gertatzen da (eta ezinbestekoa da) *ber* morfemaren ondoren kontsonante bat agertzen denean (*Konts* sinboloa definitzeko aukera emango digu konpiladoreak). *+:0* jarri beharrean + jarriko bagenu, azalean (testuan) *bir+findu* agertu beharko litzateke. Bestalde, # (hitz-hasiera) karaktererik gabe ez ginateke ziur egongo *ber* morfemaz ari garela eta ez, esaterako, *ber* bukaera duen batez (*bereber*, adibidez).

3.3 adibidea: *r* gogorra euskaraz

3.2 adibidean aipatu den bezala, *r*-z bukatutako aurrizkia bokalez hasitako aditzoin bati eranstean zaionean, *r* hori bikoiztu egiten da azalean. Erregela hori deklinabide-kasuetara hedatu daiteke, *r* bukaera duten lemetan *r* hori bikoiztu egingo baita atzizkia eranstean. Beraz, erregela hau idatz genezake:

```
0:r <=> r _ +:0 Bokal ;
        ! ber+egin:berregin
        ! zakur+a:zakurra
```

0:r bikotea horrela interpreta daiteke: lexikoan ageri ez den *r* bat azalean sortzen da testuinguru horretan.

Baina badaude salbuespenak: *ur+a:ura* edo *ager+i:ageri*. Salbuespen hauek kontuan hartzeko pentsa daiteke nahikoa dela debeku-erregela bat idaztea, baina hau posible litzateke morfema bat edo multzo konkretu baterako eta ez, kasu honetan bezala, multzoa irekia denean. Ondoko hautazko erregela ere ez da egokia:

```
0:r => r _ +:0 Bokal ;
```

hautazko erregela honekin gainsorkuntza eragingo baikenuke: *zakurra* zein **zakura* onartuko (analizatu eta sortu) litzateke *zakur* lemaaren absolutibo singular gisa. Beraz, beste zerbait behar da: ezaugarri morfologikoek baldintzatutako erregelak, Trost-ek edo Zajac-ek proposatzen duten bezala, edo diakritikoen erabilera. Formalismo honen bitartez diakritiko bat erabiliko da, itxurarenatik zein esleipen-egokitasunarenatik aukera daitekeena. Adibide honetan *r* gogorra dutenei esleituko diegu *R* marka *r* ordeztuz, azalean beti *r* gauzatuko dena, baina erregelen aplikazioa kontrolatzeko ere balio izan duena. Beraz, lexikoan *beR*, *zakuR* zein *mozkoR* agertuko dira baina *ur* zein *ager* ere bai. Ondorioz, erregela horrela geratuko da¹⁷:

```
0:r <=> R:r _ +:0 Bokal ;
        ! beR+egin:berregin
        ! zakuR+a:zakurra
        ! ur+a:ura
```

3.2 adibidean idatzitako erregela aldatu beharko da lexikoan egindako aldaketa islatzeko (beraz, esan den bezala, erregelak ez dira erabat independenteak):

¹⁷ *R* markaren azaleko baliokidea *r* izango da beti, baina *0:r* aldaketa gertatzeko *R* marka agertu beharko da. Aurrerago formalismo ahaltsuagoak azalduko dira eta orduan *R* erabili beharrean [Rgog] erabiliko da.

```
e:i <=> # b _ R:r +:0 Konts ;
        ! beR+findu:birfindu
        ! beR+egin:berregin
```

Galdera bat geratuko litzateke erregela hauei begira: zer gertatzen da *h*-ren aurrean? *ber+hasi:berrasi* behar duenez, *h* letra ez litzateke kontsonanteen artean egon beharko, eta erregela bat beharko litzateke *h*-ren galera kontrolatzeko:

```
h:0 <=> b e R:r +:0 _ Bokal ;
        ! beR+has:berras
```

Testuingurua alda liteke *h*-ren galera orokortzeko aurrizki baten atzean.

3.4 adibidea: hautazko erregela bat euskaraz

Toki-izenetan deklinabide berezia gertatzen da zenbait kasutan. Adibidez, lexikoko *Gasteiz+ra* kateak *Gasteiza* zein *Gasteizera* sortzen du azalean; beraz, *r* karakterearen galera hautazkoa litzateke. Hau egiteko toki-izena adierazten duten hautapen-marka bat zehaztuko da lexikoan (*T*, esaterako):

```
r:0 => z T:0 +:0 _ a # ;
        ! 9gasteizT+ra:9gasteiza18
```

3.5 adibidea: plurala gaztelaniaz

Demagun, pluralerako, lexikoan *s* morfema erabiltzen dela gaztelaniaz (ez *s* eta *es* alomorfoak). Kontuan hartu behar da *z* bukaera duten lemetan *z:c* aldaketa gertatzen dela pluralean. Beraz, bi erregela beharko dira, bat *e* sortzeko (epentesia deitzen zaio fenomeno honi) kontsonantez bukatutako lemetan, eta bestea pluralean *z* bukaera *c* bihurtzeko (*pez:peces*).

```
0:e <=> Konts +:0 _ s # ;
z:c <=> _ +:0 0:e s # ;
        ! pez+s:peces
        ! casa+s:casas
```

karaktereak hitz-bukaera adierazten du, kasu honetan eskuineko testuinguruan agertzen delako.

3.3.8 Zenbait fenomeno morfologikoren deskribapena

PC-KIMMO (Antworth, 1990) liburuaren 5. kapituluaren adibide asko ageri dira eta bertan hizkuntza-espektro zabal batean gertatzen diren fenomeno fonologiko anitz deskribatzen dira. Bertako sailkapenaren arabera, fenomeno fonologikoak 4 motatan bereizten dira: *asimilazioa* (edo aldaketa), *ezabaketa*, *txertaketa* eta *fenomeno ez-jarraituak*.

Asimilazioaren adibide sinple bat azaldu da euskarazko *ber/bir* aldaketan, *e:i* aldaketa deskribatzen duen erregelaren bitartez, esate baterako, baina fenomeno konplexuagoa izan daiteke, turkieraren edo suomieraren bokal-harmoniaren kasuan gertatzen den bezala. Euskaraz asimilazioa ugaria da: *bait* aurrizkia aditz jokatuarekin biltzean (*bait+gara -> baikara*), edo leku-genitiboaren kasua zenbait toki-izenen ondoren

(*Usurbil+ko* -> *Usurbilgo*). Halakoetan, aldaketa kontrolatzen duen erregela bat egon ohi da karaktere-bikote konkretu ($g:k, k:g$) edo abstraktu bakoitzeko.

Ezabaketaren adibidea h -ren eta r -ren galera izan da, baina euskaraz beste horrenbeste daude: a organikoaren galera zenbait kasutan (*ama+en* -> *amen*), bukaerako n -ren galera aditzetan (*egin+ten* -> *egiten*), etab. Kasu honetan bikotearen eskuinaldean $\emptyset$ karakterea azaldu ohi da ($a:\emptyset, n:\emptyset$).

Txertaketa edo epentesia da sailkapenaren hirugarren fenomeno, eta adibide gisa r gogorrena aurkeztu da gorago. Erregela hauetan $\emptyset$ karakterea agertzen da korrespondentziaren ezker aldean ($\emptyset:r$), eta oso ohikoa da bokalen epentesia kontsonanteen artean. Euskaraz, adibidez, e epentetikoa maiz gertatzen den fenomeno da. Fenomeno hau eta euskaraz gertatzen diren beste guztiak 5. kapituluaren aurki daitezke.

Fenomeno konplexuagoak gertatzen dira hainbat hizkuntzatan. Filipinetako *tagalog* hizkuntzan, esaterako, gertatzen den bikoizketa ikus dezakegu ondoren:

3.6 adibidea: bikoizketa tagalogez

Tagalogezko bikoizketa hau hitz-hasieran gertatzen da eta, ondorioz, erroaren *kontsonante-bokal* moduko lehen agerpena errepikatzen da batzuetan. Horrela, *pili* ('hautatu') *pipili* bihurtzen da, eta *kuha* ('hartu') *kukuha*. P, t eta k kontsonante herikari ahoskabeen aurrean bakarrik gertatzen ei da hau. Bi mailako morfologian deskribatzeko, lexikoan bi morfofonemaz (B eta I , adibidez) osatutako aurrizki bat jarriko da ($BI+$), eta erro guztiekin eraikiko da bere jarraitze-klasea. Ondoko bi erregelak beharko dira bikoizketa gauzatzeko:

```
B:HersAhosk => _ I:Bokal +:0 HersAhosk;
I:Bokal    => B:Konts _ +:0 Konts Bokal ;

          ! BI+pili: pipili
          ! BI+kuha: kukuha
```

HersAhosk sinboloak aipatutako hiru kontsonanteak biltzen ditu, herikari ahoskabe baitira, eta, bi erregelak erlazionatuta badaude ere, ez dute paraleloak izan beharrik. Esaterako, bigarren $B:Konts$ zehaztu da, baina $B:HersAhosk$ zehaztu beharko litzatekeela pentsa litekeen arren, hau ez da ezinbestekoa, B -ren bihurtzea 1. erregelan kontrolatzen baita eta nahikoa da.

3.4 Hobekuntza: lexiko-itzultzaileak

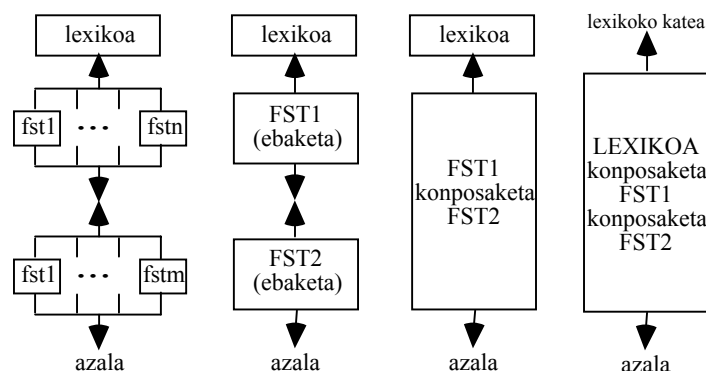
Xeroken egindako ekarpena (Karttunen eta beste, 1992; Beesley & Karttunen, 2002) bi mailako morfologiatik harantzago doa. Proposatutako *lexiko-itzultzaileak* (*lexical transducer* ingelesez) ahalmenaren aldetik dakarten iraultzaz gain (azaleko sintaxiaren

¹⁸ Erregela hau sinplifikazio bat da, e epentetikoaren erregelarekin konbinatu behar baita. $\emptyset$ karakterea maiuskulak

tratamendurako erabiltzen ari dira arrakasta handiz), bestelako abantailak ere badauzkate formalismo morfologikoaren funtsaren ikuspuntutik: batetik, lexikoko adierazpide arbitrarioak edo *diakritikoak* ekiditen dira forma kanonikoa bultzatuz, lexiko mailan karaktere berezirik egon gabe informazio morfologikoa zehatz baitaiteke; bestetik, tarteko adierazpideak eta erregela-multzo anitzak sekuentzialki konbinatuz, deskribapen morfologikoa erraz daiteke eta deskribapen-ahalmena handitu.

3.4.1 Deskribapena

Ekarpen honek lexikoan forma flexionatuak forma kanonikoekin lotzeko gaitasuna eskaintzen du. Adibidez, ingeleseko *better* eta *good* lotu daitezke, nahiz lexikoaren eta azalaren arteko distantzia handia izan. Distantzia handitze honi aurre egiteko tarteko egoerak onartzen dira lexikoko eta azaleko mailen artean, Kaplan & Kay-ren ideia zaharrak berreskuratuz. Horrez gain, informazio morfologikoa zuzenean integra daiteke lexikoan.



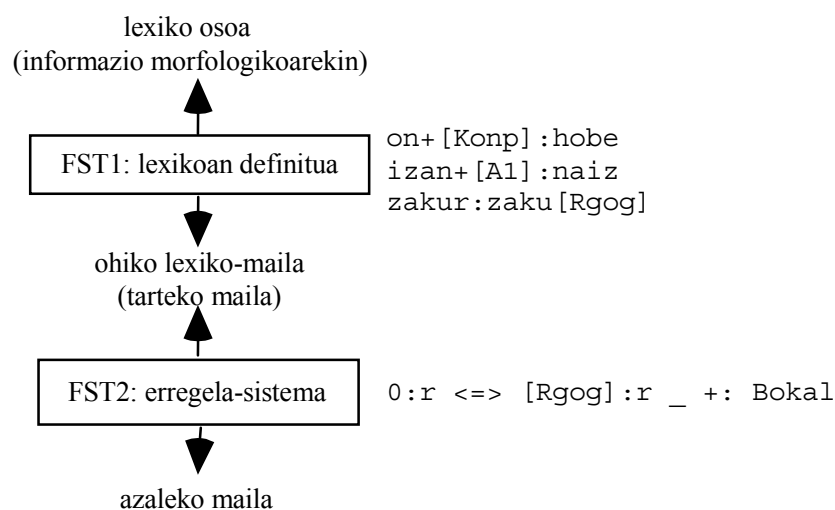
6. irudia.- Itzultzaileen arteko ebaketa eta konposaketa lexiko-itzultzaile bat lortzeko

Eredu honen funtsa ideia hauetan datza eta 6. irudian islatzen da:

- Normalean tarteko maila bat bereizten da lexikoaren eta azalaren artean, ohiko bi mailatako ereduan lexikokoa zena (adierazpide arbitrarioak, diakritikoak eta guzti eduki ditzakeena). Beraz, gutxienez 3 maila bereiziko dira: azalekoa, ohiko lexikoarena (tarteko maila izango da orain) eta lexiko maila osoa, informazio morfologiko guztiarekin. Lexikoko mailan informazio morfologikoa sartzen da, baina ez lotutako fitxategi baten bitartez, grafoko arkuetan baizik.

- Azala eta ohiko lexiko mailaren arteko aldaketak bi mailatako morfologiari dagozkion egoera finituetako itzultzaileen bidez gobernatzen dira. Horretarako *twolc* konpiladorea erabil daiteke.
- Lexikoko maila osoaren eta tarteko mailaren arteko bi mailatako erregelak idatzi beharko lirateke, itzultzaile paralelo eta bateragarriak sortzeko ohiko bidea baita. Erregela hauek asko eta oso lokalak liratekeenez, horren ordez *lexc* konpiladoreak bikoteak onartzen ditu, bi maila horien arteko parekatzea definitzen dutenak. *Be+Pres+Sg+P3+Verb:is* da horren adibide bat, *is* tarteko mailako karaktere-katea dela. Bihurketa hauei dagozkien automaten edo grafoen eraketaz *lexc* konpiladorea arduratzen da.
- Erregela-multzoen sekuentziak dakarren tarteko hizkuntzaren arazoaren aurrean, itzultzaileen arteko konposaketa konpilazio garaian burutzen da, itzultzaile bakar bat lortuz. Honetaz *lexc* konpiladorea arduratuko da, lexikoko mailaren eta aurretik zegoen itzultzailearen arteko konposaketaren bidez sistema osoa itzultzaile bakar batean uzteko. Hau funtsezkoa da eraginkortasunari begira.

Diseinu hori praktikan jartzean, espero zutena baino askoz ere emaitza hobeak lortu zituzten. Erregelen ebaketa eta konposaketaren ondorioz lortzen den egoera-kopuruaren magnitude-ordena lau edo bost zifra hamartarrekoa da, teorian izan zitezkeen hamabi zifretakoetatik oso urrun. Lexikoak murrizten omen du konplexutasuna. Bestalde, lexikoarekin konposaketa burutu ondoren egoera- zein arku-kopurua laburtzen duen optimizazioa egiten da.



7. irudia.- Lexiko-itzultzaileen ohiko 3 mailak

Lexiko-itzultzaileen ohiko mailak azaltzen dira 7. irudian (eskuinean adibide batzuk azaltzen dira). Bi automatak bakar batean konposatuko dira.

Lortutako emaitzak ikaragarriak dira. Frantseserako, adibidez, datu hauek ematen dituzte: 50 K-egoera eta 100 K-arku inguru, orotara Mega bat baino gutxiago hartzen duena eta zenbait mila hitz/segundo abiadura-ordena analisisian, PC-KIMMO baina ehunka aldiz azkarragoa.

3.4.2 Adibideak

Aurreko pasartean ikusitako adibideak, lexiko-itzultzaileen ezaugarriari esker aberastu eta argi daitezke. 3.7 adibidean azalduko dira lexikoan egin beharreko aldaketak, eta 3.8an zein 3.9an erregela-sistemarenekoak. Lexikoan egiten dira aldaketa gehienak, orain posible baita bi maila azaltzea: ezkerrean lema eta informazio morfologikoa (diakritikorik zein alomorforik gabe), eta eskuinean erregelekin arituko den lexiko-maila, azaletik gertuago, batzuetan diakritikoak edo erregelak aplikatzeko ezinbestekoak diren ezaugarri morfologikoak edukiko dituztenak.

3.7 adibidea: lexikoa euskaraz

Azpilexikoetan eta jarraitze-klaseetan ez da aldaketarik egongo, baina morfemetan bai: lemetan, ezkerreko aldean kategoria gehituko da eta, aurrizki/atzikietan, ezkerrean emaitzaren informazio morfologikoa baino ez da agertuko eta eskuinean morfema bera eta dagokion diakritikoa eta hautapen-marka sinbolikoa.

```

LEXICON Root
  IZENAK;
  ADJEKTIBOAK;

LEXICON IZENAK
ate[IZE]:ate          IZEN_JK;
etxe[IZE]:etxe        IZEN_JK;
ur[IZE]:ur            IZEN_JK;
zakur[IZE]:zaku[Rgog] IZEN_JK;
...

LEXICON ADJEKTIBOAK
gorri[ADJ]:gorri      ADJ_JK;
mozkor[ADJ]:mozko[Rgog] ADJ_JK
...

LEXICON DEKLINAB
+ [ABS] [MG]:      #;
+ [ABS] [MS]:+a    #;
+ [GEN] [MP]:+en   #;
...

LEXICON GRADUA
+ [KONP]:+ago      DEKLINAB;
+ [SUPE]:+en       DEKLINAB;

```

Makoen artean ('[' eta ']') agertzen diren datuak informazio morfologikoa dira. Horrela, *gorrien* hitza analizatzean, bi analisi posible lortuko dira: *gorri[ADJ]+[GEN]/[MP]* eta *gorri[ADJ]+[SUPE]*. *r* gogorrari dagokionez, berriz, ezker aldean ez da ezer azalduko (*zakur* edo *mozkor* zehaztuko da beren kategoriarekin batera), baina eskuinaldean orain bi aukera daude: lehen erabiltzen zen diakritiko bera, *R*, edo adibidean zehaztu den *[Rgog]* ezaugarria. Bigarren aukera honekin irakurgarritasuna irabazten da, baina erregelak egokitu behar dira 3.8 adibidean azaltzen denaren arabera.

3.8 adibidea: erregela ezaugarri morfologikoak erabiliz

Lexikoan diakritikoak baztertzen baldin badira, aldaketak egin beharko dira lehen aztertutako erregeletan. Karaktere anitzeko sinbolo gisa definituko dira lexikoaren deskribapenean, eskuinaldean, zehaztutako ezaugarri morfologiko desberdinak, tartean *[Rgog]* sinboloa, eta erregeletan ordezkatu.

Beraz, 3.3 pasarteko erregelak gutxi aldatu beharko lirateke. Bigarrenaren kasuan, esaterako, jatorrizkoa eta baliokidea hala azalduko dira:

```
0:r <=> R:r _ +:0 Bokal ;
          ! zakuR+a:zakurra

0:r <=> [Rgog]:r _ +:0 Bokal ;
          ! zaku[Rgog]+a:zakurra
```

Lexiko-itzultzaileek lexikoa eta erregela-multzoa argitzeaz eta osatzeaz gain, erregela-sistema sekuentzial anitzak bideratzen dituztenez, aipatutako 3 mailak lau, bost edo nahi dugun kopurura eraman daitezke. Honetaz balia gaitzke hainbat hobekuntza egiteko, eta horietako bat morfotaktika ez-jarraituaren kontrola izan daiteke. Kasu honetan, Bessleyk eta Karttunenek (2000) erregela-sistema bat gomendatzen dute lexikoa eta erregela morfofonologikoen artean. Tarteko sistema honetan fonologiarako pentsatutako erregelak morfotaktika murrizteko erabiltzen dira debeku-erregelak sortuz.

3.9 adibidea: urruneko mendekotasunen kontrola euskaraz

Adibide honetan *bait-+aditza+-lako* katea debekatzeko erregelaren bi bertsio agertzen dira. Kontuan hartu behar da erregela hau lexikoaren eta erregela morfofonologikoen artean dagoela; beraz, bi mailetako alfabetoa bera da, ez delako itzulpena egin behar. Ondoko adibidean bi bertsio proposatzen dira: lehenengoan aurrizki/atzizki kausalak ez dira lexikoan markatzen; bigarrenean, berriz, bai. Beraz, bigarren kasuan *bait[Kaus]+ eta +lako[Kaus]* izango dira lexikoko sarrerak, eta *[Kaus]* karaktere anitzeko sinboloa.

```
b:b /<= # _ a i t %+ [Bokal|Konts]+ %+ l a k o ;
%[Kaus%]:0 /<= _ [Bokal|Konts|%]+ % [Kaus%]: ;
```

Kasu honetan *%+* (plus karakterea) eta *+* (espresio erregularretako eragilea) bereizi dira, eta hau egin behar zen aurreko adibideetan ere (horietan *+* jarri beharrean *%+* jarri behar da) baina sinplifikatzeko ez da egin. Hurrengo kapituluaz azalduko denez, *%* karakterea izango da ihes-karakterea, eta espresio erregularretan esanahi berezia duten karaktereen aurrean jarri beharko da karaktere gisa (eta ez eragile gisa) erabili nahi direnean. Horixe gertatzen da [eta] karaktereeekin, ezaugarrietan erabili ahal izateko.

3.5 Sistema erabilgarriak

Bi mailtako morfologia erabili ahal izateko programa bat behar da, baina gomendagarria da eginda dauden tresnak erabiltzea norbera programatzen aritu gabe. Halako tresnetan bi osagai nagusi daude: erregela-multzoak edota lexiko-sistemak

automata bihurtzen dituzten konpiladoreak batetik, eta hitzen analisi/sorkuntza morfologikoa egiten duten programa eta errutina-multzoa bestetik. Programa nagusia edukitzea interesgarriagoa da probak egiteko, baina errutina-multzoa (*run-time library* izenarekin ezagutzen dena) askoz malguagoa da, halako liburutegi batek gure programetan morfologia integratzea bideratzen baitu.

Eskuragarri dauden sistema interesgarrienak hiru dira: *PC-KIMMO*, *mmorph* eta *LinguistX*. Hirurak libre erabil daitezke irakaskuntza zein ikerkuntza mailan, baina aplikazio komertzial bat garatu nahi izanez gero, zenbait kasutan, arazoak daude.

3.5.1 PC-KIMMO

Sistema zaharrena da eta, analisiaren abiaduraren aldetik motela bada ere, oso interesgarria da (Antworth, 1990). Hasiera batean ez zuen konpiladorerik, baina gaur egun eskuragarri ditu osagai guztiak: erregelen konpiladorea (*Kgen* izeneko programa dago horretarako), *PC-KIMMO* programa probak egiteko, eta liburutegia beste programetan integratzeko. Horrez gain, PATR formalismoa, sintaxirako oso interesgarria dena, eskaintzen da pakete berean.

Ez dituenex lexiko-itzultzaileen teknikak erabiltzen, analisiak poliki burutzen dira (gutxi batzuk segundoko) eta diakritikoen erabilera ezinbestekoa da. Tresna honen bidez hizkuntza askotarako prozesadore morfologikoak garatu dira modu esperimentalean.

Internet bidez dago eskuragarri helbide honetan:

<http://www.sil.org/pckimmo/v2/executables.html>

Tresnarekin batera hizkuntza batzuen deskribapen morfologikoaren adibideak ematen dira, eta ingelesaren kasuan deskribapen zabala.

3.5.2 mmorph

Europa mailako *Multext* proiektu linguistikoaren emaitza da *mmorph* (<http://www.issco.unige.ch/projects/MULTEXT.html>). Beste sistemekin konparatuz, badu eragozpen bat: konpilazio garaian forma guztiak lortzen dira; beraz, analisia zerrenda baten bilaketa baino ez da. Honek analisi-abiadura handia dakar (*hash-tauletan*¹⁹ oinarritutako teknikak erabiltzen dira), baina mugatua da euskara bezalako

¹⁹ Hash-taulak: bilaketa-abiadura handitzeko erabiltzen diren datu-egitura bereziak

flexio handiko hizkuntzetarako, forma gehiegi edo infinitu sor baitaitezke halako hizkuntzatan.

Abantailak ere baditu, batez ere lizentzia librearen eskuragarritasuna. Linuxeko *Debian* banaketan lizentzia libre bat dago eta helbide honetan eskura daiteke:

<http://packages.debian.org/stable/misc/mmorph.html>

Hizkuntza askotarako sistema errealak egin dira tresna honetan oinarriturik, tartean Europako Ekialdeko hainbat hizkuntzaren analizatzailak eta Espainiako Akademiako CREA²⁰ corpusean erabiltzen den analizatzaila morfologikoa (Sánchez León, 1997).

3.5.3 LinguistX

Xeroxen garatutako tresna-multzo hau ahaltsuena eta osoena da besteekin konparatuz, lexiko-itzultzaileak barneratzen dituelako. Horrez gain, harrigarria da emaitzen trinkotasuna eta analisiaren abiadura. Baina duela gutxi arte eragozpen bat zegoen: lizentzia lortzea ez zen erraza.

Inxight enpresaren bitartez merkaturatu du *Xeroxek* tresna hauekin garatutako hainbat prozesadore linguistiko inguruko hizkuntza hedatuenentzat, eta tresna horren bitartez garatu du IXA taldeak euskararako prozesadore morfologikoaren 2. bertsioa (lehen IXA taldeak berak egindako programan oinarritzen zen). Antzeko tresnen bitartez garatzen dituzte prozesadore morfologikoak ere arloko bi enpresa garrantzitsuk: *LingSoft*-ek eta *Conexor*-ek.

Gaur egun, lizentzien kontua aldatu da, eta Bessley eta Karttunen-en (2002) liburuarekin lor daitezke aipatutako tresnak irakaskuntzarako zein ikerkuntzarako, sistema desberdinetarako gainera: *Linux*, *Solaris* eta *Windows*-erako.

4. kapituluaren egingo da tresna honen deskribapen zehatza eta aplikazio praktikoa.

3.6 Laburpena

Bi mailatako morfologia izeneko formalismoaren deskribapen zehatza egin da kapitulu honetan, osatzen duten bi elementu nagusiak azalduz: morfemak eta morfotaktika deskribatzen dituen lexiko-sistema batetik, eta aldaketa fonologikoak zehazteko

²⁰ CREA: *Corpus de Referencia del Español Actual*.

erabiltzen diren bi mailatako erregelak bestetik. Formalismoaren gailurra diren lexiko-itzultzaileen aipamena ere burutu da.

Deskribapena sakontzeko adibide zehatzak gehitu dira, euskararenak gertutik ezagutzen direlako, eta beste hizkuntzenak fenomeno morfologiko anitzak azaldu ahal izateko. Dena den, hurrengo kapituluetan adibide gehiago eta osatuagoak azalduko dira. Horrez gain, eskuragarri dauden tresnen deskribapena eta sailkapena egin da.

4 LinguistX eta adibideak

4.1 Sarrera

Aurreko kapituluak aipatutako tresnak Xeroxen laborategietan garatu dira eta Bessley eta Karttunen-en (2002) liburuarekin lor daitezke irakaskuntza/ikerkuntzarako, sistema desberdinetarako gainera: *Linux*, *Solaris* eta *Windows*-erako. Liburu horretan deskribatzen diren funtsezko tresnak hiru izan arren, hemen bi besterik ez dira azalduko: *lexc* lexikoaren eta morfotaktikaren deskribapenerako, eta *twolc* erregela fonologikoetarako (hirugarrena, *xfst*, orokorragoa da eta espresio erregularretako tratamendura zuzenduta dago). *lexc* lexiko eta erregela fonologikoak zein itzultzaile-maila anitzak konposatzeko ere baliagarri denez, *twolc* azalduko da aurrena.

Tresna bakoitzeko bi atal nagusi egongo dira: editatu behar den fitxategiaren formatua deskribatzen duen *sintaxia* izeneko atala batetik, eta tresnaren aukera nagusiak azaltzen dituen *oinarrizko komandoak* izeneko bestetik. Horren ondoren adibide oso baten definizioa, konpilazioa eta proba zein hainbat metodologia-aspektu azalduko dira.

Oinarritzko tresnez gain, beste bi programa daude, *tokenizer* eta *lookup* izenekoak, probak eta ustiapena burutu ahal izateko.

Aurreko kapituluak azalduko adibideak programa hauei egokitzen zaizkie gutxi gorabehera, baina benetan ibil daitezten aldaketatxo batzuk beharko lirateke. Hemendik aurrera zehaztuko diren adibideak konpilatzeko moduan daude (eta probatuak izan dira kasu askotan).

4.2 *twolc* programa

Erregela morfofonologikoak konpilatzeko eta probatzeko programa da. Bere sintaxia aztertu aurretik, bertan erabiltzen diren espresio erregularren nondik norakoak azalduko dira.

4.2.1 Espresio erregularrak

Espresio erregularrak bi mailatako erregelako testuinguruan zein definizioetan erabil daitezke. Espresio erregular hauen funtsa ezaguntzat hartzen bada ere, sistematik

sistemara hainbat ñabardura dago, eta horiek azalduko dira sistema honen sintaxiaren errepaso azkar bat egitearekin batera.

Karaktere bereziak

Badaude zenbait sinbolo karaktere bereziak edo multzo bereziak adierazteko. Hona azpimarragarrienak:

kar.	azalpena	adibidea	interpretazioa
0	<i>epsilon</i> karakterea edo karaktere hutsa	$r : 0$	r lexikokoa desagertu egiten da azalean
?	ezagututako karaktere-pare guztien bildura	$? - r : 0$	pare posible guztiak aurreko adibidearena kenduta
=	maila bateko karaktere guztiak	$r : =$	r karakterearekin osa daitezkeen bikoteak r : baliokidea da
%	ihes-karakterea, esanahi berezia duten karaktereei aplikatzeko ²¹	% % % ? % +	% karakterea ? karakterea + karakterea
#	hitz-hasiera ezkerreko testuinguruan eta bukaera eskuinekoan.	: a # # e : z :	a-z bukatutako hitza ez hasiera lexiko mailan

Multzoak eta segidak

Segida baten karaktereak edo bikoteak zuriz banatu behar dira, karaktere jarraituak topatzean programak karaktere anitzeko sinbolo baten aurrean dagoela suposatzen baitu. Kontu handiz zehaztu behar dira hitz-zatiak zein morfemak, oso errore arrunta baita jarraian idaztea.

Beste aldetik, espresio konplexuagoak idatzi behar direnean, karaktereak multzokatzea gomendagarria edo ezinbestekoa da, eta horretarako makoak ([] karaktereak) erabili ohi dira, giltzak erabiltzea ere posible den arren.

Azkenik, segida baten barruan hautazko elementu bat adierazteko parentesiak erabil daitezke.

Ondoko taulan hainbat adibide azaltzen da:

²¹ Ulermena erraztearren, aurreko kapituluaren adibide gehienetan ez da erabili.

adibidea	interpretazioa
bokal	<i>bokal</i> sinboloa
b o k a l [b o k a l]	<i>bokal</i> hitz-zatia azalean eta lexikoan. aurrekoaren baliokidea.
[k : g k] o [k : g o k o]	ondoren ikusiko denez <i>ko:go</i> zein <i>ko:ko</i> bikoteak adierazten dute biek
e (d) u k i	<i>eduki</i> zein <i>euki</i> segidak

Eragileak

Espresio erregularren funtsa eragileak²² dira, eta hauek dira erabilienak:

eragilea	adibidea	interpretazioa
kateaketa: <i>zuriunea</i>	k : g o	<i>k:g</i> bikotea <i>o:o</i> bikoteaz jarraituta
bildura:	k : g k : k	lexikoko <i>k</i> , azaleko <i>g</i> edo <i>k</i> -rekin
errepikapena: *, +	[V : V] * [% + :] +	bokala 0, 1 edo n aldiz lexikoko + karakterea behin edo gehiagotan
kenketa: -	C - h	edozein kontsonante <i>h</i> izan ezik
bikote osagarria: \	\V ? - V	bokal-bikotea ez den edozein bikote baliokidea
segida osagarria: ~	~V ? * - V	bokal-bikotea ez den edozein sekuentzia baliokidea
ahaztea: /	V+ / h	bokal-segida <i>h</i> -ak kontuan hartu gabe
ebakidura: &	A&B	bi multzoetan dauden bikoteak
edukitzailea: \$	\$V ? * V ? *	bokal-pare bat duen edozein segida baliokidea

Kontu handiz erabili behar dira eragileak eta, zalantza kasuan, makoak erabili espresioak multzokatzeko.

Argi eduki behar da erregelen testuinguruan zatiak zehazten direla, eta ez hitz osoaren karaktereak. Beraz, ondoko bi erregelak baliokideak dira eta lehena da erabiltzen dena:

a : b <=> l _ r
a : b <=> ? * l _ r ? *

Horren ondorioz, erregetan hitz osoak adierazi nahi direnean, hitz-mugaren marka (#) erabiliko da ezkerreko zein eskuineko testuinguruan.

²² Gogoratu karaktere horiek lexiko edo azaleko mailan erabiltzeko '%' ihes-karakterearen ondoan jarri behar direla.

4.2.2 Sintaxia

Editatu beharreko fitxategian hainbat atal daude eta bakoitzak ingelesezko hitz bat du burukotzat: *Alphabet*, *Diacritics*, *Sets*, *Definitions* eta *Rules*. Banan-banan aztertuko ditugu, kontuan hartuta atal guztietan sententzia bakoitzaren ondoren ; karakterea zehaztu behar dela, baina burukoaren ondoren ez. Oharrak edo iruzkinak gehi daitezke ! karaktere erreserbatuaren ondoren, karaktere horretatik lerro-bukaera arte dagoena konpiladoreak ez duelako kontuan hartuko.

Alphabet

Alfabetoko bikote konkretu zuzenak (lexiko-mailakoa, ‘:’ karakterea eta azalekoa) zehazten dira bertan, zuriunez banatuta. Ohikoa da karaktere soil bat zehaztea bikotearen ordean, lexiko-maila eta azaleko maila berdina dela adieraziz. Karaktere arruntez gain, diakritikoak zein ezaugarri morfologikoak adierazten dituzten sinboloak zehatz daitezke. Espresio erregularretan esanahi berezia duten karaktereei (ikus aurrerago) % ihes-karakterea jarriko zaie aurretik.

4.1 adibidea: euskarazko alfabetoaren definizioa

Demagun, alfabeto osoaz gain, lexiko-mailan *r* gogorra adierazten duen *[Rgog]* ezaugarria eta ‘+’ morfema-muga dituen alfabeto bat zehaztu nahi ditugula.

Alphabet a b c d e f g h i j k l m n ñ o p q r s t u ü v w x y z %+:0 %[Rgog%]:r ;
--

lexikoko + desagertu eta *[Rgog]* ezaugarria *r* karakterean gauzatuko da azalean besterik esaten ez den bitartean. Gainontzeko karaktereetan ez da aldaketarik egongo erregelaren batean kontrakoa esaten ez bada.

Diacritics

Diakritikoak (aurreko orrietan hautapen-marka edo ezaugarri morfologiko izenarekin ere aipatu dira) alfabetoan defini daitezkeenez, ez ditugu atal honetan definitu beharrik. Bi baldintza hauek betetzen dituztenean bakarrik gehituko dira hemen:

- azaleko karaktere baliokiderik ez dutenean. Beraz, *R* edo *[Rgog]* marka ez da atal honetan jarriko, azalean *r* karakterean gauzatzen baita.
- erregelen testuinguruan aipatzen ez direnean, lexiko-mailan egongo ez balira bezala jokatu da. Morfema-muga adierazten duen + lexikoko karakterea ere ez da egokia atal honetarako, testuinguruetan garrantzi handia du-eta.

Orduan, noiz erabiliko da atal hau? Beste erregelatan eraginik ez duten, eta soilik erregela baten edo biren aplikaziorako sortu diren marketarako batez ere.

4.2 adibidea: diakritikoen atala

3.4 adibidean agertutako toki-izenaren marka, T batez edo *[Toki-Izen]* sinboloaz adierazia, diakritikoen atal honetan ager daiteke, bi baldintzak bete ditzakeelako: ez da azalean gauzatzen batetik, eta testuinguruan ez da, hasiera batean behintzat, kontuan hartu behar, ez bada r -ren galerrarako eta k/g bihurketarako.

Diacritics

T ; ! toki-izenen diakritikoa

Ikusten denez, ez da karaktere-bikote moduan azaltzen, pentsa baitaiteke ez dela azalean gauzatuko. Beraz, horrela azaltzeak, alfabetoan $T:0$ moduan azaltzearekin duen aldea hau da: kasu honetan $T:0$ azaltzen ez diren testuinguruetan toki-izenek testuingurua bete dezakete, nahiz eta marka hau eduki. Adibidez, kontsonantez edo bokalez bukatzen diren ala ez egiaztatzeko, T ez da kontuan hartuko.

Sets

Ezaugarri morfofonologiko komunak dituzten karaktereak (azalekoak edo lexikokoak) multzokatzeko balio du, horrela erregelen testuingurua sinpleagoa eta irakurgarriagoa izan dadin.

4.3 adibidea: euskarazko zenbait multzo

Kontsonanteak, bokalak, bokal irekiak, txistukariak etab. multzokatu ohi dira sistema askotan, portaera morfofonologiko komunak dituzte-eta.

Sets

Konts = b c d f g h j k l m n ñ p q r s t v w x y z ;
 KontsHerts = b c d f g j k l m n ñ p q r s t v w x z ;
 Bokal = a e i o u ü ;
 BokalIreki = a e o ;
 Txistu = s x z ;

KontsHerts izeneko multzo bat zehaztu da, h eta y karaktereek zenbait erregelatarako ez baitute kontsonanteen jokabidea izango.

Ikusten denez, esleipen batez definitzen dira multzoak: multzoaren izena eta '=' karaktere erreserbatua ezkerrean jarri ondoren, multzoko karaktereak zuriunez bereizirik zehaztuz.

Definitions

Identifikatzaile bat duten espresio erregularrak idazteko aukera eskaintzen du atal honek, erregelen testuingurua sinpleagoa eta irakurgarriagoa izan dadin. Testuinguru desberdinetan erabiltzen diren espresioak sinplifikatzeko erabili ohi dira eta, bertan, aurreko ataleko multzoak ere erabil daitezke.

4.4 adibidea: zenbait definizio euskararako

ts, *tx* eta *tz* afrikatuek portaera bera dute zenbait fenomenotan, baina ezin dira multzoetan sartu, bertan karaktere-bikote sinpleen multzoak baino ezin direlako definitu. Adibide honetan, silabaren definizio sinplifikatu bat eta mugaren definizioa ere egiten dira.

Definitions	
Afrik = t Txis ;	! ts, tx eta tz segidak
Silaba = Konts+ Bokal+ ;	! sinplifikazio bat da
Muga = %+: # ;	! morfema-muga edo hasiera zein bukaera

Multzoen definizioarekin antz handia du, baina atal honetan espresio erregularretako eragileak erabil daitezke eta multzoetan ez, horietan bildura eragilea suposatzen baita inplizituki eta, hauetan berriz, segida.

Rules

Funtsezko atala da erregelena, gainontzeko guztiak honen prestakuntza besterik ez dira-eta.

Erregelak idazteko jarraitzen den semantika 3. kapituluaz azaldu da; sintaxia definitzeko, berriz, kapitulu horretan azaldutakoari *twolc* programaren zehaztasunak gehitu behar zaizkio.

Erregelak idazteko 3. kapituluaz definitutako elementuez gain (korrespondentzia, erregela-mota eta testuingurua), identifikatzaile bat jartzen zaio aurretik komatxo bikoitzen artean ("e epentetikoa", esaterako). Erregela bakoitzaren ondoren oharren bidez adibideak zehaztea oso gomendagarria da, betiere lexiko-maila eta azalekoa bereiziz.

4.5 adibidea: zenbait erregela euskararako

3. kapituluaz euskararako azaldu ditugun erregela zuzenak aldatuko ditugu hona, *twolc* programak eskatzen duen sintaxira egokiturik (egia esan, oso aldaketa gutxi behar da).

"r gogorra"	
0:r <=> % [Rgog%]:r _ %+: Bokal ; ²³	
! be [Rgog]+egin:berregin	
! zaku [Rgog]+a:zakurra	
! ur+a:ura	
"e/i aldaketa ber aurritzian"	
e:i <=> # b _ % [Rgog%]:r %+: KontsHerts ;	
! be [Rgog]+findu:birfindu	
! be [Rgog]+egin:berregin	

²³ % ihes-karakere erabili behar da makoekin informazio morfofonologikoaren partea drienean, bestela, espresio erregularretako esanahia hartzen baitute makoek. Orain arte ez da horrela azaldu adibideak sinplifikatzearen. Adibideetan, berriz, ez dira ihes-karakereak erabili behar, adibideak konpilatzen ez direlako.

```
"h-ren galera"
h:0 <=> b e %[Rgog%]:r %+: _ Bokal ;
        ! be[Rgog]+has:berras

"r-ren galera"
r:0 => z T:0 %+: _ a # ;
        ! 9gasteizT+ra:9gasteiza
```

Ez da mota bereko erregela bat baino gehiago idatziko korrespondentzia bakoitzeko. Testuinguru bat baino gehiago jar daiteke korrespondentzia horretarako, baina bakoitzaren bukaeran ‘;’ karakterea zehaztu beharko da. Testuinguru bat baino gehiago zehazteak erregela testuinguru desberdinetan aplikatzea dakar.

4.6 adibidea: testuinguru bat baino gehiago

r-ren sorrera gerta daiteke *hiru* edo *lau* zenbakiak deklinatzean. Sinplifikazio bat eginda, pentsa dezagun *r* sortu behar dela bokal baten aurrean, lema hauen deklinabidean. Erregela horrela geratuko da:

```
"r gogorra"
0:r <=> %[Rgog%]:r _ %+: Bokal ;
        # [h i r u | l a u] _ %+: :Bokal ;
        ! lau+ak:laurak
        ! zaku[Rgog]+a:zakurra
```

jarri da ezkerreko testuinguruan *hamalau* bezalakoetan sorrera hau ez gertatzeko. *Bokal* jarri beharrean *:Bokal* zehaztu da bokal hori epentetikoa izan daitekeelako. Dena den, arazo bat dago: *lau* adjektiboarekin betetzen da erregela, baina ez da sortu behar. Zein da irtenbidea? Diakritiko edo ezaugarri morfologiko bat gehitzea, seguruena zenbakietan (salbuespena direlako). Lexikoan *[Rzah]* sinboloa jartzen badiogu (*hiru[Rzah]*), aldaketa hauek egin ditzakegu:

```
Diacritics
T %[Rzah%] ;
...

Rules

"r gogorra"
0:r <=> %[Rgog%]:r _ %+: Bokal ;
        %[Rzah%]: _ %+: :Bokal ;
        ! lau[ZenbZah]+ak:laurak
        ! be[Rgog]+egin:berregin
        ! zaku[Rgog]+a:zakurra
```

Ikusten denez, aukera honekin lexikoa konplexuagoa izango da, baina erregelaren testuingurua sinplifikatzen da, marka egotea nahikoa baita aldaketa mugatzeko.

Korrespondentzia gehienetan bikote konkretuak zehazten dira, hau da, karaktere konkretuekin edo karaktere hutsarekin egindako ohiko bikoteak. Dena den, bikote abstraktuak bideratzen dituzten korrespondentzia konplexuak ere zehatz daitezke aldagaiak erabiliz.

Aldagaiak, *Cx* eta *Cy* izenarekin erabili ohi direnak, zehatz daitezke korrespondentzia-bikotean. Aldagaien balioak adierazteko, erregelari dagokion testuingurua zehaztu ondoren, *where* motako sententzia bat osatzen da, aldagaiak edo bi aldagaiek har ditzaketen balioekin. Bi aldagai direnean, korrespondentzia bi motakoa izan daiteke:

banan-banan zehazten diren ordenan, edo maila bateko bakoitza beste mailako guztiekin. Lehen kasuan *matched* zehaztu beharko da sententziaren bukaeran.

4.7 adibidea: aldagaien erabilera

Pentsa dezagun hitz ez-estandarrek ezagutu nahi direla. Adibidez, *b-d-g/p-t-k* motako aldaketak, hitz-hasieran eta bokalaren aurrean. Idatzi beharreko erregela hau litzateke:

```
"HersAhosk / HerAhost (b-d-g/p-t-k) "
Cx:Cy => # _ Bokal ;
        where Cx in (b d g)
              Cy in (p t k)
        matched ;
        ! bake:pake
```

matched zehaztu denez *b* ordez *p* onartuko da, baina ez *t* edo *k*, *d* ordez *t*, baina ez *p* edo *k*; eta, *g* ordez, *k*. Parentesi artean agertzen dena multzo baten identifikatzaileak ordezka dezake.

Aipatutako sintaxia betetzen duen fitxategi bat editatu ondoren, konpilatzera eta probatzera pasa gaitezke, horretarako *twolc* programa erabiliz.

4.2.3 Oinarrizko komandoak

Komando anitz eskaintzen digu *twolc* programak, baina gutxi batzuk ezagutzea nahikoa da programari etekin handia ateratzeko.

Hona komando nagusiak eta dagokien esanahia:

komandoak	eragina
? <i>help komandoa</i>	aukeran dauden komandoen listatua. <i>komandoari</i> dagokion azalpena.
<i>read-grammar fitx</i>	editatutako fitxategi baten erregelak irakurtzea. Fitxategiaren mota <i>rul</i> edo <i>txt</i> izan ohi da. Onartzen dituen atal eta erregelak listatzen ditu.
<i>compile</i>	irakurritako erregelak konpilatzea. Mezuren bat aterako du erregelaren bat konpilatzeko arazorik baldin badu.
<i>redo</i>	azken <i>read-grammar</i> eta <i>compile</i> komandoak errepikatzea. Detektatutako errore baten ondoren erregelak aldatutakoan egiten da.
<i>save-binary fitx</i>	konpilatutako erregelei dagokien itzultzailea formatu bitarrean gordetzea geroxeago erabiltzeko. Fitxategiaren mota <i>fst</i> izan ohi da.
<i>save-tabular fitx</i>	konpilatutako erregelei dagozkien automatak gordetzea formatu sinbolikoan. Fitxategiaren mota <i>aut</i> edo <i>tab</i> izan ohi da.
<i>list-rules</i>	konpilatutako erregelen izena eta dagozkien automaten tamainen listatua.

lex-test	probak egiteko erabiltzen da. Ondoren lexiko-mailako kate bat sakatu eta dagozkion azaleko formak listatzen ditu. Funtzionamendu zuzena egiaztatzeko eta gainsorkuntza ekiditeko egokia.
pair-test	Egiaztatzeko, baina <i>lex-test</i> baino gutxiago erabiltzen da, bi mailak zehaztu behar baitira.
quit	bukatzea eta programatik irtetea.

Erregela-multzo bat konpilatzeko eta probatzeko jarraitu ohi den prozedura hau da (demagun *erregl.txt* dela erregelen fitxategia):

1. read-grammar erregl.txt
2. compile
3. lex-test zaku[Rgog] +a²⁴

Nahi adina proba egin daiteke. Urratsen batean erroreren bat detektatzen bada, fitxategian aldaketak egin beharko dira editorearen laguntzaz. Interfaze grafikoarekin ari bagara, aldaketak beste leihoren batean egingo dira eta programatik irten gabe horrela jarraituko da (bestela aurreko urratsak errepikatu beharko lirateke):

4. redo
5. lex-test be[Rgog] +has

Bukatzeko bi urrats falta dira: konpilatutako sistema gorde eta programatik irten:

6. save-binary erregl.fst
7. quit

4.2.4 Erregelen arteko eraginak eta gatazkak

Erregelak banan-banan idazten dira, baina erlazionatuta daude haien artean. Ez dago ordenarik ez tarteko lengoaiarik haien artean, 3. kapituluaren esanenez, erregela paraleloak baitira, baina elkarren artean eragina dute: erregela batean aurreikusten den bikote berriak beste erregela baten testuinguruarekin zerikusia badu, erregela berria ondo egon arren, analisis/sorkuntza ez da gauzatuko erlazionatutako erregela ukitzen ez badugu.

4.8 adibidea: erregelen arteko eragina

3.5 adibidean gaztelaniarako azaldutako erregelak dira honen adibide. Pentsa dezagun *z:c* erregela idazten dugula lehenengo, pentsatuz plurala *es* morfemaz osatuko dela.

```
"z:c erregela"
z:c <=> _ %+:0 e s # ;
          ! pez+es:peces
```

Ondoren erabakitzen dugu ezetz, pluralaren forma kanonikoa *s* dela eta horrela behar dela bokal-amaierarako. Ondorioz, *e* sortzeko erregela idatzi behar dugu:

²⁴ Edo beste edozein kate lexiko-mailan

```
"e epentetikoa"
0:e <=> Konts %+:0 _ s # ;
      ! pez+s:peces
      ! casa+s:casas
```

Konts-ekin ez dago arazorik, *z* eta *c* kontsonanteak dira-eta; dena den, agian hobe litzateke ohitzea alde bakar bat zehazten. Ezinbestez aldatu behar dena lehenengo erregela da, bertan *e* zehazten baita lexiko zein azaleko mailan, eta, aipaturiko aldaketa dela-eta, lexikoko mailan ez da egongo. Ondorioz, multzo hau litzateke zilegi:

```
"z:c erregela"
z:c <=> _ %+:0 :e s # ;

"e epentetikoa"
0:e <=> Konts: %+:0 _ s # ;
```

e agertu beharrean :*e* agertzen da azalean *e* bat behar dela adierazteko. *Konts*: jarri da aipaturiko gomendioa betetzeko.

Badago beste arazo bat, detektatzen zailago dena, baina konpiladoreari esker detektatuko dena: korrespondentzien arteko gatazka, hain zuzen. Hau <= motako erregelatan gertatzen da eta, ondorioz, <=> motakoetan ere bai. Testuinguru berean aldaketa desberdinak behartzean gertatzen da arazo hau.

4.9 adibidea: erregelen arteko gatazka

Pentsa dezagun bi erregela hauek ditugula (adibide hutsa da, morfema-muga ez da erabili kasu honetan):

```
"k:0 erregela"
k:0 <=> _ Konts ;
      ! horiektan:horietan

"k:g erregela"
k:g <=> Konts _ ;
      ! eginko:egingo
```

Ez dirudi arazorik dagoenik, baina testuinguru batean *kontsonante k kontsonante* gertatzen bada, indeterminazio bat dago, ez baitago argi konpiladorerako *k*-ren bikotea. Testuinguru hori agian ez da gertatuko definitzen ari den lexikoa kontuan hartuta, baina konpiladoreak hori ez daki. Irtenbidea testuingurua gehiago zehaztea da; adibidez, bigarren erregelan bokala zehaztuz eskuineko testuinguruan.

```
"k:0 erregela"
k:0 <=> _ Konts ;
      ! horiektan:horietan

"k:g erregela"
k:g <=> Konts _ Bokal ;
      ! eginko:egingo
```

Konpiladoreak detektatzen ditu gatazka hauek, eta errorea denean zuzendu egin beharko da; errorea ez denean, berriz, testuinguru zehatzagoa definituko da horietako batean.

4.3 *lexc* programa

Programa honen helburu nagusia lexikoaren eta morfotaktikaren definizioa da, baina bigarren erabilera bat ere badu: itzultzaile desberdinak konposatzea. Lehen helburua lortzeko zehaztasunak lehen bi pasarteetan ematen dira, eta konposaketarena hirugarrenean.

4.3.1 Lexikoaren sintaxia

Bi mailatako morfologiaren barruan, lexikoaren definizioa nola egin behar den azaldu da 3.2 pasarteetan. *lexc* programa erabiliz bi aukera gehiago dago:

- lexiko-itzultzaileetan gaudenez, lexikoan bertan bi maila azal daitezke: analisien emaitza edo sorkuntzaren iturburua izan dadin lehena, informazio morfologiko osoa edukiko duena, eta bi mailatako erregelekin elkarreragiteko egokiena bigarrena. Beraz, morfema bakoitzak bi adierazpide eduki ditzake, bata bestearen ondoren ‘.’ karaktereaz bereizirik. Bereizgarririk ez baldin badago, bi mailak berdinak direla pentsatzen du *lexc* konpiladoreak.
- karaktere anitzeko sinboloak (*Multichar_Symbols*) defini daitezke, horrekin diakritikoak ezaugarri morfologiko bihurtzen baitira, baina emaitzaren trinkotasunean eragin kaltegarririk izan gabe. Ezaugarri morfologikoak erabil daitezke lexikoko bi mailetan.

twolc konpiladorearekin alderatuta, bi aldaketa dira adierazgarrienak *lexc* konpiladorean:

- karaktere jarraituak segidak dira, kontrakoa esaten ez den bitartean.
- espresio erregularrak ez dira erabiltzen; beraz, ez da hainbeste ihes-karaktere erabili behar.

Ondoko pasarteetan sintaxiaren zehaztasunak azalduko dira: erreserbatutako sinboloak aurretik, sintaxi arrunta ondoren eta *Multichar_Symbols* atala azkenik.

Karaktere eta sinbolo bereziak

Hitz eta karaktere batzuk erreserbatuak dira, konpiladoreak esanahi berezia ematen dielako.

karakterea/sinboloa	azalpena
#	Jarraitze-klase hutsa adierazten du. Atzizki gehiagorik ez dagoela adierazteko erabiltzen da.
%	lhes-karakterea, esanahi berezia duten karaktereekin erabiltzeko. Zuriune karakterea adierazteko % eta zuriunea zehaztuko da.
0	Morfema hutsa. %0 izango da 0 karakterea.
:	Mailen arteko bereizgarria morfemaren barruan.
!	Oharra hemendik lerro-bukaera arte.
LEXICON	Jarraitze-klase edo azpilexiko baten hasiera.
LEXICON Root	Grafoaren hasiera den azpilexikoaren hasiera.
Multichar_Symbols	Fitxategiaren hasieran jarriko da, ondoren ezaugarri morfologikoak deskribatzeko.

Sarrera arruntak

Lexiko-sarrerak azpilexikoetan biltzen dira. Azpilexiko bakoitzaren hasieran LEXICON hitz erreserbatua eta azpilexikoaren izena agertuko dira eta ondoren hainbat lerro. Lerro bakoitzaren bukaera adierazteko ; jarri behar da eta aurretik morfema eta jarraitze-klasea ager daitezke zuriuneraren baten bidez bereizirik:

osagaia	adibidea	azalpena
morfema		morfemaren adierazpena. Konplexuetan : karaktereak banatuko du goiko maila (emaitzarena) eta behekoa (erregelena).
morfema sinplea	etxe +a beR+	etxe lema a atzizki arrunta ber aurrizkia diakritiko batekin
morfema konplexua	+ABS : +a ber+ : be [Rgog] +	Absolutiboaren marka goiko mailan [Rgog] ezaugarria R diakritikoaren orde
jarraitze-klasea	GRADUA	azpilexiko baten izenarekin bat etorri behar du.

Osagai bakarra agertzen bada, jarraitze-klasea izango da, ezinbestekoa baita. Morfemarik agertzen ez bada, kate hutsa suposatzen da, eta haxe da jarraitze-klase berriak definitzeko modua, azpilexiko batean (jarraitze-klasearen izena izango duena) hainbat azpilexiko zehazten baitira morfemarik zehaztu gabe.

4.10 adibidea: jarraitze-klase baten definizioa

Adjektiboen jarraitze-klasea definitzeko ondoko hiru lerroak jartzen dira:

```
LEXICON ADJ_JK
GRADUA;
DEKLINAB;
```

GRADUA eta DEKLINAB azpilexikoak izango dira eta lerroon bitartez ADJ_JK jarraitze-klasea definituko da.

4.11 adibidea: lexiko labur bat euskaraz

Euskararen lexikoaren eta morfotaktikaren azpimultzo baten deskribapena honela gera daiteke programa honekin (3. irudian azaldu da morfotaktikaren grafoa):

```
LEXICON Root
  IZENAK;
  ADJEKTIBOAK;
  ADITZAK;
  ADITZ_AUR;

LEXICON IZEN_JK25
  DEKLINAB;

LEXICON ADJ_JK
  GRADUA;
  DEKLINAB;

LEXICON IZENAK
ate[IZE]:ate          IZEN_JK;
etxe[IZE]:etxe        IZEN_JK;
ur[IZE]:ur            IZEN_JK;
zakur[IZE]:zaku[Rgog] IZEN_JK;
...

LEXICON ADJEKTIBOAK
gorri[ADJ]:gorri      ADJ_JK;
mozkor[ADJ]:mozko[Rgog] ADJ_JK
...

LEXICON DEKLINAB
+[ABS][+MG]:0      #;
+[ABS][+MS]:+a     #;
+[GEN][+MP]:+en    #;
...

LEXICON GRADUA
+[KONP]:+ago       DEKLINAB;
+[SUPE]:+en        DEKLINAB;

LEXICON ADITZ_AUR
ber[AUR]:+be[Rgog] ADITZAK;

LEXICON GRADUA
egin[ADI]:egin     #;
fin[ADI]:fin        #;
has[ADI]:has        #;
...
```

Multichar_Symbols

Atal honen barruan, lexikoan zehazten diren ezaugarri morfofonologikoak zerrendatu egin behar dira fitxategiaren hasieran. Zerrenda ‘;’ karakterearekin bukatuko da. Horrekin hainbat helburu lortzen da: diakritikoak ekiditea, forma kanonikoak goiko mailan bultzatzea, eta automata edo grafo trinkoagoak osatzea.

²⁵ Ez litzateke beharko, IZEN_JK eta DEKLINAB baliokideak direla besterik ez baita esaten.

4.12 adibidea: karaktere anitzeko sinboloak

4.11 adibidean agertu beharreko erazagupena hau litzateke:

```
Multichar_Symbols
```

```
[IZE] [ADJ] [ADI] + [ABS] + [GEN] + [KONP] + [SUPE] + [MP] + [MG] [Rgog] ;
```

Aipatutako sintaxia kontuan hartzen duen fitxategi bat prestatu ondoren, *lexc* programarekin konpilatu ahal izango da.

4.3.2 Oinarrizko komandoak

twolc programan gertatzen zen bezala, *lexc* programa abiatu ondoren ere, hainbat eragiketa burutzeko aukera dugu komando desberdinen bitartez. *lexc* programak komando anitz dituen arren, gutxi batzuk ezagutzea nahikoa da programari etekin handia ateratzeko.

Hona komando nagusiak eta dagokien esanahia:

komandoak	eragina
? <i>help komandoa</i>	aukeran dauden komandoen listatua. <i>komandoari</i> dagokion azalpena.
<i>compile-source fitx</i>	editatutako fitxategi baten lexikoa konpilatzea. Fitxategiaren mota <i>lex</i> edo <i>txt</i> izan ohi da. Konpilatzen dituen azpilexikoen osagai-kopurua zerrendatzen du.
<i>save-source fitx</i>	konpilatutako lexikoari dagokion itzultzailea gordetzea, geroxeago erabiltzeko. Fitxategiaren mota <i>fst</i> izan ohi da.
<i>read-source fitx</i>	konpilatutako eta gordetako lexikoari dagokion itzultzailea irakurtzea.
<i>quit</i>	bukatzea eta programatik irtetea.

Lexikoa konpilatu ondoren, komando hauek erabil daitezke probak egiteko:

komandoak	eragina
<i>random</i>	lexikoko 15 forma sortzen ditu ausaz.
<i>lookup</i>	lexikoko behealdea emanda, goialdea eskaintzen du.
<i>lookdown</i>	lexikoko goialdea emanda, behealdea eskaintzen du.

4.3.3 Sistemen konposaketa

Aurretik esan dugun bezala, prozesadore morfologiko bat egiteko lexikoa eta erregela-multzo bat konbinatzen dira. *lexc* programarekin, lexikoa konpilatzeaz gain, konposaketa hori burutzen da, eta, ondorioz, automata bakar bat lortuko da. Horretarako hiru maila bereizten dira: lexikokoa (*source*), erregelena (*rules*) eta emaitzena (*result*).

Komando gehigarri batzuk behar dira konposaketa aurrera eramateko:

komandoak	eragina
<code>read-rules fitx</code>	<i>twolc</i> erabiliz konpilatutako erregela-multzoari dagokion itzultzailea irakurtzea.
<code>compose-result</code>	lexikoa eta erregela-multzoa konposatzea.
<code>save-binary fitx</code>	emaitza gordetzea.
<code>result-to-source</code>	emaitza lexikora pasatzea, beste erregela-multzo batekin konposatzeko. Erregela-multzo bat baino gehiago konposatzeko erabiliko da.

Erabiliko litzatekeen ohiko komando-segida bat izan liteke:

1. `compile-source lex1.txt`
2. `random`
3. `read-rules erreg1.fst`
4. `compose-result`
5. `random`
6. `save-result morfo1.fst`

Nahi adina proba egin daiteke *random* edo *lookup* zein *lookdown* komandoekin. Arazoak detektatzen baldin badira, berriro ekin beharko zaio prozedurari, lexikoa editatu ondoren edo, arazoa erregeletan baldin badago, erregelak editatu eta konpilatu ondoren. Prozesu hau azkartzeko, posible da komando-fitxategi edo *script*-ekin lan egitea. Horretarako, komando-segida editatzen da *script* izena hartzen duen fitxategi batean, eta fitxategi horren izena jartzen da *lexc* programaren parametro gisa.

4.13 adibidea: *script* bat

Lehen azaldutako komando-segidan zeuden komandoen artean hauek dira *script1* izeneko fitxategian kokatuko liratekeenak (demagun *lex1.txt* dela 4.11 adibidean azaldutako lexikoa eta *erreg1.fst* 4.5 adibidean ikusitakoa konpilatu ondoren):

```
compile-source lex1.txt
read-rules erreg1.fst
compose-result
save-result morfo1.fst
```

Fitxategi hau edukita, konpilazioa eta konposaketa agindu bakar batez egiten da:

```
lexc script1
```

Aipatu den bezala, erregela-sistema bat baino gehiago eduki daiteke sekuentzian konposatzeko. Horretarako, *compose-result* eta *result-to-source* komandoak konbinatu beharko dira ondorengo script-ean ikus daitekeen moduan.

4.14 adibidea: *script* konplexuago bat

Pentsa dezagun errore batzuk kontuan hartzen dituzten erregela batzuk konpilatu direla (4.7 adibidean dagoena, esaterako) *erreg2.fst* itzultzailean, eta 3 automata (lexikoa, *erreg1* eta *erreg2*) konposatu nahi direla; hau izango da prozesuari dagokion scripta:

```
compile-source lex1.txt
read-rules erreg1.fst
compose-result
result-to-source
read-rules erreg2.fst
compose-result
save-result morfol2.fst
```

Konposaketaren emaitzaren ondorioz, analisi zein sorkuntza morfologikoa egiten duen itzultzaile bat izango dugu eta, ondoko pasartean azaltzen diren programak erabiliz, testuen analisisa egin ahal izango da.

4.4 Probak

Konpilazioen eta konposaketaren ondorioz lortzen den bi mailatako itzultzailea corpusetan probatzeko, bi programa laguntzaile dugu eskura: *tokenizer* eta *lookup*. Programa hauek hizkuntzarekiko independenteak dira, *lexc* eta *twolc* diren bezala, eta lehenak testua token-etan, hitzetan, banatzen duen bitartean, *lookup* programaren zeregina analisisa egitea da, azaleko mailatik lexiko-mailara itzultzea, alegia. Horretarako, parametro bat zehaztuko zaio: lortutako emaitzari dagokion itzultzailea. Beraz, analizatu nahi den corpusa *testu.txt* fitxategian baldin badago, eta analizatzailearen itzultzailea *anali.fst* fitxategian gorde bada, hau da probatzeko egin behar dena:

```
tokenizer <testu.txt | lookup anali.fst
```

4.15 adibidea: euskararako garatutako azpisistema probatzen

corpus izeneko fitxategiak ondoko edukia baldin badu:

```
zakurra berregin birfindu berrasi mozkorren
zakura beregin berfindu birrasi berhasi birhasi mozkoren
```

eta probatzeko lerro hau idazten bada:

```
tokenizer <corpus | lookup anali1.fst
```

analisiaren duten hitzetan, haien analisisa agertuko da:

```
zakurra
  zakur [IZE] + [ABS] + [MS] 26
berregin
  ber [AUR] +egin [ADI]
```

²⁶ [Rgog] ez da azaltzen, tarteko mailan baitzegoen eta, konposaketaren eraginez, tarteko maila desagertzen delako.


```
birfindu
  ber [AUR] +findu [ADI]
berrasi
  ber [AUR] +hasi [ADI]
mozkorren
  mozkor [ADJ] + [GEN] + [MP]
  mozkor [ADJ] + [SUPE]
```

emaitzan hitz hauek agertuko dira analisirik gabe, dena ondo egin baldin badugu:

```
zakura27 beregin berfindu
birrasi berhasi
birhasi mozkoren
```

4.5 Adibide oso bat

Ondoren, adibide osoa eta jarraitua azalduko da. Kasu honetan helburua ez da analisia edo sorkuntza, formen ezagutza baizik, eta horregatik ez da informazio morfologikorik zehazten. Beraz, analisiaren emaitza baino gehiago interesatuko zaigu analizatzeko gai den ala ez jakitea (hau da egiaztatzaile ortografikoen lan egiteko modua)²⁸.

4.16 adibidea: Euskarazko zenbakien idazkera, estilo zaharrean

Milioi bat baino txikiago diren zenbakiak deskribatzeko lexikoa eta erregelak azalduko dira. Lexikoarekin nahikoa izan zitekeen, baina deskribapen dotoreagoa eta egituratuagoa lortzen da aldaketa batzuk erregelen bidez adierazten badira. Horrela, *hama/heme* aldaketa edo *r* kontrolatzearena, erregelen bidez bideratuko da.

Hitz anitzeko espresioak izanik, ez dago oso argi morfologiaren barruan dagoenik, baina hizkuntza mugatu eta itxi bat definitzeko aukera ematen digu, eta, ondorioz, azaldutako tresnen erabilera argitzeko.

Has gaitezen lexikoa idazten. Orain arte ikusitakoarekin alderatuz, aldaketa nabarmenena zuriuneena da: zuriuneak egon daitezke eta horiek adierazteko % ihes-karakterea jarri behar zaie aurretik. Beste aldetik, *Q* diakritikoa erabili da *r* hori gal daitekeela adierazteko, eta ez da ezaugarriarik edo karaktere anitzeko sinbolorik definitu. Hau izango da *lex_zenb.txt* fitxategia:

```
LEXICON Root
  MILAKOAK;
  EHUNekoAK;
  HAMARREKOAK;
  BATEKOAK;

LEXICON MILAONDO
  EHUNekoAK;
  HAMARREKOAK;
  BATEKOAK;

LEXICON EHUNONDO
  HAMARREKOAK;
  BATEKOAK;
```

²⁷ Suposatzen ari gara *zaku* lema ez dagoela lexikoan.

²⁸ Zenbakiak ezagutzearekin batera, beren zenbakizko baliokideak kalkulatzeko erabil daitezke analisia, baina horretarako, ezkerrean zenbakizko balioak eta eragileak erabili beharko lirateke.

LEXICON MILAKOAK

MILA;
 bi MILA;
 hiru MILA;
 lau MILA;
 bost MILA;
 sei MILA;
 zazpi MILA;
 zortzi MILA;
 bederatzi MILA;

LEXICON MILA

mila MILABI;

LEXICON MILABI

#;
 %, % MILAONDO;

LEXICON EHUNEKOAK

EHUN; ! morfema hutsa
 berr EHUN; ! kasu hauetan ez da ebatzi erregelaren bidez
 hirur EHUN;
 laur EHUN;
 bost EHUN;
 sei EHUN;
 zazpi EHUN;
 zortzi EHUN;
 bederatzi EHUN;

LEXICON EHUN

ehun #;
 ehun% eta% EHUNONDO;

LEXICON HAMARREKOAK

hamaQ BATEKOAKHUTS;
 hogei HAMARREKOLAG;
 berrogei HAMARREKOLAG;
 hirurogei HAMARREKOLAG;
 larogei HAMARREKOLAG;

LEXICON HAMARREKOLAG

ta HOGEITA;
 #;

LEXICON HOGEITA

hamaQ BATEKOAKHUTS;
 BATEKOAK;

LEXICON BATEKOAKHUTS

BAT_ARRUNT;
 BAT_BEREZ2;
 #;

LEXICON BATEKOAK

BAT_ARRUNT;
 BAT_BEREZ;

LEXICON BAT_ARRUNT

bi #;
 hiru #;
 lau #;
 bost #;
 sei #;
 zazpi #;
 zortzi #;

```
LEXICON BAT_BEREZ
bat      #;
bederatzi #;

LEXICON BAT_BEREZ2
ika      #;
retzi    #;
```

Hainbat alomorfo erabili da (*hamaQ* bi aldiz, *bat/(hama)ika*, *bederatzi/(heme)retzi*) morfotaktika edo erregelak sinplifikatzearen.

Erregelak idazteko ez dira atal guztiak behar, baina bai *Alphabet* eta *Rules*. Horiez gain, *Muga* sinboloa definitu da *Definitions* atalean. *erreg_zen.txt* fitxategia hau da:

```
Alphabet
a b c d e f g h i j k l m n ñ o p q r s t u v w x y z Q:0

Definitions
Muga = % | # ;          ! morfema-muga edo hasiera zein bukaera

Rules

"hama/heme (1)"
a:e => h: _ m a:e Q: z o r t z i ; ! aukeran hamazortzi/hemezortzi
      h: a:e m _ Q: z o r t z i ;
      h: a: m _ Q: r e t z i ;
      h: _ m a: Q: r e t z i ;
      !hamaQzortzi:hamazortzi
      !hamaQzortzi:hemezortzi

"hamaQ-en kontrola (1)"
Q:r <=> _ Muga ;

"hamaQ-en kontrola (2)"
h:0 <=> t a _ a: m a Q: ;
      t a _ a:e m a:e QR: ;

"hamaR-en kontrola (3)"
a:0 <=> t a h: _ m a Q: ;
      !hogeitahamaQika:hogeitamaika
```

Sinplifikatzeagatik gainsorkuntza txiki bat utzi da, **hamaretzi* onartzen eta sortzen baita. *a:e* erregela bikoitza egitea edo *retzi* sarrera diakritiko batekin markatzea dira irtenbideak gainsorkuntza hau ekiditeko.

Konpilazioari ekingo diogu ondoren. Erregelekin hasiko gara:

```
twolc
>read-grammar erreg_zen.txt
>compile
>save-binary erreg_zen.fst
>quit
```

Konpilatu ondoren probak egin daitezke *lex-test* aukeraren bidez

```
>lex-test hamaQ
>lex-test hamaQzortzi
```

Orain lexikoaren konpilazioa eta konposaketa dator:

```
lexc
>read-source lex_zen.txt
>read-rules erreg_zen.fst
>compose-result zen.fst
>quit
```

zen.fst izango da proba dezakegun itzultzailea. Dena den, aurretik proba batzuk egin daitezke *lexc* barruan *random* eta *lookup* komandoen bitartez:

```
>random  
>lookup hemezortzi  
>lookup bi mila, zortziehun eta berrogeitamaika
```

Azaldutako adibidea nahiko laburra da, eta garatzeko ez da metodologia berezirik behar, baina hizkuntza baten morfologia bere osotasunean deskribatu nahi denean, metodologia egoki bat erabiltzen ez bada, lana amaiezina izan daiteke.

4.6 Metodologia

Aipatutako tresnekin eskala errealeko proiektu bat aurrera ateratzeko ez dago arazorik tresnen ahalmen aldetik, baina garrantzitsua da ondoren zehazten diren metodologia-oinarriei jarraitzea emankortasuna eta arrakasta lortu nahi bada.

2.4 atalean definitutako hiru kontzeptu izan behar dira buruan halako egitasmo baten aurrean: bat, deskribapenaren dimentsioa ematen duen *estaldura*; bi, deskribapenaren kalitatea neurtzen duen *doitasuna*; eta hiru, *gainsorkuntza*-arazoa (estaldura handitzean edo orokortzean sor daitekeen arazoa).

Ondoren azalduko diren bi lerro metodologikoetan biltzen dira oinarrizko ideiak: lexikoaren mantenua eta estalduraren kontrola.

4.6.1 Lexikoaren mantenua

Erregela-sistemen kontrola nahiko erraza den bitartean, eskala errealeko lexikoa izugarri handia izan daiteke, hiztegietan agertzen diren lemez gain, aurizkiak, atzizkiak eta hiztegietan ez dauden forma flexionatuak hartu behar baitira kontuan. Azken hauek morfemen bitartez sortzea posible bada ere, batzuetan errazagoa eta azkarragoa da osorik ematea. Horrela gertatzen da, adibidez, euskarazko aditz laguntzailearekin.

Beraz, milaka sarrerez ari gara eta hori kontrolatzeko ez da egokia ez fitxategi bakar bat, ezta fitxategi multzo bat ere. Datu-base bat dirudi irtenbide aproposena eskala honetako proiektu baterako, baina diseinua ez da sinplea izango, bertan kontrolatu behar baitira, gutxienez, osagai hauek guztiak: morfemak, jarraitze-klaseak, azpilexikoak, hautapen-markak eta ezaugarri morfologikoak. Datu-base honetatik esportazioak prestatu beharko dira tresnen sintaxia betetzen duten fitxategiak sortzeko.

Horrez gain, testuan zehar esan den bezala, alomorfoak eta diakritikoak ekidin behar dira lexikoaren deskribapenean. Beraz, diakritikoak erabili beharrean, datu-basean karaktere anitzeko sinbolo esanguratsuak erabiliko dira, gero baliatuko den programak

zuzenean trata ditzan edo, tratatzerik ez badu, esportazioan bihurketa egiteko. Alomorfoak erabil daitezke, baina salbuespen gisa, forma kanonikotik urrun daudenean, eta beti forma kanonikoari lotuta, aldaketak egiten direnean, alomorfo guztietan egin daitezten. Horrela, euskaraz, *hon*, *har* edo *hobe* alomorfoak erabil daitezke, baina *hau*, *hura* eta *on* forma kanonikoei loturik.

Erregeletarako, berriz, datu-baseak ez dirudi egitura egokiena, gutxi eta konplexuak izaten baitira erregelak; beraz, fitxategietan kokatuko dira. Metodologiaren aldetik, erregela bakoitzaren ondoan adibide batzuk eranstea gomendatzen da, eta erregela-sistema bakoitzeko bi proba-fitxategi mantentzea: bat forma egokietarako, eta forma oker nahasgarrietarako bestea. Aldaketak egiten direnean adibideak eguneratuko dira, eta fitxategi biak erabiliko dira analizatzailarekin ondoko emaitza lortzeko: batean hitz guztiek eduki behar dute analisia, eta bestean batek ere ez.

Sistema osoa corpus zabal batez egiaztatzea komenigarria da, eta bertsio batetik hurrengora emaitzak alderatzea²⁹ ere bai.

4.6.2 Estalduraren kontrola

Ondoko galdera hauen erantzuna emateak lagun dezake deskribapenaren estaldurari ekiteko garaian. Hizkuntza estandarra baino ez da deskribatu behar? Zer egin aldaera dialektalekin edo argot moduko terminoekin? Eta akatsekin? Nola egin aurre testuetan agertzen diren mailegu berriei eta pertsona- zein toki-izen arrotzei?

Erantzuna ez da erraza eta aplikazioarekin lotuta egon daiteke. Dena den, metodologikoki onena lexikoan ahalik eta deskribapen zabalena egitea da, baina beti forma estandarrekin lotuz eta ezaugarri berezi horiek markatuz: dialektoa, argota, hitz arrotzak, etab. Geroago, esportazioan edo iragazkiak eginez, aplikazio bakoitzari ez dagokion informazioa bazter daiteke.

Hizkuntzaren eta arauen eboluzioa ere kontuan hartu beharreko aspektua da. Garai batean estandar izandako formak edota forma zaharkituak ez dira kendu behar, testuetan agertuko baitira batzuetan eta hauek analizatzeko gai izan beharko baitu prozesadoreak; beraz, beste aldaerekin batera kokatu beharko dira, markatu eta forma estandar berriei lotu.

²⁹ Emaitzak alderatzeko *diff* izeneko programa interesgarria dago *Unix* sistemetan.

Erregela morfofonologikoen aldetik, 4.3.3 atalean aztertu den itzultzaileen konposaketak asko laguntzen du estaldura kontrolatzeko prozesu horretan, bai eta beste aspektu metodologiko anitzetan ere. Horrela, hizkuntza estandarrerako bi itzultzailearen konposaketa³⁰, lexikoarena eta erregela morfofonologikoa, nahikoa den bitartean, aldaera fonologikoak kontuan hartzeko erregela-multzo berri batekin egin daiteke konposaketa, erregela-multzo berri horrek desbiderapenak zehazteko balioko baitu. Beraz, erregela guztiak sistema bakar batean jarri beharrian, modulartasunean oinarrituz, erregela-multzo bat egingo dugu helburu bakoitzeko, eta multzo horien konposaketa erabiliz aurre egingo diogu corpus handien tratamenduari.

Ondoko kapituluetan aztertuko da euskararako egin den hiru urratsetako deskribapena: euskara estandarra lehen mailan, aldaera dialektalak eta desbiderapenak bigarrenan, eta gainontzeko fenomenoak hirugarrenean.

4.7 Laburpena

Xeroxek eskaintzen dituen tresnen deskribapena egin da hainbat adibidek lagunduta. Lehenik eta behin, bi tresna ahaltsuenak azaldu dira: *twolc* eta *lexc*, beraien sintaxia eta komandoak deskribatuz, eta hainbat adibide ere gehituz. Ondoren, lortutako emaitzen proba nola egiten den eta adibide oso baten garapena komentatu dira, proiektu errealak egiten direnean kontuan hartu behar diren metodologia-oinarriekin bukatzeko.

³⁰ Edo hiru, morfotaktikaren kontrolerako beste maila bat tartean sartzen bada, 4.4 ataleko 4.9 adibidean egiten zen bezala.

5 Euskararen morfologiaren deskribapena

5.1 Sarrera

Kapitulu honetan euskal morfologiaren tratamendu automatikoa ikusiko dugu, aurreko ataletan aurkeztutako tresnak baliatuz euskararen morfologiaren deskribapena egin daitekeela erakutsiz.

Hasteko, euskal morfologiaren zenbait ezaugarri aurkeztuko ditugu, aztergaia mugatzeko kontuan izan beharrekoak. Ondoren, euskararako prozesadore morfologikoa ikusiko dugu (edo, bestela esanda, bi mailatako ereduaren euskararako egokitzapena).

5.1.1 Euskal morfologiaren zenbait ezaugarri

Has gaitezen, bada, euskararen morfologiaren deskribapen laburra eginez. Aipatuko ditugun puntuak euskal morfologia konputazionalki lantzeko kontuan hartu beharreko ezaugarri nagusiak besterik ez dira:

- Euskara hizkuntza eranskaria da, nahiz ez den eredurik gardenena. Flexio-mailan osagai guztiak elkarren segidan eta independenteki datoz, ordena honetan:

hiztegi-sarrera (lema) + determinatzailea (mugatzailea) + numeroa + kasua(k)
--

Guztiekin testu-hitz bakarria osatzen dute, baina bakoitzak bere informazioa mantentzen du. Eta informazio hori ez da morfologikoa soilik —formalki bai—, morfosintaktikoa baino, erlazio-atzizkiak ere testu-hitzaren parte baitira, funtzioaren berri ematen dutenak, hain zuzen ere.

- Euskarak deklinabide-taula bakarria du: morfema-kopurua mugatua eta aldagaitza da; lema ez da, normalean, aldatzen.
- Sintagmaren barruan azken osagaiak bakarrik darama mugatasuna, numeroa eta kasua; beste guztiak —azalean— forma neutroan datoz. Berez sintaxiari dagokio hau, baina hitzaren gramatika egitean kontuan izan beharrekoa da, testu-hitza bakarka lantzen baitu morfologiak, alegia, analisi ez-osoan ematen ari gara.
- Hitz anitzeko unitate lexikalak ugariak dira, eta aurreko puntuarekin dugun arazo berarekin egingo dugu topo.

- Aditz-flexioa aberatsa da, hiru kasu gramatikalen berri ematen du. Informazio hau zehaztea garrantzitsua da gerora, syntaxian, komunztadura ondo egiteko.
- Euskarak ez du generorik morfologia mailan, eta honek sistema errazten du. Baina ez da ahaztu behar aditzean sexua markatzen dela, eta lexikoan ere bereizketak egiten direla (*aita/ama, memelo/memela, etab.*).
- Lexiko-sorkuntza aberatsa da euskaraz, eratorpena eta hitz-elkarketa erabilita. Hauen tratamendu automatikoak azpikategoriazioa modu sistematikoan landuta izatea eskatzen du. Hitz-eraketak bere ordena finkoa du gainera. Izenek, adjektiboek, izenordainek, determinatzaileak, adberbioek, partizipiotik eratorritako adjektiboek eta aditz-izenek ordena hau betetzen dute:

[[[[[[[AUR] KAT] ATZ] GRA] MUG] NUM] DEK*]
--

Alegia, aurrena aurrizki lexikala (AUR) eta hiztegi-sarrera (KAT) bilduko ditugu, ondoren bi hauen elkarketaren emaitzari atzizki lexikala (ATZ) eransteko; atzizkiak bat baino gehiago ere izan daitezke gainera. Beti behar du ordena hori. Hitz eratorria osatua dugunean has gaitezke gradua (GRA) eranstean, eta honen gainean etorriko dira mugatasuna (MUG), numeroa (NUM) eta deklinabide-atzizkia(k) (DEK).

Aditz nagusiek, bestalde, ordena hauek beharko dituzte:

[[[[[[[AUR] KAT (ADI)] ATZ] ADM] ASP] GRA]
[[[[[[[AUR] KAT (ADI)] ATZ] ADM] MEN] (DEK)]

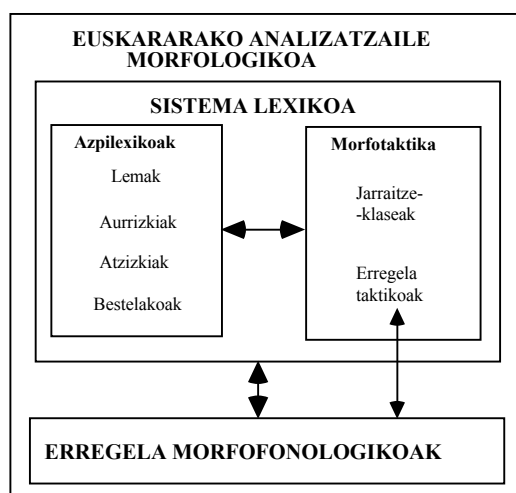
Aditzen kasuan ere gauza bera dugu: aurrizkia (AUR) eta aditza (ADI) elkartu ondoren, eratorri horren gainean ezar dezakegu atzizki lexikala (ATZ) eta, hortik, flexio-morfologia abian jarri, dela aditz-mota (ADM) eta aspektua (ASP) gehituz, dela aditz-motaren (ADM) gaineko menderagailuak (MEN) erantsiz.

Baina, guztiaren gainetik, euskararen egoera izan behar dugu gogoan, estandarizazioa / normalizazioa oso aurreratua baita, baina ez erabatekoa. Ezin dugu ahaztu erabiltzailearentzat tresna baliagarria izatea dela helburu nagusia.

5.1.2 Euskararako prozesadore morfologikoa

Hirugarren kapituluaren Koskenniemiren *bi mailatako eredua* ikusi dugu, hain zuzen gure prozesadorearen abiapuntu izan dena. Ez dugu, beraz, berriro aurkeztuko. Hurrengo orrietan euskararako egokitzapena eginez sortu dugun MORFEUS ikusiko dugu, horretarako *bi mailatako ereduaren* eskema bera jarraituz, hau da, osagaien arabera euskal flexio-morfologia deskribatzen saiatuz: sistema lexikoa eta erregelak.

Baina guretzat MORFEUS hori baino gehiago da: ez dugu zatikatzaile/sortzaile morfologiko hutsa egin nahi, osagaien metaketaren ondorioz lortzen dugun azken emaitza ere interesatzen baitzaigu. Eta horrexegatik gaindituko dugu Koskenniemiren eredua, irteera morfosintaktikoaren berri ematen saiatuko baikara.



8. irudia.- Euskararako analizatzaile morfologikoa.

Honekin estaldura handiko analizatzaile/sortzaile morfologiko orokor bat lortuko dugu, hainbat aplikaziotarako oinarri izango dena.

5.2 Sistema lexikoa

Atal honetan eta hurrengoan euskal morfologia estandarra —flexiboa batez ere— deskribatzen³¹ saiatuko gara, horretarako bi osagai nagusiak azalduz: sistema lexikoa eta erregelak. Sistema lexikoan lexikoa bera (hiztegia) eta morfemak (gramatikako osagaiak) ikusiko ditugu batetik, ondoren jarraitze-klaseak edo gramatika (alegia, morfotaktika, betiere hitzaren gramatikara mugatua) bera aurkezteko. Hori guztia azaleko mailan gauzatuko da erregelak aplikatzean, 5.3 puntuan ikusiko dugunez.

5.2.1 Sarrera lexikoak

Sarrera lexikoak oinarritzko lan-tresna dira, hau da, informazioa. Gero etorriko da morfotaktika horiek elkartzera eta testu-hitzak gauzatzera, baina fase honetan lanerako materiala dugu. Hau, ordea, antolatu egin behar da. Horretarako, bi sail nagusi bereiziko ditugu: kategoria nagusiak eta kategoria lagungarriak.

³¹ Kapitulua honetan hainbat erabaki linguistiko hartu beharko ditugu, askotan horien inguruan araurik ez dagoelako edo, maiz gertatu ohi denez, arauaren irakurketa zehatza behar dugulako. Ezinezkoa da erabaki horiek guztiak hemen justifikatzea, baina horien azalpenak Urkia (1997) tesi-txostenean aurki daitezke.

5.2.1.1 Kategoria nagusiak

Morfema askeak dira, nolabait hiztegi-sarreratzat har ditzakegunak. Kategoria-sistemaren sailkapena ikusiko dugu, hortik abiatuz informazio morfologikoa gehitu ahal izateko.

1. KATEGORIA NAGUSIAK

KATEGORIA	AZPIKATEGORIA	LEXIKOA	ADIBIDEA
IZE Izena	ARR Izen arrunta	izenak	<i>etxe</i>
	PIB Pertsona-izen berezia	izenak	<i>Aitor</i>
	LIB Leku-izen berezia	izenak	<i>Eibar</i>
	IZB Erakundeak-eta	izenak	<i>UEU</i>
IOR Izenordaina	PER Pertsona-izenordaina	izenordainak	<i>gu</i>
	IND Indartua	izenordainak	<i>geu</i>
	IZG Izenordain zehaztugabea	izenordainak	
	GAL Galdetzailea	izenordainak	<i>nor</i>
	MGB Mugagabea	izenordainak	<i>zernahi</i>
	BIH Bihurkaria	izenordainak	<i>-en_burua</i>
	ELK Elkarkaria	izenordainak	<i>elkar</i>
	ERL Erlatibozkoa	izenordainak	<i>zeina</i>
DET Determinatzailea	ERK Erakuslea (arrunta)	determinatzaileak	<i>hau</i>
	IND (indartua)	determinatzaileak	<i>berau</i>
	ZNB Zenbatzailea	determinatzaileak	
	DZH Zehaztua	determinatzaileak	<i>hiru</i>
	ZHG Zehaztugabea	determinatzaileak	<i>batzuk</i>
	ORO Orokorra	determinatzaileak	<i>guztia, oro</i>
	DZG Determinatzaile zehaztugabea	determinatzaileak	<i>edozein</i>
ADJ Adjektiboa	IZO Izenondoa	adjektiboak	<i>eder</i>
	IZL Izenlaguna	adjektiboak	<i>alegiatzko</i>
ADI Aditza	SIN Aditz sinplea (bakuna)	aditzak	<i>ibili</i>
	ADK Aditz konposatua	aditzak	<i>negar_egin</i>
	ADP Aditz-perifrasia	aditzak	<i>behar_izan</i>
ADB Adberbioa	ADO Aditzondoa	adberbioak	<i>orain</i>
	ALG Adizlaguna	adberbioak	<i>ahuspez</i>
ESK Esklamazioa		lemak	<i>eup</i>
ESR Esaera		lemak	<i>adarra_jo</i>
LOT Loturazkoa	JNT Juntagailua	lemak	<i>eta</i>
	LOK Lokailua	lemak	<i>beraz</i>
	MEN Menderagailua	lemak	<i>arren</i>
POS Posposizioa		lemak	<i>-z_kanpo</i>
PRT Partikula		lemak	<i>omen</i>

Baina datu-basean ez ditugu kategoriak eta azpikategoriak bakarrik aipatuko. Informazio morfologikoa eskaini nahi badugu, neurri batean behintzat azpikategorizazioa hartu beharko dugu kontuan. Ez da hori gure helburua, baina tasunik orokorrenak bildu nahi izan ditugu, besteak beste analisiaren irteeran beharrezko izango ditugulako. EDBL "ezaugarria/balioa" bikoteka antolatua dago, azpilexikoak ondo ulertzeko ezinbestekoa baita. Kategoriaren eremuak beti azaldu behar du, baina honen barruan tasun desberdinak daude, ondoren adierazten den moduan:

- **Izena:** IZE kategoriak lau azpikategoria ditu: ARR (izen arrunta), PIB (pertsona-izen berezia), LIB (leku-izen berezia) eta IZB (izen berezi orokorra, erakundeetarako-eta). Eta lau azpikategorizazio-ezaugarri:

*mn*l (mugatasun/numero lexikala): MS, MP, null

biz (biziduntasuna): +, -

zenb (zenbakarritasuna): +, -

pl (pluraltasuna): +, -, null

etxe (IZE, ARR, null, -, +, -)

aita (IZE, ARR, null, +, +, -)

guraize (IZE, ARR, null, -, +, +)

Aitor (IZE, PIB, MS, +, +, -)

Azpeitia (IZE, LIB, MS, -, +, -)

Ande (IZE, LIB, MP, -, +, +)

etxe³²

aita[Aorg]

guraize

9aitor[Rgog]

9azpeitia[Aart]

9ande

Izen guztiek behar dituzte *biz* eta *zenb* ezaugarriak: guztiak dira bizidun ala bizigabe, zenbakarri ala zenbakaitz; ez dago, beraz, *null* baliorik. Baina beste biek (*mn*l eta *pl*) ez dute nahitaez balio bat hartu beharrik: *mn*l ez dute izen arruntek —orokorre— eta ez da eremu hori beteko; aldiz, izen bereziek, izenordainek eta zenbait determinatzailek berezko mugatasuna dute —singularra nahiz plurala—. *Pl* ere pluralean bakarrik edo ez-pluralean bakarrik erabiltzen direnek dute, besteek, orokorre, *null* hartuko dute-eta.

- **Izenordaina:** IOR kategoriak PER (pertsona-izenordaina), IND (pertsona-izenordain indartua), IZG (izenordain zehaztugabea), GAL (galdekaria), MGB (mugagabea), BIH (bihurkaria), ELK (elkarkaria) eta ERL (erlatibozkoa) azpikategoriak ditu, eta bi eremu gehiago:

*mn*l (muga(gabe)tasun/numero lexikala): MG, MS, MP, null

per (pertsona): NI, HI, GU, ZU, ZUEK, null

gu (IOR, PER, MP, GU)

nerau (IOR, PER_IND, MS, NI)

nerau (IOR, PER_IND, MS, NI)

nor (IOR, GAL, MS, null)

elkar (IOR, ELK, null, null)

gu

nerau

neron³³

nor

elkar

Izenordain izateak muga(gabe)tasuna dakar berarekin eta horregatik dute izenen eremu bera, *MG* gehituta —zehaztugabeentzat—. Bestalde, pertsona-izenordainek zein pertsonari dagokion zehazten dute.

- **Determinatzailea:** DET kategoriak eta ERK (erakuslea), IND (erakusle indartua), DZH (zenbatzaile zehaztua), ZHG (zenbatzaile zehaztugabea), ORO (orokorra), DZG (determinatzaile zehaztugabea) azpikategoriek eremuok dituzte:

*mn*l (muga(gabe)tasun/numero lexikala): MG, MS, MP, null

*hur*m (hurbilpen-maila): HAU, HORI, HURA, null

³² Bi mailatako forma (tarteko maila lexiko-itzultzaileetan).

³³ Alomorfoa.

hau (DET, ERK, MS, HAU)	hau
hau (DET, ERK, MS, HAU)	hon
hiru (DET, DZH, MP, null)	hiru
zenbait (DET, ZHG, MG, null)	zerbait
guztia (DET, ORO, null, null)	guzti
zer (DET, DZG, null, null)	zer

mnl beharrezkoa da hemen, determinatzaile izateak berarekin baitakar muga(gabe)tasuna eta hurbilpen-maila (*hurm*), erakusleek hartuko dutena.

- **Adjektiboak** (ADJ) eta **aditzak** (ADI) kategoria eta azpikategoria besterik ez dute. Adjektibo kategoriaren barruan izenondoa (IZO) eta izenlaguna (IZL) banaketa egin da, eta aditzaren barruan aditz bakuna edo sinplea (SIN), konposatua (ADK) eta aditz-perifrasia (ADP). Aditzak askozaz informazio gehiago eskatzen du, dudarik ez da, baina horrek azpikategorizazioa sistematikoki lantzea eskatzen du eta, hori, eginkizun dago oraingoz.

txiki (ADJ, IZO)	txiki
alegiakzo (ADJ, IZL)	alegiakzo
etor (ADI, SIN)	etor[Rgog]
negar_egin (ADI, ADK)	negar_egin
behar_izan (ADI, ADP)	behar_izan

- **Adberbioa**: ADB kategoria eta ADO (aditzondoa), ALG (adizlaguna) azpikategoriaz gain adberbio-mota ezaugarria ere badute, denbora, lekua, modua, maila edo iritzia adierazten duten azaltzeko.

<i>adbm</i> (adberbio-mota): DENB, LEKU, MODU, MAIL, IRIT

gaur (ADB, ADO, DENB)	gaur[Rgog]
hemen (ADB, ADO, LEKU)	hemen
edonola (ADB, ADO, MODU)	edonola
arras (ADB, ADO, MAIL)	arras
agian (ADB, ADO, IRIT)	agian
ahuspez (ADB, ALG, MODU)	ahuspez

- **Esclamazioak** eta **esaerak**: esclamazioak kategoriaren informazioa besterik ez du eta, esaerak, kategoriaz gain, mugakizunaren kategoria, hori baita morfotaktikan eragina izango duena.

<i>m_kat</i> (mugakizunaren kategoria): ADI, IZE, ADB

kaixo (ESK)	kaixo
eup (ESK)	eup
adarra_jo (ESR, ADI)	adarra_jo
batez_ere (ESR, ADB)	batez_ere

- **Loturazkoa**: LOT kategorian hiru azpikategoria biltzen dira: JNT (juntagailua), LOK (lokailua) eta MEN (menderagailua)³⁴. Hala ere, juntagailuak eta lokailuak

³⁴ Menderagailua, gehienetan, morfema ez-askea da eta, ondorioz, kategoria lagungarrien sailean ageri da. Bikoiztuta ezarriko dugu informazioa, beraz. *Arren*, *eta* kausazkoa, *nahiz* modukoak osagai independenteak dira eta, horrenbestez, bere lekua behar dute hemen. Are gehiago, menderagailu batzuk posposizioen tratamendu bera eskatzen dute, osagai bat aditzari lotua baitute eta, bestea, askea: *-n arren*, *nahiz...* *-n*, etab.

lotura-maila adierazten dute eta menderagailuak, aldiz, erlazioa edo mendekotasun-maila. Banaka landu beharrekoak dira, beraz.

Juntagailuen eta lokailuen lotura-mailan emendiozkoak, hautakariak, aurkaritzakoak, ondoriozkoak edo kausazkoak diren zehaztuko dugu:

lotm (lotura-maila): EMEN, HAUT, AURK, ONDO, KAUS

eta (LOT, JNT, EMEN)	eta
baina (LOT, JNT, AURK)	baina
ordea (LOT, LOK, AURK)	ordea
beraz (LOT, LOK, ONDO)	beraz

Menderakuntza, bestalde, baldintzazkoa, denborazkoa, erlatibozkoa, helburuzkoa, kausazkoa, konpletiboa, kontzesiboa, moduzkoa edo zehargaldera adierazten duena izan daiteke. Erlazio-mota hauetako gutxi batzuk besterik ez dira morfema aske gisa erabiltzen, baina ez-askeak ere aintzat hartu ditugu sail honetan, erlazioak bietarako balio baitu.

erl (erlazioa): BALD, DENB, ERLT, HELB, KAUS, KNPL, KONT, MODU, ZGAL

eta (LOT, MEN, KAUS)	eta
arren (LOT, MEN, KONT)	arren

- **Partikula:** PRT kategoriak bi ezaugarri ditu:

egia (egia balioa): +, -

ziur (ziurtasun-neurria): +, -

ba (PRT, +, -)	ba ³⁵
omen (PRT, -, +)	omen

Hauk ez dute esan nahi egia adierazten duten ala ez, egia balioari erantzuten dioten ala ez baizik. Eta, ziurak diren ala ez, ez zaigu axola, balio hori duten bakarrik.

- **Kasu konplexu bat: hitz anitzeko unitate lexikalak.** Forma konplexuen berri emateko beharrezkoak dira: *gutxi_gorabehera*, *negar_egin*, etab. Lehenengo adibide hau beti ordena horretan azalduko da eta ez du morfotaktikaren aldetik arazorik planteatzen, baina *negar_egin*-en ordena aldatu egin daiteke, eta lehenengo osagaia partitiboarekin azaldu: *ez du negarrik egiten*. Egun ordena horretan datorrenean bakarrik ezagutzen du euskararako sistemak, "_" markak lotura ematen dio-eta. Sintaxiak ezagutuko ditu ordena aldatzen denean.

³⁵ Ba partikula bi lekutan aurkituko dugu: morfema aske eta ez-aske gisa; izan ere, lotuta zein bereiz idatz daiteke. Menderagailuen kasu beraren aurrean gaude hemen ere. Ba partikularen eta aditzaren artean beste partikula bat tarteka daiteke eta, orduan, askea izango da: *ba omen dator* (kategoria nagusia: morfema askea) vs. *bai, badator* (kategoria lagungarria: morfema ez-askea).

5.2.1.2 Kategoria lagungarriak

Sail honetan hiztegi-sarrera ez direnak sartu ditugu: ez-askeak edo klitikoak direlako batzuk, adizki jokatuak direlako (laguntzaileak eta trinkoak, alegia) beste batzuk, eta, azkenik, suprakategoriatzat har ditzakegulako beste hainbat (laburtzapenak, esaterako, aske diren arren, ez baitira irizpide lexikografikoen arabera hiztegi-sarrera).

2. KATEGORIA LAGUNGARRIAK

ADIZKIAK

ADL Aditz laguntzailea

ADT Aditz trinkoa

MORFEMA EZ-ASKEAK

Kategoria morfosintaktikoak

MUN Mugatasuna / numeroa

DEK Deklinabide-morfema

GRA Graduatzailea

ADM Aditz-mota

ASP Aspektua

ERL MEN Erlazioa: menderagailua

Beste morfema ez-askeak

ATZ Atzizki lexikala

AUR Aurrizki lexikala

PRT Baieztapen-partikula

ELI_IZE Elipsia: izen eliditua

ELK Elkarketa-marra

MAR Marratxoa

BESTELAKOAK (laburtzapenak)

LAB Laburdura

SIG Sigla

SNB Sinboloa

- **Adizkiak:** hauen berezko kategoria ADI den arren, zuzenean ADL (aditz laguntzailea) eta ADT (aditz trinkoa) dute, flexionatuta sartu dira-eta. Berekin duten informazioa:

oin (oinarriaren forma): *izan*, **edun*, **edin*, **ezan* (aditz laguntzaileak)

izan, *ukan*, *jakin*, *esan*... (multzo irekia aditz trinkoek)

mdn (modua/denbora): A1, A2, A3, A4, A5, B1, B2, B3, B4, B5, B6, B7, B8, C³⁶

nrk (nork pertsona): NIK, HIK-TO, HIK-NO, HARK, GUK, ZUK, ZUEK, HAIEK, null

nri (nori pertsona): NIRI, HIRI-TO, HIRI-NO, HARI, GURI, ZURI, ZUEI, HAIEI, null

nor (nor pertsona): NI, HI, HURA, GU, ZU, ZUEK, HAIEK

asp (aspektua): EZBU (ADTek bakarrik), null

³⁶ Euskaltzaindiaren (1973) arabera egindako sailkapena, gerora *EGLU-II* (1987) liburukian egokitu eta terminologikoki gaurkotu dena, baina hasierako balio-sistemari eutsiz, oinarrian bera baita. A-z hasiek orainaldia adierazten dute, B-z hasiek iragana eta C-k agintea. Zenbakien bitartez modua zehazten da: A1 (*da/du* mota, lehen indikatiboazko orainaldia izendatzen zutena), A2 (*dateke/duke*), A3 (*-dadi/-deza-*, lehen subjuntiboazko orainaldia), A4 (*-dadi/-deza-*, lehen ahalezko baldintza hurbila), A5 (*daiteke/dezake*); B1 (*zen/zuén*), B2 (*litzateke/luke*), B3 (*zatekeen/zukeen*), B4 (*-litz/-lu*), B5 (*zedin/zezan*), B6 (*-ledi/-leza-*), B7 (*liteke/lezake*), B8 (*zitekeen/zezakeen*); eta C, agintekera.

naiz (ADL, IZAN, A1, null, null, NI, null)
naiz (ADT, IZAN, A1, null, null, NI, EZBU)
generamatzion (ADT, ERAMAN, B1, GUK, HARI, HAIEK, EZBU)

Oinarriaren formak sarrerari egiten dio erreferentzia; *mdn* ezaugarriak modua eta denbora adierazten du; pertsonak *nrk*, *nri* eta *nor* ezaugarrietako balioekin³⁷ adierazten dira eta, azkenik, aditz trinkoek berekin dute aspektuaren marka, beti burutugabea. Aditz laguntzaileek ez dute eremu hori bete izango, aditz perifrastikoetan aditz nagusiak hartzen baitu aspektua, aurrerago ikusi ahal izango dugunez.

- **Morfema ez-askeak: kategoria morfosintaktikoak.** Klitikoek multzoa berez informazio morfoloikoa —are morfosintaktikoa— duena da, hor biltzen baitira flexioko osagaiak: mugatasuna, numeroa, deklinabidea, gradua, aditz-mota, aspektua eta erlazioa (menderakuntza).
 - **Mugatasuna/numeroa:** MUN kategorian (artikulu³⁸), mugagabe, mugatu singular edo plural (hurbila barne) den zehazten da.

mun (mugatasuna/numeroa): MG, MS, MP, MPH

0 (MUN, MG) +0
a (MUN, MS) +a
o (MUN, MPH) +o

- **Deklinabidea:** DEK kategoriaren barruan, kasu guztiak daude.

kas: ABS, PAR, ERG, DAT, INE, GEL, ALA, ABZ, ABU, ABL, GEN, SOZ, INS, MOT, DES, PRO, BAN.

0 (DEK, ABS) +0
i (DEK, DAT) +i
gatik (DEK, MOT) +gatik

	MG	M	S	M	P	M	PH	KM
1) Absolutiboa	0	a	0	a	k	o	k	0
2) Partitiboa	0							ik
3) Ergatiboa	0	a	0	e		o		k
4) Datiboa	0	a	0	e		o		i
5) Inesiboa	0	0	0	e		o	a/ta	n
6) Leku genitiboa	0	0	0	e		o	0/ta	ko
7) Adlatibo s.	0	0	0	e		o	0/ta	ra(t)
8) Hurbiltze adlat.	0	0	0	e		o	0/ta	rantz
9) Muga adlat.	0	0	0	e		o	0/ta	raino
10) Ablatiboa	0	0	0	e		o	0/ta	tik (r)ik
11) Jabego-genitiboa	0	a	0	0		o		en
12) Sozietiboa	0	a	0	e		o		ekin
13) Instrumentala	0	a	0	e		o		z
14) Motibatiboa	0	a	0	e	k	o	k	(en) gatik gatik
15) Destinatiboa								
Gen. g. er.	0	a	0	e		o		entzat endako

³⁷ Kontu egin, baina, *nork* eta *nori* ezaugarrien balioetan *null* ere onartu dugula; ez ordea *nor*-en kasuan, beti agertu behar baitu.

³⁸ Artikulua, berez, determinatzailea da, eta DET kategoria dagokio. Hala izango luke morfema askea balitz, hau da, kategoria nagusia balitz.

Ines. g. er.	0	0	0	e	o	0/ta	ko
Adlat. g. er.	0	0	0	e	o	0/ta	rako
Banatzailea	0						ko
Prolatiboa	0						tzat
							tako

Irudi honetan ikus daiteke mugatasuna/numeroa aldatzen dela; baina deklinabide-kasuak aldagaitzak dira: taula bakarra baitu euskarak. 0ek informazio morfologikoa ematen dute, nahiz azalean ez diren gauzatzen: zero morfemak dira hauek.

- **Gradua:** GRA kategoriak gradu-maila zehazten du: konparatiboa (-ago), superlatiboa (-en) eta gehiegizkoa (-egi).

grdm (gradu-maila): KONP, SUPE, GEHI

ago (GRA, KONP)	+ago
egi (GRA, GEHI)	+egi
en (GRA, SUPE)	+en

- **Aditz-mota:** aditz nagusiek hartzen dute aditz-mota, ADM kategorian: aditzoina, aditz-izena eta partizipioa.

adm (aditz-mota): ADOIN, ADI_IZE, PART

0 (ADM, ADOIN)	+0 (ikus)
i (ADM, PART)	+i (ikusi)
te (ADM, ADI_IZE)	+te (ikuste)

- **Aspektua:** ASP kategorian, hau ere aditz nagusiek hartzen dute: burutua (-0, -tu, -i), burutugabea (-ten, -tzen) eta geroa (-ko, -en).

asp: BURU, EZBU, GERO

tu (ASP, BURU)	+tu
ten (ASP, EZBU)	+ten
ko (ASP, GERO)	+ko

- **Erlazioa**³⁹: ERL kategoria eta MEN (menderagailua) azpikategoria *erl* ezaugarriak osatzen du. Honek aditz nagusietarako zein adizki jokatueterako balio du. Menderakuntza baldintzazkoa, denborazkoa, erlatibozkoa, helburuzkoa, kausazkoa, konpletiboa, kontzesiboa, moduzkoa edo zehar-galdera adierazten duena izan daiteke.

erl (erlazioa): BALD, DENB, ERLT, HELB, KAUS, KNPL, KONT, MODU, ZGAL

ba (ERL, MEN, BALD)	ba+
la (ERL, MEN, KNPL)	+la
la (ERL, MEN, MODU)	+la
la (ERL, MEN, DENB)	+la
tzea (ERL, MEN, KNPL)	+tzea

³⁹ Kategoria nagusien atalean ere aipatu ditugu menderagailuak, lotura-elementuen artean. Hain zuzen ere, gutxi batzuk morfema aske gisa erabili ohi dira: *arren* (*hil arren*, *lana ematen jarraitu zuen*) vs. *-tearren* (*bi baliorekin*); *eta* (*pozik nago*, *asmatu dugu(-)eta*; *hori dela eta*, *gauzak aldatu egin dira*) vs. *juntagailua* (*jan eta lo*, *ez du besterik egiten*).

- **Beste morfema ez-askeak.** Sail honetan eratorpen-hizkiak, *ba-* baieztapen-partikula eta elipsia ditugu batetik, informazio morfologikoa ematen dute-eta. Baina, hauekin batera, bi elementu grafiko ere beharko ditugu, informazio morfologikorik izan eta eman ez, baina hitzen osaera edo analisi egokirako beharko ditugunak: elkarketa-marra eta marratxoa, alegia.
- **Eratorpen-aurrizkiak eta -atzizkiak:** AUR eta ATZ kategoriak sortu ditugu, oraingoz kategoria-aldaketa eragiten ez dutenentzat. Horregatik da ez-oso eta ezaugarri batzuk ez dira orokorrak izango, ez eta balioak definitiboak ere.

<i>erat</i> (eratorria): +
<i>e_kat</i> (eratorriaren kategoria): IZE, ADJ...
<i>e_azp</i> (eratorriaren azpikategoria): ORD, DIS, FAK, null
<i>gra</i> (gradua): HAND, TXIK, null

ber (AUR, +, ADI, = ⁴⁰ , null)	ber[Rgog]+
araz (ATZ, +, ADI, FAK, null)	+araz
tzar (ATZ, +, =, HAND)	+tzar[Rgog]

Eratorpena oso modu mugatuan landu dugunez, analisiaren irteeran forma berria eratorria dela adierazteko *erat* ezaugarria sortu dugu, '+' balioa duena betiere. Eratorriaren kategoria ez da bete momentuz, baina baterakuntza-erregelatan erabiliko da, hain zuzen ere *e_kat=o_kat* adierazteko, alegia, eratorriaren kategoria oinarriaren kategoria dela esateko. Eratorriaren azpikategoria bai, aldatzen da, baina gure sistema mugatuan hiru kasutan bakarrik: zenbatzaile zehaztuak ordinal edo banatzaile bihurtzen ditu eta aditz sinplea faktitibo. Gradua ere hemen bildu dugu, atzizki apreziatiboek —handigarriek eta txikigarriek— oinarria graduatzen dutelako.

- ***Ba-* baieztapen-partikula:** morfema askeen atalean azaldu dugun informazio bera jasotzen du aurrizki honek, idazkera-kontuak besterik ez baititu bereizten: *egia* (egia balioa) eta *ziur* (ziurtasun-neurria).

ba (PRT, +, -)	ba+
----------------	-----

- **Elipsia:** zero morfema batek adierazten du ezkutuan izen bat dugula. Horregatik ezarri diogu *ELI_IZE* kategoria, *ARR* (arrunt) azpikategoria duena gainera⁴¹. Maiztasun handiarekin azaltzen denez, nahitaez aztertu beharrekoa da fenomeno hau. Alegia, genitiboak izenlagunak sortzen

⁴⁰ = marka ezarri dugu eratorriaren azpikategoria oinarriaren bera izango delako: aditz sinplea, esaterako.

⁴¹ Badira beste kategoria batzuk ere, *urrezko* (*urrez egindako*, esaterako) modukoak, baina ez ditugu hemen tratatuko. Horiek askotan hiztegi-sarrera dira gainera.

ditu, baina honek normaltasun osoz hartu ohi ditu atzizkiak: *lagunarena* (*lagunaren liburua*), *tabernakoa* (*tabernako nagusia*), etab.⁴²

0 (ELI_IZE, ARR)

+0

- **Elkarketa-marra eta marratxoa:** bi kategoria hauek ez dute informazio morfologikorik bere baitan, baina morfotaktika bideratzen laguntzen duten neurrian, ezinbesteko ditugu. Duten informazioa, beraz, kategoria hutsa izango da: *ELK* eta *MAR*.

- (ELK)

+ - (kale-kantoi)

- (MAR)

+ - (puenting-a)

- **Bestelakoak**⁴³: laburtzapenak sail orokorrean biltzen diren laburdurak, siglak eta sinboloak daude hemen, horiek ere hiztegian egon beharra dutelako —ez hiztegi tradizioaletan, izen bereziekin gertatzen den bezala, baina bai euskararako datu-base lexikalean, batez ere corpusen analisisa kontuan hartzen bada—. Hala ere, suprakategoria modukoa da hau, erreferenteak daramana, ondoren adieraziak (mugakizunak, ikus esklamazioak eta esaerak, kategoria nagusien atalean) benetan dagokion gramatika-kategoria ere bai baitarama.

UEU (SIG, IZE, IZB⁴⁴)

9u9e9u[SigB]

Laburbilduz, hona ezaugarri eta balio guztiak, taula batean:

Ezaug	Esanahia	Balioak
adbm	adberbio-mota	DENB, LEKU, MODU, MAIL, IRIT
adm	aditz-mota	ADOIN, ADI IZE, PART
asp	aspektua	BURU, EZBU, GERO
azp	azpikategoria	ARR, PIB, LIB, IZB (izenak); PER, PER_IND, IZG, GAL, MGB, BIH, ELK, ERL (izenordainak); ERK, ERK_IND, DZH, ZHG, ORO, DZG (determinatzaileak); IZO, IZL (adjektiboak); SIN, ADK, ADP (aditzak: nagusiak); ADO, ALG (adberbioak); JNT, LOK (loturazkoak); MEN (erlazio-morfemak); ORD, DIS, FAK (eratorpen-atzizkiak: determinatzaileak eta aditzak)
biz	biziduna (IZE)	+, -
egia	egiatasunezkoa (PRT)	+, -

⁴² Kasu honetan ez dugu bestelako informazio lexikorik gehituko (biziduntasuna, esaterako), testuinguru zabalaren arabera baita, aurretik aipatu dugun zerbaiti egiten dio erreferentzia: anafora baten aurrean gaude.

⁴³ “Bestelakoak” izendapena kategoria-sailkapen estandarretan erabiltzen dena da (ikus *SUC-tags* sistemako ‘Supplementary tags’ atala edo *CREA*-ko ‘Miscelánea’). Izen bera mantendu dugu guk ere, baina laburtzapenak dira berez hemen bildu ditugunak UZEIk (1988) proposatutako banaketaren arabera. Hala ere, suprakategoria bezala planteatu dugu hau, ondoren bakoitzak dagokion kategoria nagusia hartuko baitu.

⁴⁴ IZBtik, izen berezitik, alegia, izenen ezaugarriak aterako dira, noski.

erat	eratorria (AUR, ATZ)	+
e_azp	eratorriaren azpikategoria	ORD, DIS, FAK
e_kat	eratorriaren kategoria	---
erl	erlazioa (ADI, ADT, ADL)	BALD, DENB, ERLT, HELB, KAUS, KNPL, KONT, MODU, ZGAL
grdm	gradu-maila (GRA)	KONP, SUPE, GEHI
gra	gradua (ATZ: apreziatiboak)	HAND, TXIK
hurm	hurbilpen-maila (DET ERK)	HAU, HORI, HURA
kas	deklinabide-kasua	ABS, PAR. ERG, DAT, INE, GEL, ALA, ABZ, ABU, ABL, GEN, SOZ, INS, MOT, DES, PRO, BAN
kat	kategoria (nagusiak eta lagungarriak)	IZE, IOR, DET, ADJ, ADI, ADB, ESK, ESR, LOT, POS, PRT, ADL, ADT, LAB, SIG, SNB, MUN, DEK, GRA, ASP, ERL, ATZ, AUR, ELM, ELI, IZE, MAR
lotm	lotura-mota	EMEN, HAUT, AURK, ONDO, KAUS
m_kat	mugakizunaren kategoria	ADI, ADB, IZE
mdn	modua/denbora	A1, A2, A3, A4, A5, B1, B2, B3, B4, B5, B6, B7, B8, C
mnl	mugatasun/numero lexikala	MG, MS, MP
mun	mugatasuna/numeroa	MG, MS, MP, MPH
nor	nor (ADT/ADL pertsona)	NI, HI, HURA, GU, ZU, ZUEK, HAIEK
nri	nori (ADT/ADL pertsona)	NIRI, HIRI-TO, HIRI-NO, HARI, GURI, ZURI, ZUEI, HAIEI
nrk	nork (ADT/ADL pertsona)	NIK, HIK-TO, HIK-NO, HARK, GUK, ZUK, ZUEK, HAIEK
oin	oinarriaren forma (jokatuak)	IZAN, *EDUN, UKAN, JAKIN,... (multzo irekia)
per	pertsona (izenordainak)	NI, HI, GU, ZU, ZUEK
pl	plurala (IZE <i>pluralia</i> <i>tamtum</i>)	+, -
zenb	zenbakarria (izenak)	+, -
ziur	ziurtasunezkoa (partikulak)	+, -

Honaino sistema lexikoko osagaiek berekin duten informazioaren azalpena. Kategoria sistema eraiki ostean, baina, azpilexikoak definitu eta taldekatu beharko ditugu morfotaktika aplikatu ahal izateko.

5.2.2 Azpilexikoak

Aurreko atalean lexikoa prestatu dugu eta informazio morfologikoa erantsi. Hala ere, horri egitura bat eman behar zaio, informazio morfologikoaren arabera antolatu eta multzokatu behar da guztia: alegia, azpilexikoak behar ditugu.

Kategoria-sistemako sailkapena baliatu ahal izango dugu kategoria nagusiekin, sarrera lexiko diren neurrian azpilexiko berean sar baititzakegu izen guztiak, esaterako. Morfotaktikarekin lotura zuzena du honek, aurrizki lexikalek, adibidez, atzetik "izenak" har baititzakete (izen arruntak badira ere); *ber-* aurrizkiak, bestalde, aditzak (nagusiak) har ditzake ondoren. Hori da, besteak beste, azpilexikoen balioa.

Bestalde, atzizki gramatikalak oso sakabanatuta daude, multzokatzea ia ezinezkoa da kasu askotan, morfotaktikak zuzenean eragiten die-eta. Esaterako, adberbioek har ditzaketen deklinabide-kasuak ia kasuan-kasuan zehaztu behar dira eta, ondorioz, ez dago azpilexiko zabaletan biltzerik. Jarraitze-klaseek azpilexikoak biltzen dituztenez,

haiek bideratzen dute azpilexikoak batu ahal izatea ala ez. Hau da bi mailatako morfologiaren muga: morfotaktika azaltzeko mekanismoa sinpleegia baita.

Horiek horrela, azpilexikoen egitura hau da: azpilexikoek izen identifikatzaile bat hartzen dute eta horren barruan osagaia bera ageri da, informazio morfologikoarekin⁴⁵ eta dagokion jarraitze-klasearekin. Adibide batekin ikus dezakegu:

```
LEXICON izenak
Aitor[IZE] [PIB]:9aitor[IZE] [PIB] [Rgog] IB ;
```

izenak sail osoa⁴⁶: izen guztiak hemen bildu ahal izango ditugu. Azter dezagun guk jaso dugun adibide bakarra: *Aitor*. Bi mailatan jaso dugu informazioa: bi puntuen aurrekoa, ezkerrekoa, kanporatuko den informazioa da; eskuinekoa, aldiz, bi mailatako erregelei dagokiena. Lehenago ere esan dugunez, hitz-hasierako 9 maiuskularen adierazgarri da, alegia, ondoren datorrena letra larritz interpretatu behar dugula: 9a = A. 9aitor-en atzetik informazio morfologikoa dugu: [IZE] (izena), [PIB] (pertsona-izen berezia) eta [Rgog] diakritikoa (R gogorra du bukaeran, erregelek r bikoiztu beharko dute: *Aitor*, baina *Aitorrek*). Honekin, puntu honetan behintzat, informazio morfologiko nahikoa dugu fonologia bideratzeko. Azkenik, espazioaren ondoren datorren IB jarraitze-klasearen izena da, hau da, IB horren barruan *Aitor* izen bereziak atzetik har ditzakeen azpilexiko guztiak —eta bakarrak— bilduko dira. Baina hori hurrengo atalean ikusiko dugu.

Ikus ditzagun azpilexikoak, kategoria nagusiak aurrena, eta lagungarriak hurrena.

Kategoria nagusiak

Kategoria nagusietan hiztegia bera biltzen dela esan dugu gorago. Beraz, hiztegi osoa hona ekartzerik ez dagoenez, lagin erakusgarri bat bilduko dugu.

Izenak

“Izenak” atal orokorra dagokie, aurizkiek horixe baitute haien jarraitze-klasean. Baina bakoitzak du bere informazioa eta dagokion jarraitze-klasea.

```
LEXICON izenak
Eusko_Jaurilaritza[IZE] [IZB]:9eusko_9jaurilaritza[IZE] [IZB] [Aorg] LIB_IB ;
Aitor[IZE] [PIB]:9aitor[IZE] [PIB] [Rgog] IB ;
Azpeitia[IZE] [LIB]:9azpeitia[IZE] [LIB] [Aart] LIB ;
Donostia[IZE] [LIB]:9donostia[IZE] [LIB] [Aorg] LIB ;
Estatu_Batu[IZE] [LIB]:9estatu_9batu[IZE] [LIB] LIB_PL ;
surfer[IZE] [ARR]:surfer[IZE] [ARR] [Rgog] I_biz_MAR ;
zakur[IZE] [ARR]:zakur[IZE] [ARR] [Rgog] I_biz ;
kultura[IZE] [ARR]:kultura[IZE] [ARR] [Asalb] I ;
guraize[IZE] [ARR]:guraize[IZE] [ARR] I ;
```

“Eusko Jaurilaritza” izena [IZE] da, baina erakunde-izena [IZB]; honen arabera, beraz, jarraitze-klaseak bizidun eta bizigabeen morfotaktika onartuko du: *Eusko Jaurilaritzako* zein *Eusko Jaurilaritzaren*, adibidez. “Aitor” pertsona-izen berezia da [PIB] eta r gogorra du bukaeran [Rgog]. “Azpeitia”, aldiz, leku-izen berezia [LIB] eta a artikuluduna [Aart], kasu gramatikaletan mantentzen dena, baina leku-denborazkoetan galdu (*Azpeitiak*, baina *Azpeitiko*). Aldiz, *Donostiak*

⁴⁵ Hemen kategoria, azpikategoria eta diakritikoak besterik ez dira jasotzen.

⁴⁶ Beti letra xehez, letra larriak jarraitze-klasea adierazten baitu.

informazio bera du, baina *a* organikoa du [*Aorg*], hau da, berea du eta ez du inoiz galtzen (*Donostiak*, eta *Donostiako*), etab.

Aditz nagusiak

Aditzoina da oinarritzko osagaia, ez partizipioa. Izenetan ikusi dugun bezala, informazio morfologikoa, diakritikoak eta jarraitze-klaseak dituzte.

```
LEXICON aditzak
abes[ADI][ARR] A7 ;
adieraz[ADI][ARR] A12 ;
ahal[ADI][ARR] A10 ;
argitara[ADI][ARR] A3 ;
atera[ADI][ARR]:atera[ADI][ARR][Ebih] A5 ;
egin[ADI][ARR] A1 ;
ekar[ADI][ARR]:ekar[ADI][ARR][Rgog] A6 ;
sinets[ADI][ARR] A4 ;
```

Adjektiboak

“Adjektiboak” azpilexiko zabalean izenondoak [*IZO*] eta izenlagunak [*IZL*] bildu dira.

```
LEXICON adjektiboak
eder[ADJ][IZO]:eder[ADJ][IZO][Rgog] ADJK ;
merke[ADJ][IZO] ADJK ;
on[ADJ][IZO] ADJK3 ;
delako[ADJ][IZL] I ;
```

Izenordainak

Izenordainen kasuan alomorfoetara jo beharra izan dugu tarteka. *Zuek*-en kasuan ikusten denez, ezkerreko aldean forma neutroa mantendu bada ere, bi puntuen ondoren alomorfoa bera ageri da, bi mailatako forma: *zue*. Honekin, *zuek* vs. *zuei* modukoak konponduko ditugu.

```
LEXICON izenordainak
gu[IOR][PER] IO1 ;
zuek[IOR][PER] ABS ;
zuek[IOR][PER]:zue[IOR][PER] IO3 ;
edonor[IOR][IZG][GAL]:edonor[IOR][IZG][GAL][EpBer] NOR ;
elkar[IOR][ELK]:elkar[IOR][ELK][Rgog] ELK ;
```

Determinatzaileak

Hau determinatzaile erakuslearen kasuan ere alomorfoa erabili behar izan da: *hon*. Esan dugunez, analisiaren emaitzak *hau* emango du, baina morfotaktika *hon*-etik bideratzen du: *hau* vs. *honek*.

```
LEXICON determinatzaileak
hau[DET][ERK][MSL] ABS ;
hau[DET][ERK][MSL]:hon[DET][ERK][MSL] HAU ;
hiru[DET][ZNB][DZH]:hiru[DET][ZNB][DZH][Rzah] ZNB ;
edozer[DET][ZNB][ZHG]:edozer[DET][ZNB][ZHG][EpBer] IOZ ;
guzti[DET][ZNB][ORO] I4 ;
```

Adberbioak

Adberbioen multzoan, morfotaktika ia kasuan-kasuan zehaztu beharrekoa da, ondoko atalean ikusiko dugunez.

```
LEXICON adberbioak
arras[ADB][ADO] I0 ;
asko[ADB][ADO] I4_0 ;
aspaldi[ADB][ADO] I12 ;
berandu[ADB][ADO] I5 ;
berehala[ADB][ADO] KO_I0 ;
ederki[ADB][ADO] I0 ;
```

Beste kategoria nagusiak

Hainbat kasu desberdin biltzen ditu “lemak” azpilexikoak: posposizioak *[POS]*, partikulak *[PRT]*, lotura-elementuak *[LOT]* (juntagailuak, lokailuak eta aske diren menderagailuak), esklamazioak *[ESK]*, etab. Hitz anitzeko unitate lexikalak ere atal honetara ekarri ditugu: *adarra jo* aditza da, baina horren aurretik esaera *[ESR]* kategoria nagusia du. Gauza bera gertatzen da *behinik behin* formarekin, berez aditzondoa den arren.

```
LEXICON lemak
kanpo[POS] LK_KO ;
ei[PRT] I0 ;
baina[LOT][JNT] I0 ;
eta[LOT][JNT] I0 ;
haatik[LOT][LOK] I0 ;
alajaina[ESK] I0 ;
adarra_jo[ESR][ADI] A5 ;
behinik-behin[ESR][ADB][ADO] I0 ;
```

Kategoria lagungarriak

Adizki jokatuak

Aditz laguntzaileak eta trinkoak, jokatuak, alegia, zerrendatu egin ditugu gure sisteman, morfotaktikaren bidez lantzea posible zen arren, konplexua baitzen. Azpilexiko honetan bildu diren adizkiek atzizkiak har ditzakete ondoren, *LAT* jarraitze-klaseak horixe esan nahi baitu. Hain zuzen ere, *a15b178* azpilexiko-izenak adierazten du A1, A5, B1, B7 eta B8 taldeetakoak biltzen direla hemen: A1 (*da/du* motakoak: *izan/*edun*-en orainaldia), A5 (*daiteke/dezake*), B1 (*zen/zuen*), B7 (*liteke/lezake*) eta B8 (*zitekeen/zezakeen*).

```
LEXICON a15b178
da[ADL] LAT ;
da[ADT] LAT ;
dio[ADL] LAT ;
dakit[ADT] LAT ;
dakarkio[ADT] LAT ;
```

```
darama[ADT]:darama[ADT] [Aadi] LAT ;
dator[ADT]:dator[ADT] [Rgog] LAT ;
nentorren[ADT] LAT ;
```

Siglak

Siglak ere bereiz jaso ditugu, nahiz morfotaktikaren aldetik izenen (berezien, erakundeenak batez ere) portaera bera duten.

```
LEXICON siglak
EDBL[SIG] [IZE] [IZB]:9e9d9b91[SIG] [IZE] [IZB] [SigB] LIB_IB ;
UEU[SIG] [IZE] [IZB]:9e9h9u[SIG] [IZE] [IZB] LIB_IB ;
```

Kategoria morfosintaktikoak

Morfema ez-askeen kasuan morfema-muga (+) ezarri beharko dugu aurretik (atzizkia bada) edo atzetik (aurrizkia bada). Baina formalismoa kategoria nagusien bera da. Ia morfema bakoitzak bere azpilexikoa behar du, kasu gutxi batzuetan izan ezik. Guk, halere, kategoria-sisteman ikusi dugun sailkapenaren arabera aurkeztuko ditugu, nahiz bakoitzak bere azpilexiko-izena izango duen. Morfema hauek hobeto ulertzeko, bakoitzaren ondoren adibide bat erantsi dugu.

Mugatasuna/numeroa

Informazio morfologiko bera eskaintzeko azpilexiko desberdinak sortu behar izan ditugu hainbat kasutan, mugatasunean batez ere. Honen arrazoia sinplea da: mugatasunak eta numeroak bideratzen dituzte deklinabide-kasuak.

```
LEXICON adb_a
+a ABS ; !seguruen a

LEXICON mg
+0 [MUN] [MG] MG ; !etxe 0

LEXICON mg2
+0 [MUN] [MG] MG5 ; !handi 0
```

```
LEXICON mg_bat
+0 [MUN] [MG] MG1 ; !bat 0

LEXICON mg_biz
+0 [MUN] [MG] MG3 ; !haur 0

LEXICON mg_zein
+0 [MUN] [MG] MG4 ; !zein 0, edozein 0

LEXICON ms
+a [MUN] [MS] MS ; !etxe a

LEXICON ms_bat
+a [MUN] [MS] MS1 ; !bat a

LEXICON ms_biz
+a [MUN] [MS] MS3 ; !haur(r) a
```

```

LEXICON ms elk
+0 [MUN] [MS] MS4 ; !elkar 0

LEXICON ms lekden
+0 [MUN] [MS] MS2 ; !etxe 0 (ko,...)

LEXICON ms zein
+a [MUN] [MS] MS5 ; !zein a

LEXICON mp
+e [MUN] [MP] MP ; !etxe e (k, ta-n)

LEXICON mph
+o [MUN] [MPH] MP ; !etxe o (k, ta-n)

LEXICON mph2
+o [MUN] [MPH] MP2 ; !handi o (ta-n, en-gan)

LEXICON mph abs
+ok [MUN] [MPH] ABS ; !etxe ok

LEXICON mph biz
+o [MUN] [MPH] MP3 ; !haur(r) o (k, en-gan)

LEXICON mp abs
+ak [MUN] [MP] ABS ; !etxe ak (0)

LEXICON mp batzu
+e [MUN] [MP] MP2 ; !batzu e (k)

LEXICON mp biz
+e [MUN] [MP] MP3 ; !haur(r) e (k, en-gan)

```

Mg hasitako azpilexikoak kontuan hartzen baditugu, hau ikus dezakegu: *mg* soila izen arrunt bizigabeek hartzen dute; *mg_biz*, berriz, izen arrunt bizidunek; *mg2* adjektiboek, bizidun zein bizigabeen deklinabidea har baitzakete; *mg_bat bat*-ek hartzen du, mugagabe eta mugatu izateaz gain, portaera berezia duelako, alegia, ez du partitiborik hartzen⁴⁷ eta leku-denborazko kasuak zuzenean, *ta*-tik pasata zein genitibotik abiatuta osa ditzake; *mg_zein*, azkenik, *zein* eta bere eratorriek hartzen dutena da, determinatzaile izanda ere, mugagabe zein mugatu izan baitaitezke.

Deklinabidea

Mugatasunean alomorfoak erabili dira, informazio morfologiko bera eskaintzeko, batzuetan azpilexikoak bikoiztu (edo laukoiztu) behar izan dira. Deklinabidean kasuren batean besterik ez da halakorik gertatzen, informazioa errepikatzen bada ere, osagaiak desberdinak dira-eta: alegia, aldaerak. Instrumentalerako, adibidez, hiru azpilexiko sortu ditugu, hiru aldaera jasotzeko: *-z*, *-taz* eta *-zaz*.

```

LEXICON abs
+0 [DEK] [ABS] IO ; !etxe 0 0, etxe a 0, etxe ak 0

LEXICON par
+ik [DEK] [PAR] :+ik [DEK] [PAR] [Rep] IO ; !etxe 0 (r)ik, zuhaitz 0 ik

LEXICON erg
+k [DEK] [ERG] :+k [DEK] [ERG] [Eep] IO ; !etxe 0 k, ume a k, zuhaitz 0 (e)k,
gu k, AEK-k

LEXICON dat
+i [DEK] [DAT] :+i [DEK] [DAT] [Rep] IO ; !zuhaitz 0 i, ume 0 (r)i, zuhaitz a
(r)i, zu (r)i, elkar (r)i, AEK
(r)i, zuhaitz e i, ume o i

```

⁴⁷ Egia da *batik bat* modukoak baditugula, baina hori hitz anitzeko unitate lexikaltzat hartuko dugu eta, beraz, sarrera propioa izango du.


```

LEXICON ine
+n[DEK] [INE] IO ; !etxe 0 a n, zuhaitz e ta n

LEXICON ine_biz
+gan[DEK] [INE] IO ; !seme a ((r)en) gan, zu ((r)e) gan

LEXICON gel
+ko[DEK] [GEL] :+ko[DEK] [GEL] [Eep] I1_ELIP ; !etxe 0 ko, zuhaitz e ta ko

LEXICON gel2
+ko[DEK] [GEL] I1_ELIP ; !gaur ko, lan a ekin ko

LEXICON ala
+ra[DEK] [ALA] :+ra[DEK] [ALA] [Eep] KO_IO ; !etxe 0 ra, mendi e ta ra

LEXICON ala_biz
+gana[DEK] [ALA] KO_IO ; !lagun a ((r)en) gana

LEXICON abz
+rantz[DEK] [ABZ] :+rantz[DEK] [ABZ] [Eep] AGO_EGI ; !etxe 0 rantz, mendi e ta rantz

LEXICON abz_biz
+ganantz[DEK] [ABZ] AGO_EGI ; !lagun a (ren) ganantz

LEXICON abu
+raino[DEK] [ABU] :+raino[DEK] [ABU] [Eep] KO_IO ; !etxe raino

LEXICON abu2
+daino[DEK] [ABU] KO_IO ; !gaur daino

LEXICON abu_biz
+ganaino[DEK] [ABU] KO_IO ; !lagun a (ren) ganaino

LEXICON abu3
+rarte[DEK] [ABU] :+rarte[DEK] [ABU] [Eep] KO_IO ; !hurrengorarte

LEXICON abl
+tik[DEK] [ABL] :+tik[DEK] [ABL] [Eep] KO_IO ; !zuhaitzetik, gaurtik

LEXICON abl_biz
+gandik[DEK] [ABL] KO_IO ; !ama(ren)gandik

LEXICON abl2
+danik[DEK] [ABL] KO_IO ; !gaurdanik

LEXICON abl3
+ik[DEK] [ABL] :+ik[DEK] [ABL] [Rep] KO_IO ; !seitarik

LEXICON gen
+en[DEK] [GEN] :+en[DEK] [GEN] [Rep] GN ; !etxearen

LEXICON gen_biz
+en[DEK] [GEN] :+en[DEK] [GEN] [Rep] GN_biz ; !haurraren

LEXICON gen_ber
+e[DEK] [GEN] :+e[DEK] [GEN] [Rep] GN_biz ; !gure, bere

LEXICON soz
+ekin[DEK] [SOZ] :+ekin[DEK] [SOZ] [Rep] KO_IO ; !zurekin, hurrekin

LEXICON ins
+z[DEK] [INS] :+z[DEK] [INS] [Eep] IO ; !paperez, alabaz, batez, zerez

LEXICON ins1
+taz[DEK] [INS] IO ; !horretaz, gutaz, zertaz

LEXICON ins2
+zaz[DEK] [INS] IO ; !horrezaz, nihaurezaz

LEXICON mot
+gatik[DEK] [MOT] :+gatik[DEK] [MOT] [Eep] KO_IO ; !semea(ren)gatik, zu(re)gatik

LEXICON des
+entzat[DEK] [DES] :+entzat[DEK] [DES] [Rep] KO_IO ; !semearentzat, haranarentzat

LEXICON des1
+etzat[DEK] [DES] :+etzat[DEK] [DES] [Rep] KO_IO ; !zuretzat, guretzat

LEXICON des2

```

```

+endako[DEK][DES]:+endako[DEK][DES][Rep] I0 ;      !semearendako, haranarendako
LEXICON des3
+etako[DEK][DES]:+etako[DEK][DES][Rep] I0 ;      !zuretako, guretako
LEXICON des4
+ko[DEK][DES] I0 ;      !haraneko, zertako
LEXICON des5
+rako[DEK][DES]:+rako[DEK][DES][Eep] I0 ;      !haranerako, zertarako
LEXICON des6
+ganako[DEK][DES] I0 ;      !semea(ren)ganako, harenganako
LEXICON pro
+tzat[DEK][PRO] KO_I0 ;      !semetzat, gizontzat, Eibartzat
LEXICON pro2
+tako[DEK][PRO] I0 ;      !semetako, gizontako, Eibartako
LEXICON ban
+ko[DEK][BAN]:+ko[DEK][BAN][Eep] I0 ;      !semeko, haraneko, hiruko
LEXICON a_ine
+a:+a[Eep] INE ;      !etxe 0 a n, haran 0 (e)a n
LEXICON ta
+ta:+ta[Eep] LK1_KO ;      !etxe 0 ta n, haran 0 e ta ra
LEXICON ta2
+ta:+ta[Eep] LK2_KO ;      !gu ta ko, gu ta rik, gu ta tik

```

Gradua

Gradua aurreko kategoriaren arabera bikoiztu da: adjektiboez hartzen dutena (*adj_ago...*), adberbioez (*adb_ago...*). Horrez gain, deklinabide-kasuen ondokoentzat hirugarren bat behar izan da (*ago_ko...*).

```

LEXICON adb_ago
+ago[GRA][KONP] I6 ;      !beranduago (-ko, ...)
LEXICON adb_ago_0
+ago[GRA][KONP] I0 ;      !ederkiago; ikustenago
LEXICON adb_egi
+egi[GRA][GEHI] I6 ;      !beranduegi (-ko, ...)
LEXICON adb_egi_0
+egi[GRA][GEHI] I0 ;      !ederkiegi
LEXICON adb_en
+en[GRA][SUPE] I6_A_IK ;      !beranduen (-ko, -a, -ik, ...)
LEXICON adb_en_ik
+en[GRA][SUPE] A_IK ;      !ederkien (-a, -ik)
LEXICON adj_ago
+ago[GRA][KONP] I2 ;      !handiago, txikitxoago
LEXICON adj_egi
+egi[GRA][GEHI] I2 ;      !handiegi, txikitxoegi, onegi
LEXICON adj_en
+en[GRA][SUPE] I2 ;      !handien, txikitxoan, onen
LEXICON ago_ko
+ago[GRA][KONP] KO_I0 ;      !harantzago(ko)
LEXICON egi_ko
+egi[GRA][GEHI] KO_I0 ;      !harantzegi(ko)
LEXICON en_ko
+en[GRA][SUPE] KO_I0 ;      !gehiago, gutxiago

```

Aditz-mota, aspektua

Aditzoinak beti 0 marka du, baina ondoren har ditzakeen morfemak desberdinak dira aditz-motaren arabera. *-te* eta *-tze* bi atzizki desberdintzat hartu ditugu, Mitxelenaren bidetik⁴⁸.

```
LEXICON adoin
+0 [ADM] [ADOIN] IO ; !nahi

LEXICON adoin1
+0 [ADM] [ADOIN] TEN ; !esan, ikus

LEXICON adoin2
+0 [ADM] [ADOIN] TZEN ; !etor, pentsa

LEXICON adoin3
+0 [ADM] [ADOIN] TEN_KI ; !izan

LEXICON adoin4
+0 [ADM] [ADOIN] TZEN_KI ; !ibili

LEXICON partiz
+0 [ADM] [PART] PARTI ; !nahi, esan

LEXICON partiz1
+i [ADM] [PART] PARTI ; !etorri

LEXICON partiz2
+tu [ADM] [PART] PARTI ; !pentsatu

LEXICON partiz3
+0 [ADM] [PART] GERO_KO ; !ahal, ari

LEXICON adi_ize
+te [ADM] [ADI_IZE] I2 ; !egite

LEXICON adi_ize2
+tze [ADM] [ADI_IZE] I2 ; !nekatze

LEXICON ezbu
+ten [ASP] [EZBU] AGO ; !erosten dut; esaten(ago) du

LEXICON ezbu2
+tzen [ASP] [EZBU] AGO ; !etortzen da, nekatzen(ago) naiz

LEXICON buru
+0 [ASP] [BURU] AGO ; !esan du, etorri da, nekatu(ago?) da

LEXICON gero
+ko [ASP] [GERO] IO ; !esango du, etorriko da, nekatuko naiz

LEXICON gero1
+en [ASP] [GERO] :+en [ASP] [GERO] [Rep] IO ; !esanen du, etorriren da
```

Menderagailuak: adizki jokatuak

la-n azpilexikoak, bakarrak, adizki jokatuak atzetik har ditzaketen osagai guztiak biltzen ditu.

```
LEXICON ba_bald
ba [ERL] [MEN] [BALD] + BABALD ;
! (baldin) banaiz, banadi
```

⁴⁸ “En algunos dialectos *-te* y *-tze* alternan en el sustantivo verbal; guip. *egite* de *egin*, pero *gertatze* de *gertatu*, etc. No debe pensarse, sin embargo, que *tz* sea el resultado de *t+t*: se trata probablemente de sufijos de distinto origen.” (Mitxelena, 1961, 346. or.).

```

LEXICON bait
bait[ERL] [MEN] [KAUS] + BAIT ;
    !erosi egin dut, merkea baitzen

LEXICON la_n
+la[ERL] [MEN] [KNPL] :+la[ERL] [MEN] [KNPL] [Eep] KO_I0 ;
    !Etorriko dela esan du
+lako[ERL] [MEN] [KAUS] :+lako[ERL] [MEN] [KAUS] [Eep] IO ;
    !Lo hartu dudalako etorri naiz berandu
+na[ERL] [MEN] [KNPL] :+na[ERL] [MEN] [KNPL] [Eep] IO ;
    !Badakit etorriko zarena
+nik[ERL] [MEN] [KNPL] :+nik[ERL] [MEN] [KNPL] [Eep] IO ;
    !Ez zuen esan etorriko zenik
+la[ERL] [MEN] [MODU] :+la[ERL] [MEN] [MODU] [Eep] IO ;
    !Frakak urratuta zituela etorri zen
+larik[ERL] [MEN] [MODU] :+larik[ERL] [MEN] [MODU] [Eep] IO ;
    !Frakak urratuta zituelarik etorri zen
+la[ERL] [MEN] [DENB] :+la[ERL] [MEN] [DENB] [Eep] IO ;
    !Autobusean zetorrela gogoratu zen
+larik[ERL] [MEN] [DENB] :+larik[ERL] [MEN] [DENB] [Eep] IO ;
    !Etxean sartu delarik, lapurrak ibili direla ohartu da.
+n[ERL] [MEN] [ERLT] :+n[ERL] [MEN] [ERLT] [Eep] AL1 ;
    !Gehien aspertzen nauen pertsona etorri da
+nean[ERL] [MEN] [DENB] :+nean[ERL] [MEN] [DENB] [Eep] IO ;
    !Etorri naizenean euria ari zuen
+nean[ERL] [MEN] [BALD] :+nean[ERL] [MEN] [BALD] [Eep] IO ;
    !Zuk esaten duzunean, egia izango da
+nean[ERL] [MEN] [KAUS] :+nean[ERL] [MEN] [KAUS] [Eep] IO ;
    !Pozik egon daitezke suak dena kiskali ez duenean
+neko[ERL] [MEN] [DENB] :+neko[ERL] [MEN] [DENB] [Eep] IO ;
    !Ikusi zueneko hil zuen erbia
+nerako[ERL] [MEN] [DENB] :+nerako[ERL] [MEN] [DENB] [Eep] IO ;
    !Jon datorrenerako, hor izango naiz
+nera[ERL] [MEN] [DENB] :+nera[ERL] [MEN] [DENB] [Eep] IO ;
    !Hemendik datorrenera, afaria prestatzeko astia izango dut
+netik[ERL] [MEN] [DENB] :+netik[ERL] [MEN] [DENB] [Eep] IO ;
    !Ezkondutik denetik aldatuta dago
+naz[ERL] [MEN] [MODU] :+naz[ERL] [MEN] [MODU] [Eep] IO ;
    !Ikusten duzunaz, bueltatu naiz
+nez[ERL] [MEN] [MODU] :+nez[ERL] [MEN] [MODU] [Eep] IO ;
    !Ikusten duzunez, bueltatu naiz
+nez[ERL] [MEN] [KAUS] :+nez[ERL] [MEN] [KAUS] [Eep] IO ;
    !Euria ari duenez, ez gara aterako
+n[ERL] [MEN] [ZGAL] :+n[ERL] [MEN] [ZGAL] [Eep] IO ;
    !Ikusi duzun jakin nahi dut
+nentz[ERL] [MEN] [ZGAL] :+nentz[ERL] [MEN] [ZGAL] [Eep] IO ;
    !Ez diguzu esan etorriko zarenentz
+nez[ERL] [MEN] [ZGAL] :+nez[ERL] [MEN] [ZGAL] [Eep] IO ;
    !Ez diguzu esan etorriko zarenez

```

Menderagailuak: aditz nagusiak

tea eta *tzea* azpilexikoek aditz nagusi jokatu gabeen morfotaktikaren atalik handiena biltzen dute. Ondoren datozen *ki_l* eta *ki_n* besterik ez dira falta. Azken azpilexikoak (*ik_ta_z* eta *agatik*), aldiz, partizipioaren menderagailuak dira, moduzkoak, denborazko bat eta kontzesiboa sortzeko, alegia.

```

LEXICON tea
+tea[ERL] [MEN] [KNPL] IO ;      !Komeni da berria zuk ematea
+teak[ERL] [MEN] [KNPL] IO ;     !Tranpa egiteak min eman zidan
+teari[ERL] [MEN] [KNPL] IO ;    !Beldurra dio bakarrik egoteari
+tean[ERL] [MEN] [DENB] IO ;     !Etxera iristean kontatuko dizut guztia
+ten[ERL] [MEN] [MODU] IO ;      !Porrusalda egiten ahaztu zaio

```

+ten[ERL] [MEN] [KNPL] IO ;	!Ane ikasten ari da; esaten erraza da, egiten zaila
+teko[ERL] [MEN] [HELB] IO ;	!Gauzak ondo egiteko arreta jarri behar da
+teko[ERL] [MEN] [KNPL] IO ;	!Amak etxera joateko esan du
+tekotan[ERL] [MEN] [BALD] IO ;	!Biok joatekotan erosiko nuke, bestela ez
+tekotan[ERL] [MEN] [DENB] IO ;	!Irtetekotan ginen, euria hasi zuelarik
+tekotan[ERL] [MEN] [HELB] IO ;	!Hara joatekotan erosi dut
+tera[ERL] [MEN] [BALD] IO ;	!Jakitera, han gelditu nintzatekeen
+tera[ERL] [MEN] [HELB] IO ;	!Ogia eroatera joan da
+terako[ERL] [MEN] [DENB] IO ;	!Zu irteterako, hamarrak izango dira
+terakoan[ERL] [MEN] [DENB] IO ;	!Etxetik irteterakoan ikusi zuen
+teagatik[ERL] [MEN] [KAUS] IO ;	!Eskolan irratia entzuteagatik zigortu ninduten
+teagatik[ERL] [MEN] [HELB] IO ;	!Biharko zure hitzaldia entzuteagatik etorri naiz
+tearren[ERL] [MEN] [KAUS] IO ;	!Eskolan irratia entzutearren zigortu ninduten
+tearren[ERL] [MEN] [HELB] IO ;	!Biharko hitzaldia entzutearren etorri naiz
LEXICON tzea	
+tzea[ERL] [MEN] [KNPL] IO ;	!Ondo legoke zu ere etortzea
+tzeak[ERL] [MEN] [KNPL] IO ;	!Goiz jaikitzeak asko laguntzen du
+tzeari[ERL] [MEN] [KNPL] IO ;	!Lapurtzeari gaizki deritzot
+tzean[ERL] [MEN] [DENB] IO ;	!Lanetik ateratzean joango natzaizu
+tzen[ERL] [MEN] [MODU] IO ;	!Zenbatzen ahaztu zaio
+tzen[ERL] [MEN] [KNPL] IO ;	!Ane pentsatzen ari da; ulertzen erraza da, azaltzen zaila
+tzeko[ERL] [MEN] [HELB] IO ;	!Gauzak ondo ulertzeko arreta jarri behar da
+tzeko[ERL] [MEN] [KNPL] IO ;	!Amak etxera etortzeko esan du
+tzekotan[ERL] [MEN] [BALD] IO ;	!Zu etortzekotan erosiko nuke, bestela ez
+tzekotan[ERL] [MEN] [DENB] IO ;	!Abiatzekotan ginen, euria hasi zuelarik
+tzekotan[ERL] [MEN] [HELB] IO ;	!Hona etortzekotan jaiki naiz
+tzera[ERL] [MEN] [BALD] IO ;	!Susmatzera, han gelditu nintzatekeen
+tzera[ERL] [MEN] [HELB] IO ;	!Ogia ekartzera joan da
+tzerako[ERL] [MEN] [DENB] IO ;	!Zu etxeratzerako, hamarrak izango dira
+tzerakoan[ERL] [MEN] [DENB] IO ;	!Etxera etortzerakoan esan zigun
+tzeagatik[ERL] [MEN] [KAUS] IO ;	!Zu agurtzeagatik galdu nuen trena
+tzeagatik[ERL] [MEN] [HELB] IO ;	!Gero belarretan laguntzeagatik eman dizut dirua
+tzearren[ERL] [MEN] [KAUS] IO ;	!Zu agurtzearren galdu nuen trena
+tzearren[ERL] [MEN] [HELB] IO ;	!Gero belarretan laguntzearren eman dizut dirua
LEXICON ki_1	
+ki[ERL] [MEN] [MODU] IO ;	!Kontu handiz ibilki lortuko duzu zure helburua
LEXICON ki_n	
+ki[ERL] [MEN] [MODU] IO ;	!Hala izaki, guztia jakinki, goazen aurrera
LEXICON ik_ta_z	
+ik[ERL] [MEN] [MODU] :+ik[ERL] [MEN] [MODU] [Rep] KO_IO ;	!Bizitza arriskaturik etorri zen
+ta[ERL] [MEN] [MODU] KO_IO ;	!Bizitza arriskatuta etorri zen gugana
+takoan[ERL] [MEN] [DENB] IO ;	!Gela txukundutakoan, komuna garbitu
+z[ERL] [MEN] [MODU] IO ;	!Ohitura jatorrak galduz doaz
+az[ERL] [MEN] [MODU] IO ;	!Zurekin hitz eginaz asko ikasi dut
LEXICON agatik	
+agatik[ERL] [MEN] [KONT] IO ;	!Zuk esanagatik ez dut sinisten

ba- baieztapen-partikula ez-askea

Kategoria nagusien artean ere aipatu dugu *ba* baieztapenekoa, baina bitara jokatzeko du honek: morfema aske eta ez-aske bezala. Horregatik bikoiztu behar izan dugu, morfotaktikan eragina du-eta: *ba* askeari ez zaio morfemarik erantsen ondoren (*ba omen dator*), baina ez-askearen atzetik adizki jokatuak datoz (*badator*).

LEXICON bai	
ba[PRT] + BABAI ;	!bai, badakit; badatorrela

Elipsia

Elipsia zero morfema bat da, informazio lexikoa (izen eliditua) bakarrik eskaini eta izenen morfotaktika bideratzen duena.

```
LEXICON eli_ize
+- [ELI_IZE] ELIP ;          !umearen 0 a, txikiko 0 ek, dakidan 0 txo a
```

Elkarketa-marra

Elkarketa-marrak ez du inongo informazio morfologikorik ematen, baina ‘-’ osagaiaren ondoren *izenak* joan daitezke (izen+izen moduko elkarteak besterik ez ditugu onartzen, oraingoz).

```
LEXICON elk_marra
+- [ELM] izenak ;          !botoi-zulo, serora-etxe, kale-mutiko
```

Marratxoa

Marratxoa ere elementu grafiko hutsa da, baina aldaketa morfofonologikoak baldintzatzen ditu. Ez da berdin *surferrak* edo *surfer-ak*, esaterako: atzizkia lotzean *r* bikoizten da, salbu marratxoa tartekatzen denean.

```
LEXICON marra1
+- [MAR] IB1 ;          !Ronan-ek, Descartes-ek

LEXICON marra2
+- [MAR] LIB ;          !Alger-eko

LEXICON marra3
+- [MAR] LIBK_MARRA ;    !AEK-k, AEK-ko (baina AEKri)

LEXICON marra4
+- [MAR] I1 ;          !cm-a, stock-ean

LEXICON marra5
+- [MAR] I1_BIZ ;        !surfer-ak

LEXICON marra6
+- [MAR] LIB_IB2 ;        !Peugeot-ek
```

Aurrizki lexikalak

Oraingoz landu ditugun bi aurrizki lexikalek informazio bera dute. Har dezagun kontuan *ber-* aurrizkiaren jarraitze-klaseak adierazten duena: aditzak (nagusiak) har ditzake ondoren.

```
LEXICON ber
ber [AUR] :ber [AUR] [Rgog] + ADITZAK ;    !berregin, birmargotu

LEXICON ez
Ez [AUR] + EZ_JK ;          !ez-uste, ez-entzun
```

Atzizki lexikalak

```

LEXICON garren
+garren[ATZ][ORD] I1_ELIP ;                !bigarren, zenbatgarren

LEXICON na1
+na[ATZ][DIS] I0 ;                          !bina liburu

LEXICON na2
+na[ATZ][DIS] ABS ;                         !liburu bana

LEXICON txo
+txo[ATZ][TXIK] TXO ;                       !etxetxo

LEXICON txo_biz
+txo[ATZ][TXIK] TXO1 ;                      !haurtxo

LEXICON adj_txo
+txo[ATZ][TXIK] ADJK2 ;                     !txikitxo

LEXICON ño
+ño[ATZ][TXIK] TXO ;                        !etxeño

LEXICON ño_biz
+ño[ATZ][TXIK] TXO1 ;                       !haurño

LEXICON adj_ño
+ño[ATZ][TXIK] ADJK2 ;                      !txikiño

LEXICON tzar
+tzar[ATZ][HAND]:+tzar[ATZ][HAND][Rgog] TXO ; !etxetzar

LEXICON tzar_biz
+tzar[ATZ][HAND]:+tzar[ATZ][HAND][Rgog] TXO1 ; !gizontzar

LEXICON adj_tzar
+tzar[ATZ][HAND]:+tzar[ATZ][HAND][Rgog] ADJK2 ; !tontotzar

LEXICON araz
+araz[ATZ][FAK] A12 ;                       !agerraraz, ...

LEXICON te
+te[ATZ][IZE] I2 ;                           !egite

LEXICON tze
+tze[ATZ][IZE] I2 ;                           !egokitze

```

5.2.3 Morfotaktika

Sistema lexiko osoa definitu eta antolatu ondoren, morfotaktikaren berri emateko jarraitze-klaseak beharko ditugu, alegia, lema edo morfema batek atzetik har ditzakeen morfemak —eta horiek bakarrik— bilduko dituztenak. Neurri batean gramatika —hitzaren gramatika— osatuko dugu fase honetan.

Jarraitze-klase baten osagaiak morfema ez-askeak lantzean aipatu ditugun azpilexikoak dira, baina hauekin batera definitua dagoen beste jarraitze-klase bat ere sar dezakegu. Azpilexikoak minuskulaz azalduko dira eta jarraitze-klaseak maiuskulaz. Ezkerretara datorrena jarraitze-klasearen izena da, azpilexiko guztiek eramango dutena, eta eskuinetara datorrena, berriz, jarraitze-klase hori osatzen duten osagaien zerrenda —azpilexikoak eta beste jarraitze-klase batzuk, hainbat kasutan—.

Gure sistemarako 120 bat jarraitze-klase sortu ditugu. Beti gertatu ohi den bezala, hauetako gutxi batzuek izen guztien morfotaktikaren berri eman badezakete ere, adberbioak lantzeko, adibidez, mordoxka bat sortu behar izan dugu, kasuan-kasuan aztertu eta bakoitzak har zitzakeen morfemak bereizi behar izan ditugulako.

Alomorfoek aukera ematen dute, morfotaktika konplexuegi egin gabe, informazio guztiaren berri emateko. Baina, gure ustez, linguistikoki gardenagoa —eta txukunagoa— da gramatikak berak konpontzea, lexikoa ukitu gabe, nahiz jarraitze-klaseen kopuruak gora egingo duen ezinbestean. Halere, zenbait kasutan ezinbestean baliatu beharko ditugu.

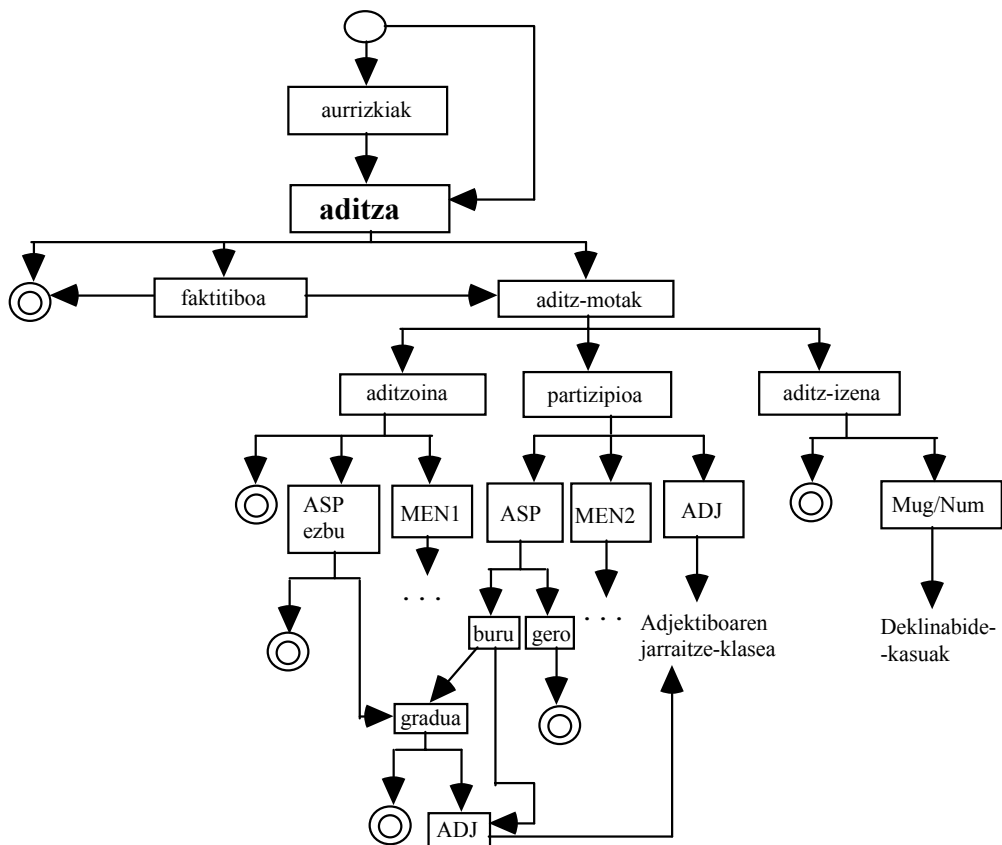
Morfotaktika osoaren berri hemen ematea ezinezkoa denez, adibide erakusgarri batzuk hautatu ditugu, horien bitartez hurbilpen bat eginez. Aditzak, izen arrunt bizigabeak eta adberbioak ikusiko ditugu adibideen bidez.

5.2.3.1 Aditzak

Aditz nagusien kasuan, esaterako, aditz-motaren araberako sailkapena egiten badugu, konbinazio guztiak kontuan hartuta hamabi jarraitze-klase beharko dira (aditz defektiboak barne hartuta, noski). Sailkapen erregular zabala hau da, azpilexikoetan banatu duguna:

Aditza (bukaera)	Aditzoina	Partizipioa	Aditz-izena	Adibidea
-N	0	0	te	<i>egin, izan</i>
-I	0	i	tze	<i>ireki, bizi</i>
AlboSud, Dard.	0	i	tze	<i>bidali, eskaini, jarri</i>
Txis, Afr.	0	i	te	<i>ikusi, idatzi</i>
Bokal, LehGor	0	tu	tze	<i>pentsatu, lehertu</i>
Txis, Afr.	0	tu	te	<i>abestu, hoztu</i>

Baina hamabiek erantzuten diete eskema berari. 9. irudian agertzen da grafo baten bidez adierazita.



9. irudia.- Aditz nagusiaren morfotaktika

Adibidez, *egin* motakoak (*egin, egin, egite*), A1 jarraitze-klasea dute:

```
LEXICON A1
I0 ;                !bukaera
araz ;              !faktitiboa
adoin1 ;            !aditzoia
partiz ;            !partizipioa
adi_ize ;           !aditz-izena
```

Lexikoa ikusitakoaren arabera, *I0* jarraitze-klase bat da (letra larrian baitago), elementu terminala duena (#), hau da, ondoren ez du ezer hartzen. Beste elementu guztiak azpilexikoak dira, aurretik definitu ditugunak: *araz* azpilexikoa *araz*[ATZ][FAK] dela esan dugu, atzizkia, aditzak faktitibo bihurtzen dituen. Honek, aldi berean, *A12* jarraitze-klasea hartuko du, hain zuzen ere beste faktitibo bat aplikatzen utziko ez duena. *Adoin1* azpilexikoa *0* morfema bat da, [ADM][ADOIN] informazioa duena (aditz-mota: aditzoia) eta jarraitze-klaseak dioskunez, *TEN* baita, aspektua *ten*-etik bideratzen duena: *egiten*, ez *-tzen*. *Partiz* azpilexikoa ere *0* morfema da, partizipioa inongo morfema espliziturik gabe bideratzen baita, nahiz [ADM][PART] informazioa gehitzen duen. Azkenik, *adi_ize* azpilexikoa, aditz-izenak bideratzen dituen, *te* morfemaren bidez gauzatzen da, [ADM][ADI_IZE] informazioarekin. Honen jarraitze-klasean *I2* dugu, alegia, izenen morfotaktika (murriztapen batzuekin) baliatuko duena.

Azpilexiko hauek sortuko dira *egin* motako aditzek har ditzaketen forma posible guztiak.

Honela zatika daiteke jarraitze-klasea, modu grafikoagoan:

egin	0 (I0)		<i>egin</i>
	araz (araz)	i, ten, ...	<i>eginaraz, eginarazi, eginarazten</i>
	0 (adoin1)	ten (asp ezbu)	<i>egin (dezaket), egiten</i>
		tea, teagatik, tean, tearren, teko,...	<i>egitea, egiteagatik, egiteko...</i>

	0 (partiz)	0 (asp buru)	<i>egin (dut)</i>
		ko (asp gero)	<i>egingo</i>
		en (asp gero1)	<i>eginen</i>
		agatik, ik, z, ta,...	<i>eginagatik, eginik, eginda...</i>
	te (adi_ize)	0, a, ak, ari,...	<i>egite, egitea, egiteak...</i>

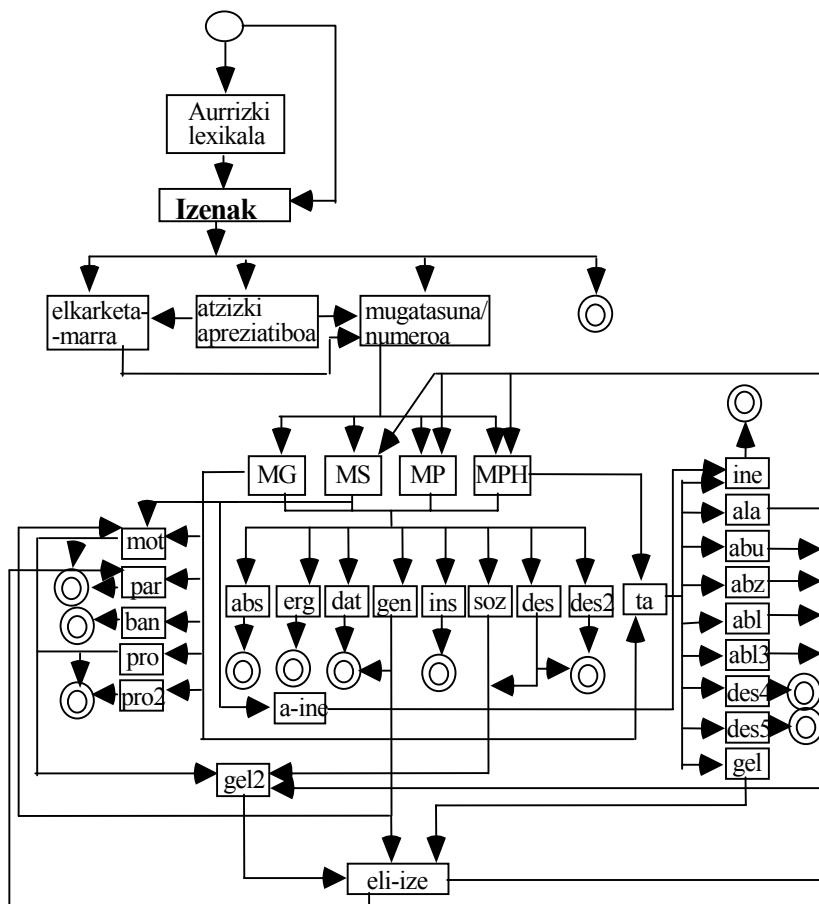
edan, eman, entzun, eragin, eraman, erantzun, esan, irakin, jan, jardun multzo honetan landu dira. Erregela bidez konponduko da bukaerako *n*-ren galera: aditz-izena, aspektu burutugabea eta menderagailuak hartzean (*-tea, ...*) bukaerako *n* galtzen baita (*egin + -te: egite*).

5.2.3.2 Izen arrunt bizigabeak

Adibide zabalago batekin uler dezakegu sistema lexiko osoa: izenekin. Izenek hartzen dituzten jarraitze-klaseak nahiko modu sistematikoan bil daitezke.

Izen arruntei dagokienez, bizidun/bizigabe da banaketa nagusia. Izen bereziek, bestalde, mugatasun lexikala dutenez, zuzenean hartzen dituzte deklinabide-kasuak, mugatasunetik pasa gabe eta, hortaz, pertsona-, leku- eta erakunde-izenak banatuta nahikoa da.

Izen arrunt bizigabeen morfotaktikaren grafoa 10. irudian azaltzen da.



* abz-k ago-egi ere har dezake oraindik, baina ez dugu irudian sartu.

10. irudia.- izen arrunten morfotaktika.

Grafoan ikus dezakegunez, aurrizki lexikaletik has gaitezke izena osatzen, segidan izenari eransteke, edo izenetik beretik. Kasu honetan, izenaren morfotaktika interesatzen zaigunez, izenetik abiatuko gara. Hor lau aukera ditugu: a) izenari elkarketa-marra gehitu eta, beste izen bat erantsi, ondoren mugatasuna/numeroa bideratzeko, hirugarren elkarketa-osagairik ez da-eta onartzen; b) atzizki apreziatiboak, hemendik mugatasuna bideratzeko edo, elkarketa-marraren bidez, hitz-elkarketa abiarazteko; c) izena maila lexikoan utzi eta informazio morfologikorik erantsi ez (biribilek adierazten digutena). Hiru aukera hauek maila lexikoan ditugu, oraindik ez dugu analisi edo sorkuntza morfologikoa bideratu. Horretarako d) puntura joan beharko dugu, alegia, mugatasuna/numeroa aplikatzera, hortik bideratuko baitira deklinabide-kasuak. Hori bera, jarraitze-klasea ikusita, hau dugu:

LEXICON I

elk_marra ;	! etxe- (etxe-orratz)
txo ;	! etxetxo (etxetxoa, etxetxoetan...)
ño ;	! etxeño (etxeñoa, etxeñoetan...)
tzar ;	! etxetzar (etxetzarra, etxetzarretan...)
II ;	! etxe (etxea, etxetzarretan...)

Beraz, mugatasuna/numeroa (II jarraitze-klaseak bideratuko duena), hala irakur dezakegu grafoan:

Mugagabea (*MG*), mugatu singularra (*MS*), mugatu plurala (*MG*) eta mugatu plural hurbila (*MPH*). Nahitaez horietako batetik pasa beharra dago aurrera egiteko; alegia, izen arrunt bizigabeez ari garenez, mugatu egin behar ditugu kasua erantsi aurretik, eta hori ezinbestekoa da. Ondoren, horiek guztiek kasuak hartuko dituzte. Kasu batzuk terminalak izango dira, hau da, ez dute atzetik ezer hartuko, ez dute birdeklinabidea onartuko. Kasu gramatikalak (absolutiboa (*abs*), ergatiboa (*erg*) eta datiboa (*dat*), partitiboa (*par*), destinatiboa (aldaera bat izan ezik: *des*), instrumentala (*ins*), inesiboa (*ine*), banakaria (*ban*) eta prolatiboaren aldaera bat (*pro2*, *pro1*-ek leku-genitiboa har dezake atzetik) terminalak dira. Baina beste kasuek leku-genitiboa (*gel*, *gel2*) har dezakete eta honek, genitibo edutezkoarekin (*gen*) batera, elipsia (*eli_ize*) har dezake atzetik, berriro ere mugatasuna/numeroa martxan jarriaz (baina kasu honetan mugatuak besterik ez ditu kontuan hartuko, izen eliditua izateak nahitaez ezaguna eta, ondorioz, mugatua izatea eskatzen baitu).

Hau da, modu zabalean, goiko grafoaren irakurketa eta, beraz, izen arrunt bizigabeen morfotaktikaren azalpena. Baina beste horrenbeste beharko dugu izen arrunt bizidunekin (beste jarraitze-klase bat), etab. Esaterako, oraindik egokitu gabe dauden zenbait maileguk marratxoaren beharra izaten dute irakurgarritasunaren mesedetan. Kasuotarako *I_MAR* (*stock-ean*) eta *I_biz_MAR* (*surfer-ak*) jarraitze-klaseak daude, baina mugatuta erabiliko dira, ez edozein izen arruntetarako. Eta mailegu hauek, noski, elkarketa bideratzeko aukera ere izango dute (*elk_marra*) (*stock-arazo*, *surf-ikastaro*).

Leku-genitiboak⁴⁹, erlatiboak eta zenbatzaile ordinaletan elipsia eragiten dute, bizidunek eta bizigabeek har dezaketena gainera: *etxe 0 ko 0 a*, *du n 0 a*, *bi garren 0 a*. Hemen ez dugu bizidun/bizigabe banaketarik egin, elipsiaren ondoren biak etor daitezke-eta.

Atzizki apreziatiboei deklinabide-kasuak eta elkarketa-marra erants dakieke, baina ez beraien mailako atzizkirik (*haurtxo*, *etxetxo*, baina **haurtxotxo*, **haurtxoño*, **etxetxotzar*).

5.2.3.3 Adberbioa

Bukatzeko, adberbioen morfotaktikaren adibide bat ikusiko dugu, hauek, esan dugunez, ia kasuan-kasuan aztertu behar dira-eta. *Adberbioak* azpilexikoan *berehala* aditzondoa “berehala[ADB][ADO] KO_I0” gisa definitu dugu; alegia, morfotaktikan *KO_I0* jarraitze-klasea dugu:

LEXICON KO_I0	
gel2 ;	! leku-genitiboa
I0 ;	! bukaera

non

```
LEXICON gel2
+ko [DEK] [GEL] I1_ELIP ;
LEXICON IO
# ;
```

alegia, *berehala* eta *berehalako* izango ditugu, aditzondo hutsa eta izenlagun bilakatua. Azken honek, baina, I1_ELIP jarraitze-klasea har dezake berriro ere eta, ondorioz, izen gisa funtzionatzen has daiteke (*berehalakoa izan da erantzuna* modukoetan), azpian izen eliditua dagoela adierazten baitu jarraitze-klase horrek.

5.3 Erregelak

Aurreko atalean gramatika osatu dugu, baina horri azaleko forma eman behar zaio eta horretarako erregelen beharra izango dugu.

Koskenniemiren eredua ikustean aipatu dugu bi mailatako erregelek formalismoan duten lekua, bai eta linguistika konputazionalari ekarri dioten aurrerakuntza ere. Baina ondoren izan dituzten kritikak ezin ahantz ditzakegu. Guk bi mailatako erregela hobetuak aurkeztuko ditugu ondoren, lehenengo bertsioak baino askoz ere irakurgarriagoak direlako. Izan ere, bertsio zahar horretan, PC-KIMMO estiloko diakritiko asko erabiltzen da. Gu nagusiki Karttunen eta Beesleyk (1992, 1994; Karttunen 1993) proposatutako eredu berriari jarraituko gatzazkio, testinguruan karaktere-bikote hutsak gainditu eta ezaugarri morfologikoak erabiliz. Autore hauek proposatutako *twolc* eta *lexc* sintaxiak erabili ditugu, aurreko kapituluaren aurkeztu ditugunak. Hori izango da Koskenniemiren ereduarekiko desberdintasun nagusia.

Pragmatikari erantzuten dioten erregela-multzoak aurkeztuko ditugu, alegia, Koskenniemik (1983) erabilitako irizpide berari jarraituta: eransketa-, ezabaketa- eta trukaketa-erregelak. Gogora dezagun karaktere mailako aldaketa dela erregelak multzokatzeko irizpidea, baina baldintza eta testinguru desberdinek egiten dute erregela mota desberdinekoa izatea: karaktere-parea erregela osatzeko helduleku besterik ez da. Adibidez, *n*-ren galera hainbat testingurutan gertatzen da, baina arrazoi oso desberdinei erantzunez; ez da berdin esplikatzen *egite*, *zegoela* edo *harekiko* adibideetako sudurkariaren galera.

⁴⁹ Jabego-genitiboak ere elipsia hartzen du, baina honen parean motibatiboa ere onartu dugu (*ume a en gatik*), elipsirik gabea. Horregatik ez dugu hemen landuko, jarraitze-klase desberdina hartzen du-eta.

5.3.1 Aurredefinizioak: erregeletako sinboloak

Erregelen egitura ikusi dugu 3. atalean eta horixe erabiliko dugu hemen ere. Baina horien erabilera egokirako aurredefinizio batzuk zehaztu beharko ditugu: alfabetoa, ezaugarri morfologikoak, diakritikoak, multzoak eta definizioak⁵⁰. Ikus ditzagun banaka⁵¹:

Alfabetoa: karaktere arruntek osatzen dute alfabetoa, azaleko mailan zein lexikokoan agertzen direnak.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	ñ	o	p	q	r	s	t	u	ü	v	w	x	y	z
9	-	_	.	+																							

Euskal alfabetoa⁵² esatea nahikoa genuke printzipioz; baina, horrez gain, marra —elkarketakoa edo marratxoa—, ordinalen puntua (.), maiuskulak adierazteko aurretik behar dugun 9 zenbakia eta forma konplexuak detektatzeko erabiliko dugun azpimarra () ere hemen bildu ditugu, hauek azalean ere aurki ditzakegulako. Alfabetoaz haratago doa lehen multzo hau, beraz.

Ezaugarri morfologikoak: maila lexikoan bakarrik ageri dira, alegia, ADB:0 moduan ulertu behar dira.

ADB	ADI	ADL	ADT	DET	IOR	IZB	LIB	MEN	ORD					
B1	B2	B3	B4	B5	B6	B7	B8	MOT	MG	MP	MPH	MS		

Erregelen testuinguruetan erabili direlako zehaztu dira hemen, baina ez dira guztiak, baldintza bezala behar direnak bakarrik. Morfologikoki baldintzatutako erregelak daudela erakusten dute ezaugarriok. Mota desberdinetako ezaugarri morfologikoak daude hemen: kategoria nagusiak (ADBerbioa, ADItza, DETerminatzailea eta IzenORDaina), azpikategoriak (IZB erakunde-izena, LIB leku-izen berezia, MENderagailua eta ORDinala), kategoria lagungarriak (ADL aditz laguntzailea eta ADT aditz trinkoa) eta kategorien barruko informazio morfologikoa (iragana: B1-B8, MOTibatiboa, MG mugagabea, MP mugatu plurala, etab.)

Diakritikoak: azaleko mailan gauzatzen ez diren sinboloak dira diakritikoak (horregatik ez dago 0 gauzatzea aipatu beharrik, berez diakritiko bezala ezagutzen dira-eta; baina erregelako testuinguruan beti lexiko-mailan daudela zehaztu beharko da, bi puntuen ezkerrean, *Eep*: esaterako) eta ez dute eraginik erregelen testuingurua egiaztatzean aipatzen ez badira. Berez, marka fonologikoak dira.

Aadi	Aorg	Aart	Asalb	Eep	Ebih	EpBer	Rep	Rgog	Rzah
SigB	SigBK								

⁵⁰ Erregelen testuingurua modu egokian aplikatzeko arazoak daude ezaugarri morfologiko eta diakritiko batzuekin eta, tarteka, alfabetoan txertatu beharko ditugu erregela ondo aplikatzeko.

⁵¹ Espresio erregularrekin ez nahastearren, deskribapena modu sinplifikatuan aurkeztuko dugu, markak kenduta. Horrela, irakurgarriago izango dira guztiak. Hiru sailetan du soilte honek eragina: 1) alfabetoaren sailean, % ihes-karakterea soildu dugu azken karaktereetan: %9, %- , % , %, %+ beharko zuketzen berez; 2) ezaugarri morfologikoez ere, berez %[ADB%] modukoak izan beharko zuketzen, baina soilduta aurkeztuko ditugu, ADB moduko hutsak erabiliz; azkenik, c) hautapen-markak edo diakritikoak, berez, makoen artean ere azaldu beharko lukete, %[Aadi%] moduan, adibidez, baina Aadi bezala azaleratuko ditugu.

⁵² Euskaltzaindiaren 17. arauaren arabera, nahiz ü gehitu den (*Altziurükü* moduko izen bereziak ere euskal izenak direlako, eta hori ez da arauan jasotzen).

Aadi: -a bukaera duten, baina *a:e* aldaketarik egiten ez duten aditz trinkoak: *darama*: *daramala*, baina *da:del*).

Aorg: *a* organikoa edo berezkoa duten lemak.

Aart: *a* bukaera duten leku-izen berezi batzuek *a* organikoa ez, baina artikulua dute berekin. Eta deklinatzerakoan, kasu gramatikaletan mantendu egiten badute ere (*a* organikoak bezala), lekuzko kasuetan galdu egiten dute: *Azpeitia*, *Azpeitiak*, *Azpeitiari*, baina *Azpeitiko*.

Asalb: hitz-elkarketan bukaerako *a* gal dezaketen sei formek hartzen duten marka (-ia bukaera dutenak ez dira hemen sartzen): *burdina*, *eliza*, *hizkuntza*, *kultura*, *literatura* eta *natura*. Hauek *kultura*, baina *kultur etxe* egin dezakete (nahiz *kultura etxe* eta *kultura-etxe* ere ontzat ematen diren (Euskaltzaindia, 25. araua).

Eep: *e* epentetikoa.

Ebih: *a-z* bukatutako aditz defektiboek dute, partizipioari osagai bat eranstea *a:e* egiteko, *ateraEbih+a:aterea*, etab.

EpBer: epentesi berezia, *har-* alomorfoak (*hura*) eta *nor-* (*inor*, *edonor*) nahiz *zer-* (*ezer*, *edozer*) modukoek eskatzen dutena, *hark* baina *harez*, *zerk* baina *zerez* egiteko.

Rep: *r* epentetikoa eska dezake.

Rgog: *r* gogorrez bukatutako lemei dagokien marka, ondoren bokal bat badator bikoiztera behartzen duena: *hurRgog+a*: *hurra* (baina *ur+a:ura*).

Rzah: *hiru* eta *lau* zenbatzaile zehaztu zaharrena da hau, *hiruretan*, *laurek*, etab. egiteko, baina *seietan*, ez **seiretan*.

SigB: bokalez bukatutako izenen moduan deklinatzen diren siglek behar dute diakritiko hau (*SigBK* markak ez bezala).

SigBK: siglak bokalez nahiz kontsonantez bukatu bezala deklinatzeko.

Hauek gain, sistemari berari dagozkion +, # eta ! ere azalduko dira: plus zeinuak (+) morfemalotura markatzen du; sostenituak (#), berriz, hitz-muga, ezkerreko testuinguruan hitz-hasiera eta eskuinekoan bukaera; heia-zeinuak (!), azkenik, ez du erregelan eraginik, oharrak markatzeko-eta erabil daiteke.

Multzoak: ezaugarri komunak dituzten azaleko alfabetoko karaktereak biltzen ditu, betiere karaktere arruntak elkartuz.

```
Bokal = a e i o u ü ;
Bireki = a e o ;
Bitxi = i u ü ;
Konts = b c d f g h j k l m n ñ p q r s t v w x y z ;
Txis = z s x ;
AlboSud = l m n ñ ;
HersAhosk = p t k ;
HersAhost = b d g ;
```

Definizioak: errepikatzen diren definizio erregularrak biltzen dituen multzoa, ez karaktere arrunt hutsak.

```
LekKonts = AlboSud | Txis | t | r (0:r) ;
Afrik = t Txis ;
LekKas = [ r a | t i k | k o ] Eep: | k o ;
Dipt = a i | e i | o i | u i | a u ;
Rez = (%:.) ( %+: %-: ) ;
Emend = \[ u | t ] e | i | \d \i o | \[ g | z ] u ;
MorfBuk = #: | %+: ;
Hasiera = #: (%*:) ;
Mm = [%+:]+ ;
Morfe = [ Bokal: | Konts: | %-: | %*: | %_: | %:~: ]+ ;
Irag = [ B1: | B2: | B3: | B4: | B5: | B6: | B7: | B8: ] ;
```

Emend definizioaren esanahia hau da: menderagailua eranstea *e* epentetikorik har ez dezaketen adizki jokatuak. Ukapena erabili dugu definizio hau osatzeko, hain zuzen ere hemen zehaztu ditugun bukaerek ez baitute epentesirik hartzen: *du:duela*, baina *dugu:dugula*, ez **duguela*.

Ezaugarri morfologikoen erabilera ezinezkoa zen hemen, eta diakritikorik ere ez genuen datu-basean sartu nahi kasuotarako bakarrik. Beraz, definizioan zehaztutako karaktere-segidek ez dute erregela aplikatuko, definizioan bertan ukapena baitaramate (\). Hemen bildutako kasuak hauek dira: *ue*, *te*, *i*, *dio*, *gu* eta *zu* bukaera dutenak, *alegia*, *duzue*, *dute*, *nadi-*, *dio*, *dugu* eta *duzu* motakoak.

Aurreko karaktere eta sinbolo hauen guztien esanahia argituta, posible dugu erregelak irakurri eta ulertzea. Guztira 16 metaerregela sortu ditugu: eransketari dagozkion hiru (*0:r*, *0:e* eta *0:a*), ezabaketari dagozkion zazpi (*r*, *e*, *a*, *k*, *t*, *n* eta *h*), bost trukaketa-erregela (*a:e*, *k:g*, *t:d*, *HersAhost:HersAhosk* eta *e:i*) eta, azkenik, ortografia hutsa den marratxoaren galerakoa (*-:0*). Horiekin guztiekin testu-hitz baten analisisian edo/eta sorkuntzan gertatzen diren aldaketa guztiak jasotzen dira.

5.3.2 Erregelak

a) **Eransketa:** bi osagai elkartzean karaktere berri bat sortzen da.

1. *r* epentetikoa

```
0:r <=> Bokal (LIB:) (Rez) Mm _ Morfe Rep: ;
      r Rgog: _ Mm (h:0) :Bokal ;
      r LIB: Rgog: _ Mm :Bokal ;
      Rzah: _ Mm :Bokal ;
0:r <= IZB: SigB: (Rez) Mm _ Morfe Rep: ;
0:r => IZB: [ SigB: | SigBK: ] (Rez) Mm _ Morfe Rep: ;

      ! seme++iRep:semeri (baina !seme+eMP+iRep:semeei;
      !          zuhaitz++iRep:zuhaitzi)
      ! amaAorg+a+ekinRep:amarekin
      ! 9bilboLIB+iRep:9bilbori
      ! zaharRgog+ikRep:zaharrik (baina !plater+a:platera)
      ! agerRgog+araz:agerraraz
      ! berRgog+has+i:berrasi
      ! 9eibarLIBRgog+iRep:9eibarri
      ! hiruRzah+ak:hirurak
      ! 9e9d9b9lIZBSigB+enRep:9e9d9b9lren
      ! 9m9i9tIZBSigBK+enRep:9m9i9tren (9m9i9tIZBSigBK+enRep:9m9i9ten)
```

Bi osagaien lotura bideratzen duen karakterea da epentesia, *r* kasu honetan. Baina lau modutakoa izan daiteke: a) bi bokalen artekoa, bokalez bukatutako lemari bokalez hasitako atzizkia gehitzean behar dena (*etxeRi*); b) *r*-z bukatuak bokalez hasitako beste morfema bat gehitzean eskatzen duena, *r* gogorra (*zakurRa*); c) *hiru* eta *lau* zenbatzaile zaharrek bokala eranstea eskatzen dutena (*lauk*, baina *lauRak*); eta d) siglen tratamendu berezia.

2. *e* epentetikoa

```
0:e <=> [ Knts (%.:) | t:0 | 0:r ] Mm _ Morfe Eep: ;
      [ [ Emend [ ADL: | ADT: ] ] | [ d i o ADT: ] ] Mm _ Morfe MEN: ;
      [ r t | b a i t ] ADB: Mm _ HersAhosk | r ;
      EpBer: Mm _ z (Morfe) Eep: ;
      #: g a u Mm _ [ MS: [ a | LekKas ] ] Eep: ;
      #: (Knts)* Dipt Mm _ MG: Morfe INS: Eep: ;
0:e => Knts [ IOR: | DET: (EpBer:) | Mm MG: ] Mm _ Morfe MOT: Eep: ;
      LekKnts LIB: (Rez) Mm _ Morfe (MOT:) Eep: ;
      ORD: Mm _ HersAhosk (Morfe) Eep: ;
      SigBK: (Rez) Mm _ Morfe Eep: ;
0:e <= Knts [ IOR: | DET: ] Mm _ Morfe MOT: Eep: ;
      LekKnts LIB: (Rez) Mm _ Knts (:0)* Eep: ;
```

```

! haran++koEep:haraneko
! polit++koEep:politeko (! seme++koEep:semeko)
! zaharRgog++raEep:zaharrera
! zenADLB1+laMENEep:zela
! duADL+laMENEep:duela (baina ! duguADL+laMENEep:dugula)
! naizADL+nMENEep:naizen
! dutADL+laMENEep:dudala
! dukADL+laMENEep:duala
! daADT+laMENEep:dela (baina ! daramaADTAadi+laMENEep:daramala)
! dioADT+laMENEep:diola (baina ! dioADL+laMENEep:diola)
! bartADB+ko:barteko (baina ! bartADB+danik:bartdanik)
! nolabaitADB+ko:nolabaiteko
! zerEpBer+zEep:zerez (baina ! zerEpBer+kEep:zerk (ez *zerek))
! gau+MS+aEep+nEep:gauean (baina ! gau+MG+kEep:gauk)
! dei+MG+zEep:deiez
! norbaitIOR+gatikMOTEep:norbaitegatik
! 9usurbilLIB+koEep:9usurbileko (nahiz ! 9usurbilLIB+koEep:9usurbilgo)
! bi+garrenORD+koEep:bigarrenko (nahiz ! bi+garrenORD+koEep:bigarreneko)

```

Aurreko erregelari bezala, zehaztuko ditugun morfema-pilaketa batzuetan *e* bokala sortzera behartuta gaude. Erregularik —metaerregularik— konplexuena da, testuinguru askotan gertatzen delako batetik eta, bestetik, beste hainbat erregelarekin erlazionatua dagoelako —eta, noski, horren berri testuinguruan eman beharra dago—.

3. Adizki jokatueta *a* azalekoa

```

0:a <=> [ t:d | k:0 | n ] [ ADL: | ADT: ] Mm _ ;

! dutADL+nMENEep:dudan
! dunADT+laMENEep:dunala (baina ! zuenADLB1+laMENEep:zuela)
! dukADT+laMENEep:duala

```

0:a hau *e* epentetikoaren pareko dela ere esan genezake, baina *Eep* beharrean *a* egiteko orduan aurreko kontsonanteek dute eragina. Bestela, **duten*, **duken* eta **dunen* egingo genuke, paradigmari jarraituz. Adizkiak baditu murriztapen batzuk: *dut* modukoek azken karakterea ahostun bihurtuta azaleratzen dute (*t:d*, 13. erregela), *duk* modukoek *k* galduta (*k:0*; 7. erregela) eta *dun* modukoek azken karakterea bere hartan mantentzen dute.

b) Ezabaketa: azaleko karaktere bat desagertzen da erregela aplikatzean.

4. *r*-ren galera

```

r:0 <=> [Konts-r] LIB: Mm _ :a ;

! 9usurbilLIB+raEep:9usurbila
! 9eibarLIBRgog+raEep:9eibarra
! 9usurbilLIB+rantzEep:9usurbilantz

```

Erregela bakarra da lexiko mailako *r* dardarkaria azalera iristen uzten ez duena: lekuzko izen bereziki gertatzen dena.

5. *e*-ren galera

```

e:0 <=> [ MP: | MPH: ] Mm _ ;

! seme+eMP+ekinRep:semeekin
! seme+eMP+kEep:semeek
! amaAorg+eMP+entzatRep:amentzat
! zue [IOR] [PER] MP+iRep:zuei
! haieDETERKMP+ekinRep:haiekin

```

Lexiko mailako *e* azalera pasa aurretik "ezabatu" egiten da, alegia, bakarrera bildu, ezkerreko testuinguruan mugatu plurala badu morfema-mugaren aurretik. Erregela hau aplikatu izan ez balitz, *seme+e+ekin:semeekin* izango genuke. Are gehiago, eskuineko testuinguruan diakritikoa ezarri

izan bagenu (*Rep*) **semeerekin* aterako litzaiguke. Hori konpontzen du erregela honek, *Rep* lexikoan definitua dugulako eta, beraz, testuinguruan aipatzen ez denez, ez da betetzen.

Bestalde, zenbait sarrerak mugatasun lexikala dute: *zuek* izenordain pertsonala eta determinatzaile erakusle pluralak (*hauek, horiek, haiek, beroriek,...*). Hauen alomorfoak *zue-*, *haue-* etab. dira, *e* eta guzti, hor ikus baitaiteke pluraltasuna eta hori baita aldagaitza den unitate minimoa: *zue-i* ...

6. *a*-ren galera (bokal-bilkura)

```
a:0 <=> _ [Aorg: |Asalb:] Mm :Bireki (Morfe) (MP:|MPH:) MorfBuk ;
_ LIB: Aart: Mm LekKas ;
_ a Mm _ r: a: z ;
a:0 => i _ Aorg: #: ;
_ Asalb: #: ;
r a Mm a:0 r:0 _ z ;

! amaAorg+a:ama
! amaAorg+eMP+entzatRep:amentzat
! kulturaAsalb+a:kultura
! 9azpeitia[IZE] [LIB]Aart+koEep:9azpeitiko (baina
! 9azpeitia[IZE] [LIB]Aart+iRep:9azpeitiari)
! pentsa+araz+i:pentsarazi
! filologiaAorg:filologi (nahiz ! filologiaAorg:filologia)
! kulturaAorgAsalb:kultur (nahiz ! kulturaAorgAsalb:kultura)
! kulturaAorgAsalb+a:kultura
```

Kasuistika arras desberdinak biltzen ditu metaerregela honek, baina guztietan maila lexikoko *a* bokala ondokoari bilduko zaio testu-hitza azalean gauzatu aurretik.

7. *k*-ren galera

```
k:0 <=> _ Mm HersAhosk ;
_ [ ADL: | ADT: ] Mm ( 0:a | 0:e ) Morfe MEN: ;

! etxe++tikEep+ko:etxetik
! dukADL+laMENEep:duala
```

Herskari ahoskabeek eta adizki jokatu batzuek eragiten dute maila lexikoko *k*-ren galera, eta salbuespenik gabea da dagoen testuinguruetan (<=>).

8. *t*-ren galera

```
t:0 <=> _ Mm [ :HersAhosk | AlboSud | h ] ;
_ Txis LIB: Mm HersAhosk (Morfe) Eep: ;
_ Txis ADI: Mm t ;
n _ Txis Mm k ;
Txis _ Mm Morfe [ ORD: | DIS: ] ;

! bait+gara:baikara
! bait+naiz:bainaiz
! bait+haiz:baihaiz
! ezagut+tu:ezagutu (baina ! polit++tikEep:politetik)
! 9zarautzLIB+tik:9zarautzik
! utzADI+te:uzte
! etxe++rantzEep+ko:etxeranzko
! bost+garrenORD:bosgarren (baina ! bat+garrenORD:batgarren)
! bost+naDIS:bosna
! zenbat+naDIS:zenbana (baina ! zenbat+garrenORD:zenbatgarren)
```

Ezkerreko testuinguruko *t* herskari ahoskabea ezabatu egiten da: a) herskari ahoskabea, albokaria, sudurkaria edo *h* badu atzetik; b) afrikatuz bukatutako leku-izen bereziei herskari ahoskabeek hasitako atzizkia eranstea; c) bai eta bukaera bereko aditzekin ere, ondoren *t* hartzean; d) hurbiltze-adlatiboari *-ko* atzizkia eranstea; eta e) *bost* zenbatzaile zehaztuari ordinala gehitzean.

9. *n*-ren galera

```

n:0 <=> _ ADI: Mm [ t e | k i MorfBuk ] ;
      k i _ (Rep:) Mm k: o ;
      _ [ ADL: | ADT: ] Irag Mm (0:a) Morfe MEN: ;
      _ (ADB:) Mm n ;
n:0 => Bokal _ t:0 Txis ADI: Mm t ;

! eginADI+te:egite
! izanADI+ki:izaki
! amaAorg+a+ekinRep+ko:amarekiko
! zenADLB1+laMENEep:zela
! zuenADLB1+nMENEep:zuen
! ehun+na:ehuna
! nonADB+nahi:nonahi
! jantzADI+te:jazte
! jantzADI+te:janzte

```

n sudurkaria hainbat kasutan ez da azaleratzen: a) aditz nagusiari herskariz hasitako atzizki zehatz batzuk gehitzean; b) atzizki artean; c) iraganeko adizki jokatuaren menderagailua ezartzean; eta d) *n+n:n* pilaketan. Hauek guztiek beti gauzatu behar dute aldaketa. Baina testuinguruaren murritzapena bakarrik dakarren beste erregela batean, aurreko erregelari ikusi dugun *jantzi* moduko aditzetako *n:0* konponduko da.

10. *h*-ren galera

```

h:0 <=> #: b e r (0:r) Rgog: Mm _ ;

! berRgog+has+i:berrasi

```

15. erregelari ikusiko denez, *h* ez da azaleratzen aurretik *ber-* aurrizkia badu (# hitz-hasiera da), eta beste erregela batean (1. erregela) sortu dugun *r* epentetikoa, *r* gogorraren aurrean baikaude (bestela **berasi* modukoak ontzat emango genituzke), morfema-mugaren aurretik.

c) **Trukaketa:** karaktere baten aldaketa beste bati elkartzean.

11. *a:e* aldaketa

```

a:e <=> _ Ebih: Mm [ Bireki | LekKas ] ;
      _ [ ADL: | ADT: ] Mm Morfe MEN: ;

! ateraEbih+a:aterea
! ateraEbih++tikEep:ateretik
! daADL+nMENEep:den
! garaADT+laMENEep:garela (baina ! daramaADTAadi+laMENEep:daramala)

```

Atera moduko aditz defektibo batzuek partizipioari mugatasuna eranstean *a+a:e* egiten dute, bai eta *a-z* bukatutako zenbait adizki jakatutak ere, menderagailua eranstean. Jokatuaren kasuan, ezabaketa-erregela ikusten duenik ere bada, adizkiaren *a* bukaerakoa galdu eta epentesia hartzen duela dioenik, alegia. Baina ez dirudi hala gertatzen denik, indar handiagoa baitu adizkiak berak epentesi hutsak baino.

12. *k:g* aldaketa

```

k:g <=> [ AlboSud: LIB: | n (ORD: | ADB: ) ] Mm _ o ;

! 9usurbilLIB+ko:9usurbilgo
! 9irunLIB+ko:9irungo
! egin+ko:egingo53

```

⁵³ Lehenago aipatu dugunez, *n-z* bukatu aditzetan herskari ahoskabea ahostun bilakatzen da flexio-morfema gehitzean, baina eratorpenean ez da beti gertatzen: *emankor*, ez **emakor*, *eginkizun*, ez **egikizun* nahiz *-te* erator-atzizkia gehitzean aditz-izenak bezala jokatu duen: *egite*, ez **eginte*.

```
! hemenADB+ko:hemenko
! nonADB+ko:nongo (baina ! gaurADB+ko:gaurko)
! bi+garrenORD+koEep:bigarrenko (nahi ! bi+garrenORD+koEep:bigarreneko)
```

Lexiko mailako *k* herskari ahoskabeak ahostun bezala azaleratu behar du lekuzko izen berezi batzuekin, eta *n-z* bukatutako ordinaletan eta adberbioekin, *Mm*-ren ondoren *o* bokal irekia badago. Baina erregelan zehaztu beharrik ez dugun beste kasu batzuetan ere gertatzen da hori, aditzekin esaterako: *egin+ko:egingo*.

13. *t:d* aldaketa

```
t:d <=> AlboSud: LIB: Mm _ ;
_ [ ADL: | ADT: ] Mm (0:a | 0:e) Morfe MEN: ;
:AlboSud ( ADB: ) Mm _ [ =:a | Bitxi ] ;

! 9irunLIB+tikEep:9irundik
! 9usurbillLIB+tikEep:9usurbildik
! dutADL+laMENEep:dudala
! dakitADT+laMENEep:dakidala
! egunADB+tikEep:egundik (baina ! egun++tikEep:egunetik)
! egin+ta:eginda54
! agin+tu:agindu
```

t herskari ahoskabeak *d* ahostuna azaleratzen du beti hiru kasu betetzen badira eta, gainera, kasuok beti eman behar dute ahostuna: a) albokari edo sudurkariz bukatutako leku-izen bereziak; b) menderagailua hartzen duten *t-z* bukatutako adizki jokatuak (tartean beste aldaketa batzuk ere badira, 3. erregelan aipatu dugun moduan, oinarrian *da* egon baitaiteke Mitxelenaren (1961) ustez); eta c) albokariak eta sudurkariak oro har, adberbioa aukeran, baina aditza ere izan daiteke, beti testuinguru zehatz batzuetan.

14. Herskari Ahostuna/Ahoskabea aldaketa

```
Cx:Cy <=> #: b a i t: Mm _ ;
where Cx in HersAhost
      Cy in HersAhosk
      matched ;
! bait+dugu:baitugu
! bait+gara:baikara
```

Bait- aurrizki kausala adizki jokatuari itsastean, azken hauen hasierako herskari ahostuna ahoskabe bihurtzen da. Bi norabidetan gertatzen da erregela hau.

15. *e:i* aldaketa

```
e:i <=> #: b _ r Rgog: Mm [Konts-h] ;

! berRgog+sorADIRgog+tu:birsortu
! berRgog+pitzADI+tu:birpiztu
! berRgog+hasADI+i:berrasi
! berRgog+eginADI:berregin
```

Aurritzia kontsonantez hasitako aditz batekin elkartzean, *bir-* azaleratuko da, salbuespen bakarrarekin⁵⁵: kontsonante hori *h* bada, orduan ez du kontuan hartu behar ([Konts-h]).

d) Erregela ortografiko hutsa

⁵⁴ Gogora dezagun berriz ere, eratorpenean ez dela ahostun bilakatze hau sistematikoa: *eginda*, *jakinda*, baina *jakintza*.

⁵⁵ Bada beste salbuespen bat: *berpiztu*; baina aditz hau bere hartan egongo da hiztegitratua. *Piztu* (sua, etab.) aditzak beste erabileretan *e:i* egingo du.

16. Marratxoaren desagertzea

```
%-:0 <=> _ (%+:) #: ;
! 9ararat+-:9ararat
! 9ararat+-+:9ararat
```

Marratxoa ez da azaleko mailan azalduko hitz-bukaeran (#) dagoenean.

5.3.3 Erregela taktikoak

Aurreko kapituluetan aipatu dugunez, urruneko mendekotasuna ez da bi mailatako formalismoak proposatzen duen morfotaktikarekin —jarraitze-klaseekin— konpontzen. Beraz, erregelak baliatu behar ditugu horretarako, sistema bera indartuz gainera: debeku-ezarpenak edo ukatze-erregelak.

Bi erregela taktiko sortu ditugu, hain zuzen ere adizki jokatuena aurrizki vs. atzizki erabilera zaintzeko eta elkarketa-marra behin bakarrik onartzeko. Hartara, **baitelako* eta **liburu-azoka-denda* modukoak ukatuko ditugu, erabilera zuzena —edo arruntena— bakarrik onartuz eta gainsorrera ebatziz elkarketaren kasuan.

Erregela hauek aplikatzeko aurredefinizioak zertxobait aldatzen dira: alfabetoa berdin mantentzen da; hiru ezaugarri morfologiko behar dira: BALD (baldintzazkoa), KAUS (kausazkoa) eta ELM (elkarketa-marra). Eta bi definizio zehaztuko dira:

```
Karak2 = ~[ %+ | %-] ;
Karak = ~[%+] ;
```

Horrekin, erregela taktikoak hauek dira:

1. ba-/bait- galera

```
b /<= #: _ Karak* [ BALD | KAUS ] Karak* %+ Karak* %+ ;
! baBALD+naizADL+laMEN:naizela
! naizADL+nMEN:naizen
```

Adizki jokatuetakoa aurrizkiak —*ba-* baldintzazkoa eta *bait-* kausazkoa— ezin dira azaldu (/<=, ukapen-erregela) *BALD* eta *KAUS* ezaugarri morfologikoak azaltzen direnean eta atzizkien morfema-mugaren ondoren (%) hainbat karaktere —adizkiak berak— eta berriro morfema-muga badago —horrek esan nahi baitu beste zerbaitek datorrela adizkiaren atzetik, alegia, atzizki menderagailuak—. Urruneko mendekotasuna kontrolatzeko erabiltzen da erregela hau, adizkiek aurrizkia eta atzizkia har baitezakete, baina bata erabiltzeak bestea ukatzen du.

2. Elkarketa-marraren kontrola

Hitz-elkarketara egin dugun hurbilpen sinpleenetakoa *izen + izen* motakoa izan da, bai bukaerako -a organikoaren galera kontrolatuz, bai elkarketa-marra onartuz. Bada, azken honen berri emateko erregela taktikoen beharra dugu; izan ere, izen arrunten jarraitze-klasean '-' onartzen dugu, baina bigarren izenari ukatu egin behar zaio berriro erabiltzea, bestela nahi beste izen izango genuke elkarren segidan, **haur-liburu-etxe*, esaterako. Gogora dezagun, halere, oraingoz edozein izen dela beste batekin elkar daitekeena, ez dugula murriztapenik ezarri (eta hau, neurri batean, gainsortzailea da).

```
%- /<= %- %+ Karak2* %+ _ ELM ;
```

! haur +-ELM+liburu+-ELM+denda:haur-liburu

Elkarketa-marra ez da onartuko aurretik beste marra bat (%) eta hainbat karaktere (*Karak2**) — izenari dagozkionak, nahiz ez dugun zehaztu beharrik, morfotaktikan baitago— badaude eta ondoren ELM ezaugarria ageri bada.

Hamasei erregela hauekin analisia eta sorkuntza egiteko gai izango da prozesadorea, hau da, azaleko mailako testu-hitzak analizatuko ditu, bai eta hiztegi-sarrerak flexionatu eta testu-hitz bilakatuko.

5.4 Emaizak

Formalismoaren euskararako aplikazioa burututa, hitzaren analisia eta sorkuntza egiteko gai da prozesadorea. Analizatzailearen estaldura-tasa orokorra % 95 ingurukoa da, oinarri gisa azken urteak biltzen dituzten corpusak kontuan hartuta. Hutsen arrazoi nagusiak testu-motaren araberakoak badira ere, euskara ez-estandarren erabilera, terminologia berezituégia, osaera berriko hitzak, maileguak eta idazkera-akatsak dira. Ondoko kapituluan hobekuntza-proposamenak egingo ditugu, batez ere aldaeren eta akatsen detekzioa bideratu eta lexikorik gabeko analisiak ahalbidetuko dituztenak. Horrela, estaldura % 100 izatea lortuko dugu. Zehaztasunari edo doitasunari dagokionez, % 99 gaingitzen dugula esan dezakegu, alegia, % 100etik oso gertu gaudela.

Ikus ditzagun hori frogatuko duten adibide batzuk:

5.4.1 Sorkuntza morfologikoa

Sorkuntza morfologikoa, teorian *ad infinitum* irits daiteke euskaraz, praktikan mugak ezartzen badira ere. Aurreko ataletan ikusi dugunez, hiztegi-sarrerari⁵⁶ mugatasuna/numeroa ezartzen zaio eta, ondoren, aldian aldiko kasua. Baina, esan dugunez, genitiboek (jabego- eta leku-genitiboak) kasuak har ditzakete atzetik, birdeklinabidea onartuz, eta berriro mugatasuna eta kasua hartzeko aukera ere ematen dute, kasurik gehienetan elipsia baitago analisiaren azpian. Beraz, teoriar mugarik gabe jarrai badezakegu ere, praktikan birritan baino gehiago aplikatzea ez da normala. Eta, hala ere, kasua behin bakarrik aplikatuta, izen arruntek 77⁵⁷ forma osa ditzakete. Biak

⁵⁶ Hiztegi-sarrera forma soila izan daiteke, baina aurizkia edo/eta atzizkia ere erants dakióke mugatasuna/numeroa aplikatu aurretik.

⁵⁷ Izen arrunt bizigabeek, esaterako: 15 deklinabide-kasu ikusi ditugu, baina 22 aldaera guztira. Hauetako 17 aldaerek mugagabea, mugatu singularra, plurala eta plural hurbila hartzen dituzte (17 x 4 = 68). Partitiboak eta destinatiboko banatzaileak eta prolatiboak (azken honetako bi aldaerek) mugagabea bakarrik har dezakete (4

aplikatuta, aldiz, 275⁵⁸ hitz sor ditzakegu sarrera bakarretik. Eta bi aplikaziotan jarri dugu muga. Are gehiago, ez ditugu atzizki lexikalak kontuan izan.

Sorkuntzan mugak ezartzea arriskutsua izan daiteke, forma onargarriak ez onartzea gerta baitaiteke, baina, inongo murriztapenik ezartzen ez bada, gainsorkuntzak arazo larriak sor ditzake, inoiz ez baita bukatzen. Erregela taktikoak balia daitezke hori kontrolatzeko, elkarketarekin egiten den bezala, genitiboak bi aldiz aplikatu eta hortik aurrera sorkuntza debekatzuz.

Ikus dezagun ondoko adibidean, lema edo hiztegi-sarrera bat emanda, morfotaktika ondo definitua badago, eskura dezakeguna zenbat zabalitzen den:

```

gener(ate) > 1. etxe
2. etxeak
3. etxeagatik
4. etxeagatiko
5. etxeagatikoak
6. etxeagatikoaganaino
7. etxeagatikoaganako
8. etxeagatikoaganantz
9. etxeagatikoaganantzago
10. etxeagatikoaganantzegi
11. etxeagatikoagana
12. etxeagatikoagandik
13. etxeagatikoagan
14. etxeagatikoagatik
15. etxeagatikoak
16. etxeagatikoaz
17. etxeagatikoa
18. etxeagatikoarekin
19. etxeagatikoarendako
20. etxeagatikoarentzat
21. etxeagatikoari
22. etxeagatikoei
23. etxeagatikoek
-- <CR> for more, q to quit --990. etxeotakootarako
991. etxeotakootarantz
992. etxeotakootarantzago
993. etxeotakootarantzegi
994. etxeotakootara
995. etxeotakootatik
996. etxeotakootarik
997. etxeotakooz
998. etxeotakookin
999. etxeotakoondako
1000. etxeotakoontzat
1001. etxeotakoan

```

sarrera, beraz); ablatiboaren aldaerak mugagabea eta plurala (hurbila barne) har ditzake (3 sarrera) eta, azkenik, motibatiboaren pluraleko aldaerak bi dira (hurbila ere behar dugulako). Hortaz, $68 + 4 + 3 + 2 = 77$ sarrera oinarritzko izango ditugu.

⁵⁸ Aurreko 77 sarreretatik abiatuta, zazpi kasuk, lau muga(gabe)tasun-zutabeak kontuan hartuta, *-ko* har dezakete, eta 5 aldaerek ere bai ($7 \times 4 = 28$; $28 + 5 = 33$ forma berri). Hauek, mugatu singularrean 17 forma berri osa dezakete, mugatu pluralean 19 eta mugatu plural hurbilean beste horrenbeste ($17 + 19 + 19 = 55$ forma gehiago). Aurreko 33ei lotuta, beraz, 88 forma ditugu. Eta jabego- eta leku-genitiboak mugatasuna (3 zutabe, singularra, plurala eta plural hurbila) har dezakete ($2 \times 55 ((17 \times 3) + 2 + 2) = 110$). Hortaz, bigarren maila horretan $77 + 88 + 110 = 275$ sarrera izango genituzke. Eta *mendikoan* modukoak aztertuko genituzke, testu-hitz arruntak, nolanahi ere.

```

1002. etxeotakoko
1003. etxeotakoraino
1004. etxeotakorako
1005. etxeotakorantz
1006. etxeotakorantzago
1007. etxeotakorantzegi
1008. etxeotakora
1009. etxeotakotik
1010. etxeotakorik
1011. etxeotan
1012. etxeotaraino
-- <CR> for more, q to quit --2002. etxetxoetarainokootara
2003. etxetxoetarainokootatik
2004. etxetxoetarainokootarik
2005. etxetxoetarainokooz
2006. etxetxoetarainokookin
2007. etxetxoetarainokoondako
2008. etxetxoetarainokoontzat
2009. etxetxoetarainokoan
2010. etxetxoetarainokoko
2011. etxetxoetarainokoraino
2012. etxetxoetarainokorako
2013. etxetxoetarainokorantz
2014. etxetxoetarainokorantzago
2015. etxetxoetarainokorantzegi
2016. etxetxoetarainokora
2017. etxetxoetarainokotik
2018. etxetxoetarainokorik
2019. etxetxoetarako
2020. etxetxoetarantz
2021. etxetxoetarantzago
2022. etxetxoetarantzagoko
-- <CR> for more, q to quit --q

```

5.4.2 Analisi morfologikoa

Testu-hitz batek analisi bat edo gehiago izan ditzake, informazio morfologikoaren eta honen egituratzearen arabera. Analizatzailerak aukera posible guztiak eman behar ditu, osagai bakoitzak bere informazioa duela. Gero, ondoko aplikazioen arabera, desanbiguate-lanak egin beharko dira, baina beste urrats batzuk dira horiek. Puntu honetan, testu-hitz bat idatzi eta analisiaren emaitza guztiak eskainiko dizkigu prozesadoreak. Hori frogatzeko hainbat adibide ekarri ditugu, argigarri izan daitezkeelakoan. Analisisirik ematen ez duenean, alegia, forma hori txartzat ematen duenean, izartxoak azaleratuko ditu.

```

bada
1. ba [ERL] [MEN] [BALD] +da [ADL]
2. ba [ERL] [MEN] [BALD] +da [ADT]
3. ba [PRT] +da [ADL]
4. ba [PRT] +da [ADT]

badelako
1. ba [PRT] +da [ADL] +lako [ERL] [MEN] [KAUS]
2. ba [PRT] +da [ADL] +la [ERL] [MEN] [KNPL] +ko [GEL]
3. ba [PRT] +da [ADT] +lako [ERL] [MEN] [KAUS]
4. ba [PRT] +da [ADT] +la [ERL] [MEN] [KNPL] +ko [GEL]

baitelako
*****

```

Lehenengo adibidean *ba-* baldintza eta partikula onartzen ditu, bai eta aditz laguntzailea zein trinkoa ere. Bigarrenean, aldiz, menderagailua gehitu diogunez, partikula besterik ez du kontuan hartzen. Eta gauza bera dugu hirugarrenean ere: alegia, erregela taktikoa aplikatu du, urruneko mendekotasuna kontrolatuz eta, beraz, analisia txartzat emanez.

```

ume-liburu
1. ume [IZE] [ARR] +- [ELM] +liburu [IZE] [ARR]
2. ume [IZE] [ARR] +- [ELM] +liburu [IZE] [ARR] + [MG] + [ABS]

ume-liburu-seme
*****

```

Marratxoak bi izen elkartzea onartzen du, baina ez gehiago.

```

AEK-k
1. AEK [SIG] [IZE] [IZB] +- [MAR] +k [ERG]

AEKek
*****

```

Sigletan, *k*-z bukatuek tratamendu berezia dutela ikusi dugu erregeletan. Ezin dira kontsonante arruntzat hartu, marratxoa eskatzen dute.

```

Igori
*****

Igorri
1. Igor [IZE] [PIB] +i [DAT]

igorri
1. igor [ADI] [ARR] +i [ADM] [PART]
2. igor [ADI] [ARR] +i [ADM] [PART] + [ASP] [BURU]
3. igor [ADI] [ARR] +i [ADM] [PART] + [MG] + [ABS]

```

Igor-ek bi analisi ditu: pertsona-izen berezia (*Igor*[IZE]/[PIB]) eta aditza (*igor*[ADI]/[ARR]). Lehenengo adibidean izena eskatu dugu, letra larriz idatziz, baina ez du *r* gogorraren erregela bete; bigarrenean, aldiz, bai. Hirugarren adibidean ez dugu maiuskularen markarik ezarri eta, beraz, aditza besterik ez du kontuan hartu, hiru analisi posibleak emanez.

```

Azpeitiako
*****

Azpeitiari
1. Azpeitia [IZE] [LIB] +i [DAT]

```

Azpeitiak a artikulua du, leku-denborazko kasuetan galdu egiten duena, baina kasu gramatikaletan mantendu. Horregatik ez du lehenengo testu-hitzarentzat analisirik eman, *Azpeitiko* behar zuen-eta.

```

Jonengatik
1. Jon [IZE] [PIB] +en [GEN] +gatik [MOT]

Jongatik
1. Jon [IZE] [PIB] +gatik [MOT]

```

Pertsona-izen bereziek motibatiboa genitibotik pasata nahiz pasa gabe osa dezakete. Bi analisiak, beraz, egokiak dira.

```

Zarautzko
*****

```

Zarautzek bi analisi posible ditu leku-genitiboa aplikatzean: *Zarautzeko* (*e* epentetikoa gehituz: 2. erregela) edo *Zarauzko* (afrikatua txistukari bihurtuz: 8. erregela). Kasu honetan ez da baldintza bete.

```

amarekinko
*****

```


9. erregelak dio bi atzizki metatzen direnean, lehenengoak sudurkaria galtzen duela. Ez da, beraz, baldintza bete: *amarekiko* behar zuen.

amentzat
1. ama [IZE] [ARR] +e [MP] +entzat [DES]

a organikoa duen ama formari plurala gehitzean, 6. erregelak dioskunez, *a* galdu egiten da eta *r* epentetikoa ez da azaleratzen.

atzoko
1. atzo [ADB] [ADO] +ko [DES]
2. atzo [ADB] [ADO] +ko [GEL]

atzo aditzondoak bi analisi posible ditu: destinatiboa (*atzoko egin behar zenuen lana*) eta leku-genitiboa, izenlagungilea (*atzoko filma*).

beranduena
1. berandu [ADB] [ADO] +en [GRA] [SUPE] +a

Aditzondoak gradua har dezake eta informazio morfologikorik ez duen *a* ere bai (berez, absolutibotzat jo izan dena, partitiboarekin batera (*beranduenik*) aditzondoei erants baitakieke).

berhasi

berrasi
1. ber [AUR] +has [ADI] [ARR] +i [ADM] [PART]
2. ber [AUR] +has [ADI] [ARR] +i [ADM] [PART] + [ASP] [BURU]
3. ber [AUR] +has [ADI] [ARR] +i [ADM] [PART] + [MG] + [ABS]

hasi aditza aurrizkiarekin bilduz (eta 10. erregela aplikatuz), hiru analisi eskuratuko ditugu: a) partizipio hutsa, jokatugabea; b) aspektu burutua; eta c) adjektibo gisa jokatzeko duen partizipioa, absolutibo mugagabearen deklinatua.

bestela
1. bestela [ADB] [ADO]
2. bestela [LOT] [LOK]

bestela-k bi analisi ditu: aditzondoa eta lokailua. Alegia, bi sarrera lexiko dira.

bosgarren
1. bost [DET] [ZNB] [DZH] +garren [ZNB] [ORD]

boskarren

bostgarren

bost zenbatzaile zehaztuak zuzenean hartzen du ordinala, eta, ondorioz, azpikategoria berria hartzen du. Baina aldaketak eragiten ditu honek: *t* galdu egiten du (8. erregela).

daramala
1. darama [ADT] +la [ERL] [MEN] [DENB]
2. darama [ADT] +la [ERL] [MEN] [KNPL]
3. darama [ADT] +la [ERL] [MEN] [MODU]

daramela

eraman aditzaren *daramala* forma trinkoak (*darama* zuzenean hiztegitratuak) hiru analisi posible ditu, hiru menderagailu desberdin: denborazkoa, konpletiboa eta moduzkoa. Bestalde, adizkietan *da:del*a modukoak gertatzen badira ere, [*Aadi*] diakritikoari esker ez da 11. erregela aplikatzen.

dioela

1. dio [ADT] +la [ERL] [MEN] [DENB]
2. dio [ADT] +la [ERL] [MEN] [KNPL]
3. dio [ADT] +la [ERL] [MEN] [MODU]

diola

1. dio [ADL] +la [ERL] [MEN] [DENB]
2. dio [ADL] +la [ERL] [MEN] [KNPL]
3. dio [ADL] +la [ERL] [MEN] [MODU]

dio adizki jokatuak bi dira, homografoak: laguntzailea eta *esan* aditzaren forma trinkoa. Batari zein besteari *-la* atzizkia eranstea bueltatuko digun informazioa, halere, bera da: denborazkoa, konpletiboa eta moduzkoa. Baina laguntzaileak *e* epentetikoa behar du gauzatzeko eta trinkoak, berriz, ez.

eginten

egiten

1. egin [ADI] [ARR] + [ADM] [ADOIN] +ten [ASP] [EZBU]
2. egin [ADI] [ARR] + [ADM] [ADOIN] +ten [ERL] [MEN] [KNPL]
3. egin [ADI] [ARR] + [ADM] [ADOIN] +ten [ERL] [MEN] [MODU]

egin aditzak hiru analisi posible ditu *-ten* atzizkia gehitzean: aspektu burutugabea, konpletiboa eta moduzkoa, azken bi hauek menderakuntzazkoak. Horrez gain, 9. erregelari esker, azaleratzeko egokitzapena egin du, *n* galaraziz eta, ondorioz, *eginten* forma baztertuz.

egundik

1. egun [ADB] [ADO] +tik [ABL]

egunetik

1. egun [IZE] [ARR] + [MS] +tik [ABL]

eguntik

egun bi sarrera dira datu-basean: izena eta aditzondoa. Biei ablatiboa gehituz, aditzondoak kasu-marka zuzenean hartuko du, mugatasuna berezkoa du-eta; izenak, aldiz, mugatu singularra erantsi behar dio aurretik, nahiz *0* morfema bat izango den (baina informazioa emango dio). Azaleratzean ez dute berdin jokatzen: aditzondoak ez du epentesirik hartzen eta atzizkiaren leherkari gorra ozen egiten du, 13. erregela aplikatuta. Izenak, aldiz, *e* epentetikoa eransten du, atzizkia bere osotasunean gordeaz. Hau 2. erregelak bideratzen du.

etxeko

1. etxe [IZE] [ARR] + [MG] +ko [BAN]
2. etxe [IZE] [ARR] + [MS] +ko [DES]
3. etxe [IZE] [ARR] + [MS] +ko [GEL]
4. etxeko [IZE] [ARR]
5. etxeko [IZE] [ARR] + [MG] + [ABS]

etxekoa

1. etxe [IZE] [ARR] + [MS] +ko [GEL] + [ELI_IZE] +a [MS] + [ABS]
2. etxeko [IZE] [ARR] + [MS] + [ABS]

etxeko formak bi hiztegi-sarrera, bi izen erakusten ditu: *etxe* eta *etxeko*. Baina bost analisi ditugu: *etxe* sarrerak hiru (banatzailea, mugagabea, destinatiboa eta leku-genitiboa, biak singular) eta *etxeko*-k bi (forma lexiko hutsa eta absolutiboa, mugagabea). Aldiz, *etxekoa* formak bi besterik ez ditu: *etxe* leku-genitiboaren gain sortutako izenlagunari gehitutako izen eliditu mugatua, absolutiboan (*etxekoa*: *etxeko atea*, *esaterako*) eta *etxeko* izena (senitartekoa), absolutibo singularrean.

ezjakin

1. ez [AUR] +jakin [ADI] [ARR]
2. ez [AUR] +jakin [ADI] [ARR] + [ADM] [ADOIN]

3. ez [AUR] + jakin [ADI] [ARR] + [ADM] [PART]
4. ez [AUR] + jakin [ADI] [ARR] + [ADM] [PART] + [ASP] [BURU]
5. ez [AUR] + jakin [ADI] [ARR] + [ADM] [PART] + [MG] + [ABS]
6. ezjakin [ADJ] [IZO]
7. ezjakin [ADJ] [IZO] + [MG] + [ABS]

Zazpi analisi ditugu *ezjakin* formarentzat: *ez-* aurrizkia hartu duen *jakin* aditzarentzat bost: 1) forma lexiko hutsa; 2) aditzoina; 3) partizipioa; 4) aspektu burutua; eta 5) adjektibo funtzioan, absolutibo mugagabea. Izenondoak bi analisi posible ditu: forma lexiko hutsa, informazio morfologikorik gabea, eta absolutibo mugagabea.

```
harek
*****
hark
1. hura [DET] [ERK] [MSL] +k [ERG]

hura
1. hura [DET] [ERK] [MSL] + [ABS]
```

Hura determinatzaile erakusleak alomorfoak erabili behar izaten ditu absolutiboa ez beste kasu guztietan, baina analisiaren emaitzan ez dugu alomorfoaren berri izan beharrik, bigarren adibidean ikusten denez. Determinatzaileek mugatasun lexikala dute eta, ondorioz, kasu-markak zuzenean hartzen dituzte. Bestalde, *[EpBer]* diakritikoari esker, ez da epentesiaren erregela aplikatzen.

```
lehenengo_eta_behin
1. lehenengo_eta_behin [ESR] [ADB] [ADO]
```

Hitz anitzeko unitate lexikaletako osagaiak azpimarraren bidez lotzen dira. Kasu honetan, esaera baten aurrean gaude, baina morfotaktikarako behar dugun informazioa azpikategoriak damaigu: aditzondoa.

```
milana
*****
```

Analisirik ez du, *mila* sarreraren morfotaktikan *-na* banatzailea falta delako. Izan ere, *bana*, *bina*... baina *mila bana* agindua du Euskaltzaindiak.

```
txikitxoena
1. txiki [ADJ] [IZO] +txo [ATZ] [TXIK] +en [GRA] [SUPE] +a [MS] + [ABS]
2. txiki [ADJ] [IZO] +txo [ATZ] [TXIK] +e [MP] +en [GEN] + [ELI_IZE] +a [MS] + [ABS]
```

Bi analisi ditugu izenondoari erantsitako atzizki txikigarritik abiatuta: a) gradu superlatiboan, absolutibo mugatu singularra (*txikitxoena etorri da*); eta b) genitibo plurala izenlagun bilakatu da eta honi elipsia erantsi zaio, azken emaitza absolutibo singularra bada ere (*haur txikitxoena jolasleku*) a da hori).

```
txikitxoagoarekikoek
1.
txiki [ADJ] [IZO] +txo [ATZ] [TXIK] +ago [GRA] [KONP] +a [MS] +ekin [SOZ] +ko [GEL] +
[ELI_IZE] +a [MS] + [ABS]
```

Atzizki-metaketa handiagoa da adibide honetan, baina zuzena da analisisa: sozietiboari leku-genitiboa erantsiz izenlaguna lortu dugu. Hemendik abiatuta, izen eliditua erabiliz, kategoria berria eskuratu dugu, ergatibo pluralean.

```
zorroztenago
1. zorrotz [ADI] [ARR] + [ADM] [ADOIN] +ten [ASP] [EZBU] +ago [GRA] [KONP]
```

Aditzoinaren gain sortzen den aspektu burutugabeak gradua har dezakeela frogatzen du adibide honek.

Adibide hauekin frogatu dugu, neurri batean behintzat, kapituluaren zehar aurkeztu duguna. Hala ere, adibide hutsak dira hauek eta prozesadorearen benetako balioa corpus

oso batean probatzean ikusten da, adibideak ez baitira “aukeratuak” izaten, testu errealak baizik. Dena dela, aurreko orrietan aipatu dugu corpus errealean estaldura eta zehaztasunaren emaitzak oso onak direla.

Kontuan izan beharreko beste atal bat eraginkortasunarena da. Izan ere, adibide gutxiarekin abiadura garrantzizkoa ez bada ere, corpusetan abiadura egokia mantentzea da helburua, eta hala gertatzen da lexiko-itzultzaileek duten ezaugarriei esker. Horren erakusgarri, *Xuxen* zuzentzaile ortografikoa bera da. Prozesadorea *Xuxenen* integratua dago eta ez du eraginkortasunean eraginik.

5.5 Hobekuntza-proposamena: informazio egituratuaren beharra

Aurreko atalean sorkuntza eta analisiaren emaitzak ikusi ditugu eta formalismoa bietarako egokia dela erakutsi ere bai, nahiz muga batzuk aipatu ditugun. Beraz, gure helburua, neurri batean behintzat, bete da. Baina, morfologia hutsa gainditu eta morfosintaxiaren beharraz ere aritu gara gorago. Bada, ez dugu morfema-zatikatzailerik bakarrik osatu nahi, informazio-metaketa garrantzitsua baita: alegia, osagaiak egituratzea. Puntu honetan sorkuntza alde batera utzi eta analisisira mugatuko gara: analisi egokia eta osoa eskuratzen saiatuko gara, hau da, azken emaitzak erakusten edo “goratzen” duena azaleratzen, analisiaren irteerak sendoa eta osoa behar baitu.

Sistema lexikoa aurkeztean, azpilexikoak guztiz banatuta ezarri ditugu morfotaktika modu gardenean aplikatzeko. Baina analisiaren emaitza osoa nahi badugu, ez digu askorako balio genitiboaren gain eraikitako motibatiboak genitiboa ere baduela esateak, adibidez, osaerari baitagokio kontua —ez da, gainera, beti genitibotik pasata osatzen, ikusi ahal izan dugunez—. Hortaz, halako kasuak soildu egingo ditugu maila honetan, nahiz zatikatzailerik beti guztiaren berri emango digun. Ber gauza dugu hitzaren barruko elipsian ere, ondo egituratua izatea garrantzitsua baita. Honek erregela morfosintaktikoak sortzera eramango gaitu, baterakuntza-formalismoan oinarrituta, bi mailatako formalismoan hau egitea ezinezkoa baita (edo, posible izanez gero, oso modu zeharkakoan egin beharko bailitzateke).

Morfologia bera gainditzeko badu ere, hitzaren azken analisi osoa eta egituratua lortzeko morfosintaxira jo beharko dugu. Hiru kasuistika desberdin konpondu beharko ditu honek: a) kasu-marken metaketa, bai automatikoa denean, bai testu-hitz oso baten

gainekoa denean, birdeklinabidea deitu izan dena; b) muga(gabe)tasun lexikala eta analisitik datorren muga(gabe)tasuna; eta c) hitzaren barruko elipsia.

5.5.1 Kasu-marken metaketa

Kasu-markak maila desberdinetan bil daitezke; izan ere, ez da gauza bera aipatu berri dugun motibatiboaren osaera edo birdeklinabidetzat hartu izan dena —aurrekoa ere birdeklinabidea dela esan daitekeen arren (Rebuschi, 1984)—. Lehenengoan testu-hitz baten osaera-oinarrian gaude, baina bigarrenean sortua dugun testu-hitz deklinatu bati beste atzizki bat erantsi diogu, izenlagungilea, hain zuzen ere. Ikus ditzagun banaka:

a) Genitibotik abiatuta osatzen diren kasuak: bilakaera automatikoa⁵⁹

1. Mugatasunaren eta motibatiboaren artean genitiboa dagoenean

Motibatiboa ez da, bakarrik, genitiboaren gain osatzen; kategoria hauekin aukeran dago: pertsona-izen bereziak (*Erkuden(en)gatik*), leku-izenak (*Urkulu(ren)gatik*), izen arruntak⁶⁰ —bizidunak (*semea(ren)gatik*) zein bizigabeak (*mendia(ren)gatik*)—, pertsona-izenordainak (*ni(re)gatik*) eta determinatzaile erakusleak (*har(e(n))gatik*).

Hortaz, proposa dezakegun erregela orokorra hau da: genitiboa eta motibatiboa elkarren segidan ditugunean, azken analisisan motibatiboa bakarrik jasoko dugu, hori baita benetako informazio baliagarria.

$$\text{GEN} + \text{MOT} = \text{MOT}$$

Adibidez, *Erkudenengatik* testu-hitzaren informazio gordina —segmentatzaile morfologikotik datorriguna— hau da:

Erkudenengatik

(Erkuden [IZE][PIB][+biz][MS]) (en [GEN]) (gatik [MOT])

baina erregela aplikatu ondoren hau emango digu:

Erkudenengatik

(Erkudenengatik [IZE][PIB][+biz][MS][MOT])

⁵⁹ Genitiboaren gain sortutako kasu-markak automatikoak izan ohi dira —elipsirik ez dagoenean, noski—, baina hauen artean ere banaketa egin beharrean gaude. Izan ere, batzuk lexikoan elkartuta eskaini ditugu, sozietiboa eta destinatiboa, esaterako, azpian dagoen genitiboa erabat ihartua dagoelako (*-en+-kin*: *-ekin*; *-en+-tzat*: *-entzat*), besteak beste *-kin* eta *-tzat* (ez da prolatiboarekin nahasi behar) ez direlako berez ezer, genitiboarekin batera ez bada. Ez dugu beharrezko ikusi puntu horretarainoko banaketa egitea, ez baita diakronia gure helburua. Baina beste kasu batzuetan ez da hori gertatzen. Adibidez, motibatiboa osatzeko genitibotik pasa gaitzke, eta hala eskatzen da batzuetan gainera —mugatu pluralean, besteak beste—, baina beste batzuetan mugatasunetik zuzenean iritsi gaitzke kasu honetara (*amagatik* nahiz *amarengatik*).

⁶⁰ Bizidunek aukeran dute genitiboa bai mugatuta daudenean, bai mugagabeen. Ez, aldiz, bizigabeek, mugatu pluralak beti eskatzen baitu genitiboa motibatiboaren aurretik (*mendiengatik*, ez **mendiegatik*, nahiz *mendiakatiko* moduko osakeretan ez den genitiboa azaltzen, hain zuzen ere, mugatasuna eta numeroa absolutiboan bezala mantendu delako (*-ak*)). Baina hau morfotaktikak kontrolatzen duenez, ez dugu hemen arazorik izango.

Eta analisi hau izango da ondoko edozein aplikaziotarako erabiliko duguna.

2. Mugatasunaren eta bizidunen leku-denborazko kasuen artean genitiboa ageri denean

Jakina da leku-denborazko kasuetan deklinabide-sistemak bi bide jarraitzen dituenak, bizidunak eta bizigabeak bereiziz. Bada, bizidunek *-gan-* daramate kasuari itsatsita, eta guk ere halaxe jaso ditugu kasuok, banaketarik gabe. Baina hauek genitibotik pasata osa daitezke, motibatiboarekin ikusi dugun bezalatsu. Hain zuzen ere, aurreko atalean aipatutako kasu gehienetan gertatzen da hau: pertsona-izen berezietan (*Erkuden(en)gana*), izen arruntetan⁶¹ —baina bizidunetan bakarrik (*semea(ren)gan*)—, pertsona-izenordainetan (*gu(re)gana*) eta bizidunekin erabiltzen diren determinatzaile erakusleetan (*har(en)gandik*). Murritzapena ezarri dugu halabeharrez, leku-izen bereziek, izen arrunt bizigabeek eta erakusleek (bizigabeekin doazenean) ez dute-eta hemen lekurik, [+biz] tasuna ez dutelako. Proposa dezakegun erregela, beraz, hau da:

$$\text{GEN} + \text{LekKas} (+\text{biz}) = \text{LekKas} (+\text{biz})$$

Adibidez, Erkudenengan:

Erkudenengan
(Erkuden [IZE][PIB][+biz][MS]) (en [GEN]) (gan [INE])

baina erregela aplikatu ondoren:

Erkudenengan
(Erkudenengan [IZE][PIB][+biz][MS][INE])

b) Forma deklinatuei leku-genitiboa erantsiz izenlagunak sortzen direnean

Desberdina da metaketa-kasu hau; izan ere, lehendik osatua —analizatua— dugun testu-hitz bat izenlagun izatera pasako da *-ko* leku-genitiboa eranstea. Beraz, analisiaren irteerarako azken kasuaren informazioa (*GEL*) nahikoa da, lehen osagaiaren mugatasuna eta numeroa mantenduz. Ez da bi kasuen artean informaziorik tartekatzen, erregela honetan ez dago-eta elipsirik. Osaera-prozesua aurrekoaren desberdina den arren, emaitza bera izango da eta azken kasua bakarrik goratuko dugu. Erregela, hortaz, hauxe izango da:

$$\text{DEK} (-\text{ELI_IZE}) + \text{GEL} = \text{GEL}$$

Zein kasurekin gertatzen da hau? Adlatiboarekin (*etxerako*, *amarenganako*) eta honen eratorri diren muga- eta hurbiltze-adlatiboarekin (*etxerainoko*, *etxeranzko*);

⁶¹ Motibatiboarekin mugagabeen genitiboaren erabilera librea bazen, hemen ez daude kontuak hain garbi; bai, ordea, mugatu singularrean eta pluralean. Euskaltzaindiak *gizonengan* baina *alaba(ren)gan* adibideak ematen ditu, kontsonantez bukatuei ia genitiboa eskatzen. Baina hori ere morfotaktika bidez konpondu dugu eta, hortaz, segmentatzaileak egingo du lan hori. Hemen osagaiak pilatu besterik ez ditugu egingo.

ablatiboarekin (*etxetiko*); soziatiboarekin (*etxearekiko*); motibatiboarekin (*etxeagatiko*); destinatiboarekin (aldaera batzuetan bakarrik: *amarentzako*, *zuretako*) eta prolatiboarekin (*amatzako*), leku-denborazko kasuetan bizidun edo bizigabe diren kontuan hartu gabe, bietan gertatzen baita—.

Genitiboarekin —bai jabego-genitiboa, bai lekuzkoa— lotura berezia dute kasu hauek guztiek. Aurreko orrietan aipatu ditugun kasu berekin egingo dugu topo berriro ere: leku-denborazkoak, motibatiboa, soziatiboa eta destinatiboa. Hauek jabego-genitiboaren gain eraikitzen dira —batzuk nahitaez, besteak aukeran—, eta guztiek (eta ez bestek, prolatiboa salbu) har dezakete ondoren leku-genitiboa elipsirik gabe.

Baina *-ko* izenlagungilea ez dute aipatu ditugun kasu-markek bakarrik hartzen; izan ere, menderagailua duten adizkiak ere izenlagun bilaka baitaitezke atzizki hau berau gehituta: *esanda* -> *esandako*, *duela* -> *duelako*. Kasuotan aditzaren informazioa goratuko dugu, ondoko aplikazioetan baliagarri izan daitekeelako.

5.5.2 Mugatasun-metaketa

Mugatasun lexikala aipatu dugu gorago eta lexikoan tasun hau dutenek kasu-marka zuzenean hartzen dutela baieztatu dugu gainera. Baina determinatzaile zehaztuek arazoak dituzte hau aplikatzean; izan ere, determinatzaile eta zehaztu izateak berez mugatu lexikal izatea eskatzen duen arren, hauek mugagabeen erabil baitaitezke (ikus Euskaltzaindia, 1993:81). Horien tratamendua argitzen saiatuko gara atal honetan.

Gogora dezagun, halere, noiz dugun muga(gabe)tasun lexikala, kasuan kasuan printzipioa ezarri ahal izateko:

1. Izen bereziak —pertsona-, leku- eta erakunde-izenak— eta pertsona-izenordainak beti dira mugatu lexikal, erreferentzia bakarrekoak⁶² dira-eta. Hauek, beraz, kasu-marka hartuko dute zuzenean eta ez dute mugatasun/numerotik pasa beharko, tasunen barruan *mnl* (mugatasun/numero lexikala) baitute (mugatu singularra, *MS*, edo plurala, *MP*). Baina *mugatu plurala* dutenen morfotaktika pluraletik osatua da eta, beraz, bi mugatasun-marka izango dituzte (*mnl* eta *mun*). Hauen analisiaren irteeran lexikala besterik ez dugu eskainiko, besteak ez du-eta garrantzirik.

$$(LIB/IZB) \text{ MNL} + \text{MUN} = \text{MNL}$$

⁶² Erreferentzia bakar hori multzo bati dagokiona ere izan daiteke, *Alpeak* bezalakoetan; hauek mugatasunean plurala dela zehaztuta izango dute eta morfotaktika ere pluralarekin osatuko dute.

Pertsona-izenordainek *Pertsona* eremua ere badute, zein den zehaztuz. Izenordain zehaztugabeek *mnl mugagabea* dute, aurrekoen prozesu bera jarraituz.

2. Determinatzaileak: "determinatu" egiten dutenez, *mnl* eremua dute eta zuzenean hartzen dute kasua: zehaztugabeak beti dira mugagabe eta erakusleak eta zenbatzaile zehaztuak beti mugatu. Baina zenbatzaile zehaztuak, mugatasun lexikala izan arren, mugagabeen zein pluralean deklina daitezke (*hiruk* eta *hirurek*), nahiz kasu honetan izenordain bezala funtzionatu duten⁶³. Muga(gabe)tasunari buruzko informazioa lexikoan eta analisitik jasoa dugunean, beraz, analisiaren irteeran biak eskainiko ditugu, izen berezietan ez bezala, informazio osoa eskaintzearren, Euskaltzaindiaren irizpideak jarraituz. Baina aipatu berri dugun kasuan besterik ez dugu halakorik izango; bestelako kasuetan muga(gabe)tasun lexikala da analisiaren irteerak emango duena, besterik ez duelako.

Printzipio orokor bezala, beraz, hau proposa dezakegu: EDBLko *mnl* eremua hutsa denean, morfotaktikatik datorkiguna goratuko dugu analisiaren irteerarako; bestela, datu-basekoa da nagusi —izen berezietan aipatu duguna—.

Zer egin, beraz, determinatzaile zehaztuekin? Ez dugu mugatasun lexikala galdu nahi, baina mugagabe ala plural bezala jokatzaren duten ere esan nahi dugu —horregatik landu ditugu mugagabea eta mugatu plurala morfotaktikan, mugatasun lexikala duten besteei kasua hartzen baitute zuzenean—. Hortaz, erregela bat sortu beharko dugu: datu-baseko informazioa beti jasoko da, baina honekin batera deklinabidekoa ere azken emaitzara pasako da: alegia, *hiru* zenbatzaile zehaztuaren analisisa hau izango da:

hiruk

(hiru [DET][DZH][MP]) (0 [MG]) (k [ERG])
(hiruk [DET][DZH][MP][MG][ERG])

eta *hirurek*-ena⁶⁴,

hirurek

(hiru [DET][DZH][MLP]) (e [MP]) (k [ERG])
(hirurek [DET][DZH][MLP][MP][ERG])

Erabakia, beraz, hau da:

⁶³ Baina ez da beti betetzen baieztapen hau; *bi* zenbatzailea, esaterako, izenaren ezkerrean zein eskuinean erabil baitaiteke, eta orduan ez da izenordain moduan azaltzen: *lagun bik* / *lagun biek* (ZNB) edo *bi lagunak* (ZNB) -> *bik* / *biek* (IOR). *Bat* ere, hainbat erabileratakoa den arren, eskuinean ageri da beti (**bat lagun*, baina *lagun bat* (ZNB) -> *bat* (IOR)). Besteetan, aldiz, beti dago izen eliditua: **lagun hiruk*, baina *hiru* (ZNB) *lagunek* -> *hiruk* / *hirurek* (IOR).

⁶⁴ *Hiruek* testu-hitza ere bada, baina orduan *hiru* izena da eta ez zaio *r* zaharraren erregela aplikatzen.

(DET ZNB DZH) MNL + MUN = (DET ZNB DZH) + MNL + MUN

Azkenik, determinatzaile orokorrek ez dute *mnl* eremua beteta, *null* balioarekin baizik, eta morfotaktikatik datorkiena besterik ez dute jasoko. Hauek muga(gabe)tasuna baino gehiago orokortasuna beharko lukete, semantikoki [-sing] edo izango litzatekeena, baina singularrean ere erabiltzen dira bestalde: *guztia*, *guztiak*.

5.5.3 Hitzaren barruko elipsia

Hitzaren barruan elipsi-kasuak ditugunean, zein da eskaini beharreko azken analisia? Analisi batek elipsia noiz dagoen esan behar du, besteak beste analisi sintaktikoan-eta erabilgarri izango delako.

Kasu-pilaketan ikusi dugun bezala, hemen ere genitiboaren aurrean gaude, baina elipsia jabego- zein leku-genitibotik abiatuta dugu gainera: genitiboaren atzetik mugatasuna ezartzen dugunean elipsiaren aurrean gaude, eta beti behar du mugatu izatea, ezkutuan dagoena ezaguna baita —eta izena ia kasu guztietan—. Baina badira hemen ere salbuespenak: partitiboa eta prolatiboa, mugagabetik abiatzen diren kasu-marka zalantzazkoak.

Ez da deklinabide-kasuen metaketa elipsia ezkututzen duen bakarra; izan ere, adizki jokatuak beti dute izen bat isilpean erlatiboaren ondoren: *zuen0a*. Hori ere kontuan hartu beharrekoa izango da. Are gehiago, determinatzaileek ere, izenordain bezala funtzionatzen dutenean, izen bat dute ezkutuan eta, beraz, elipsiaren aurrean gaude berriro (*hiru lagunek: hiruk, hirurek*).

Kasu hauei guztiei irtenbidea eman beharko diegu. Baina, zein da eskainiko dugun azken emaitza? Elipsiaren berri emango dugu eta hori izango da, hain zuzen ere, banaketa-irizpide: elipsiaren aurrekoa landuko dugu batetik eta ondokoa bestetik, gerora erabilgarri izan dadin. Bi analisi batean eskainiko ditugu, hortaz. *Erkudenena* formaren analisia —segmentatzaitetik jasoa lehendabizi, analisiaren ondokoa gero— hau izango da:

Erkudenena

(Erkuden [IZE][PIB][+biz][MS]) (en [GEN]) (0 [ELI_IZE]) (a [MS]) (0 [ABS])

Erkudenena

(Erkuden [IZE][PIB][+biz][MS][GEN]) (a [ELI_IZE][MS][ABS])

Aurreko erabaki guztiak aplikatuz, *Erkudenenganakoarekin*, alegia, "Erkudenenganako (maitasun)arekin" horrela analizatuko dugu zatikatzailearen informaziotik abiatuz:

Erkudenenganakoarekin
(Erkuden [IZE][PIB][+biz][MS]) (en [GEN]) (gana [DES]) (ko [GEL]) (0 [ELI_IZE])
(a [MS]) (ekin [SOZ])

Erkudenenganakoarekin
(Erkudenenganako [IZE][PIB][+biz][MS][GEL]) (arekin [ELI_IZE][MS][SOZ])

Printzipioa, beraz, horrela defini dezakegu:

$$\text{KAT} + \text{MNL/MUG} + \text{GEN/GEL} + \text{ELI_IZE} + \text{MUG} + \text{KM} = \\ \text{OSAG1 (KAT MNL/MUN GEN/GEL)} + \text{OSAG2 (ELI_IZE MUN KM)}$$

Eta ber gauza adizki jokatuekin⁶⁵ ere:

zuena
(zuen [ADL]) (en [ERL][MEN][ERLT]) (0 [ELI_IZE]) (a [MS]) (0 [ABS])

zuena
(zuen [ADL][ERL][MEN][ERLT]) (a [ELI_IZE][MS][ABS])

Azken emaitza ez da, berez, adizkia, baina izen eliditua bakarrik esatea ere ez da zilegi kasu askotan —lemaren informazioa eskaintzerakoan, adibidez—. Printzipio orokorra goikoaren parekoa izango da:

$$\text{ERL MEN} + \text{ERLT} + \text{ELI_IZE} + \text{MUN} + \text{KM} = \\ \text{OSAG1 (ERL MEN + ERLT)} + \text{OSAG2 (ELI_IZE MUN KM)}$$

Hiru puntuok argituta, eta hartu beharreko erabakiak hartuta, analizatzaile morfosintaktikorako bidea prest dugu, beste guztia baterakuntza-erregela bidez konponduko dugu-eta. Baina ez gara hemen baterakuntza-formalismoa⁶⁶ azaltzen hasiko, haren sintaxia azaltzea konplexua da-eta. Beraz, adibide batzuk besterik ez ditugu bistaratuko. Dena den, analisiaren irteera sendoa dela erakusten du formalismo honek.

5.5.4 Adibideak

Ondoko orrietan erregela morfosintaktikoen konplexutasuna eta ahalmena adierazten duten lau adibide azaltzen dira. Erregelak aplikatu ahala informazioa hautatzen eta

⁶⁵ Adizki jokatuekin bada, halere, arazotxo bat: ezar al daiteke erlatiboaren ondoren edozein mugatasun- edo kasu-marka, adizkia osatzen duten pertsonak kontuan hartu gabe? Alegia, *etorri denek, dizkiodanekin* modukoak ontzat eman behar al ditugu? Jakina, aditzaren azpikategoriazioa landu beharko genuke aurrena murriztapenak ezarri ahal izateko, baina oraingoz zabal jokatu dugu eta edozer onartu. Badirudi kasu gramatikalek gutxienez bat etorri behar dutela, numeroan behintzat, elipsiaren ondoko numeroarekin, baina hor ere badira salbuespenak: *ditugunarekin*, esaterako, *guk(nrk)-haiek(nor)* da, pluralean pertsona gramatikal guztiak, baina gaitzetsiko al genuke "liburuak erosi ditugun lagunarekin joan gara" modukoak? Alegia, "liburuak erosi *ditugunarekin* joan gara"? Are garbiagoa da lekua eta denbora adierazten duten isileko izenekin erabiltzen denean: *zenituztena*, adibidez ("liburuak erosi zenituztena" (zenituzten eguna)), edo *nintzenak* ("oporretan egon nintzen asteak"), etab. Oraingoz bide irekia da hau, bitxiak izan arren gramatikalki ondo eratuak baitaude gure ustean; baina azpikategoriazioa lantzean berrikusi beharra izango da.

⁶⁶ Erregelen azalpen osoa ikusteko, Urkia (1997), 6. kapitulura jo daiteke.

goratzen doa, azken erregelaren ondorioz lortutako informazio morfosintaktikoa hitzari esleitzen zaiona izanik.

1. adibidea: umearenganako ((((ume) a) (en + gana)) ko)

ume	a	en	gana	ko
kat:IZE	kat:MUN	kat:DEK	kat:DEK	kat:DEK
azp:ARR	mun:MS	kas:GEN	kas:DES	kas:GEL
biz:+	hurm:	sar:en	sar:gana	sar:ko
zenb:+	sar:a			
pl:				
sar:ume				

1. erregela

engana
kat:DEK
kas:DES
sar:engana

2. erregela

umea
kat:IZE
azp:ARR
biz:+
zenb:+
mun:MS
sar:ume

3. erregela

umearengana
kat:IZE
azp:ARR
biz:+
zenb:+
mun:MS
kas:DES
sar:umearengana

4. erregela

umearenganako
kat:IZE
azp:ARR
biz:+
zenb:+
mun:MS
kas:GEL
sar:umearenganako

2. adibidea: txikitxoagook ((((txiki) txo) ago) o) k)

Erregela honetan garbi asko ageri da ezkerretik eskuinerako ordena, hizkuntza eranskarien ezaugarri dena.

txiki	txo	ago	o	k
kat:ADJ	kat:ATZ	kat:GRA	kat:MUN	kat:DEK
azp:IZO	e_kat=o_kat	grdm:KONP	mun:MPH	kas:ERG
sar:txiki	gra:TXIK	sar:ago	hur:+	sar:k
	erat:+		sar:o	
	sar:txo			

19. erregela

txikitxo

kat:ADJ
azp:IZO
gra:TXIK
erat:+

sar:txikitxo

13. erregela

txikitxoago

kat:ADJ
azp:IZO
gra:TXIK
erat:+

grdm:KONP
sar:txikitxoago

2. erregela

txikitxoagoo-

kat:ADJ
azp:IZO
gra:TXIK
erat:+

grdm:KONP
mun:MPH
hur:+

sar:txikitxoagoo

3. erregela

txikitxoagook

kat:ADJ
azp:IZO
gra:TXIK
erat:+

grdm:KONP
mun:MPH
hur:+

kas:ERG
sar:txikitxoagook

3. adibidea: *pentsaraztearren* (((pentsa) araz) 0) tearren)

pentsa	araz	0	tearren
kat:ADI	kat:ATZ	kat:ADM	kat:ERL
azp:SIN	o_kat:ADI	adm:ADOIN	azp:MEN
sar:pentsa	e_azp:FAK	sar:0	erl:HELB ⁶⁷
	erat:+		sar:tearren
	sar:araz		

20. erregela

pentsaraz

 kat:ADI
 azp:FAK
 erat: +
 sar:pentsaraz

5. erregela

pentsaraz

 kat:ADI
 azp:FAK
 erat: +
 adm:ADOIN
 sar:pentsaraz

7. erregela

pentsaraztearren

 kat:ADI
 azp:FAK
 erat: +
 adm:ADOIN
 erl:HELB
 sar:pentsaraztearren

⁶⁷ Helburuzkoa aukeratu dugu analisi honetarako, baina kausazkoa ere izan zitekeen. Bi analisi ditugu, beraz.

4. adibidea: *dakidantxoa* (((((dakit) n) txo) a) 0)⁶⁸

dakit	n	0	txo	a	0
kat:ADT oin:JAKIN mdn:A1 nrk:NIK nor:HURA asp:EZBU sar:dakit	kat:ERL azp:MEN erl:ERLT sar:n	kat:ELI_IZE sar:0	kat:ATZ gra:TXIK e_kat=o_kat erat:+ sar:txo	kat:MUN mun:MS hurm: sar:a	kat:DEK kas:ABS sar:0
<i>10. erregela</i>		<i>19. erregela</i>			
dakidan kat:ADT oin:JAKIN mdn:A1 nrk:NIK nor:HURA asp:EZBU erl:ERLT sar:dakidan		txo kat:ELI_IZE gra:TXIK erat:+ sar:txo			
		<i>2. erregela</i>			
		txoa kat:ELI_IZE gra:TXIK erat:+ mun:MS sar:txoa			
		<i>3. erregela</i>			
		txoa kat:ELI_IZE gra:TXIK erat:+ mun:MS kas:ABS sar:txoa			
		<i>16. erregela</i>			
		dakidantxoa kat:ADT_EIZ gra:TXIK erat:+ mun:MS kas:ABS osag1_kat:ADT oin:JAKIN, mdn:A1 nrk:NIK, nor:HURA asp:EZBU, erl:ERLT sar:dakidan osag2:elipsia sar:dakidantxoa			

⁶⁸ Erlatiboa hartu duen adizki jokatuari ere ezar dakiok atzizki apreziatiboa, tartean izen eliditua baitugu. Beraz, atzizkia ezkutuan dagoen izenari dagokio, ez adizkiari.

6 Estaldura osorako urratsak

6.1 Sarrera

Aurreko kapituluaren euskararako prozesadore morfologiko estandarra ikusi dugu, hizkuntza estandarra analizatzeko/sortzeko beharko litzatekeen bakarra. Baina kapituluaren hasieran ikusi dugu estandarizazioa ez dela erabatekoa eta, hori kontuan hartuta, irtenbide bat bilatu behar da euskara-erabiltzaileen erabilera guztiak tratatu ahal izateko eta, bide batez, egokiak proposatzeko. Horiek horrela, euskararen tratamendu automatikorako arkitektura orokorra ikusiko dugu, hiru urrats inkrementaletan gauzatzen dena eta, orain arte, lehenengoa besterik ez dugu ikusi. Ondoren, horiek banaka aurkeztuko ditugu, alegia, tratamendu estandarretik abiatu eta, euskararen egoera kontuan hartuta, aldaeren tratamendua egiten saiatuko gara, aldaera dialektalak edo/eta ezagutza faltak eragindako akatsak detektatu ahal izateko. Hala ere, izango dira lexikoan bildu gabeko hitzak ere, sistemak ezagutzen ez dituenak: horiek lexikorik gabeko analisia eginez lortuko ditugu, aurkeztu dugun sistema sendoa eta osoa dela erakutsiz.

Euskararen normalizazioan aurrerapen handia egin bada ere, oraindik bada idazkerako azken arauak bere egin ez dituenik (edo arautze bidean diren proposamenak, erabiltzailearengana iritsi ez direnak); horrez gain, euskalkietako erabilerak baliatzen dituenik ere bada. Beraz, ezezagutzak eragindakoak izan zein berariazko aldaerak izan, eragina dute analizatzailearen diseinuan, eta horiek guztiak kontuan hartu behar dira tresna erabilgarria eta sendoa izango bada. Euskara erabat estandarrarekin ez dago beste ezeren beharrik, baina desbiderapen horiek sistema ez-osoak egiten dute egungo euskal erabiltzailearentzat, gorago aipatu ditugun arrazoiengatik. Horiek horrela, atal honetan euskararako prestatu diren hiru urrats inkrementaletan oinarritutako metodoak azalduko ditugu, orotariko analizatzailea lortu ahal izateko, hau da, sendoa eta gainsorkuntzarik txikienarekin.

6.1.1 Lexiko-itzultzaile estandarra

Bosgarren kapitulu osoa eskaini diogu analizatzaile estandarri, lexiko estandarra eta erabiltzaileak gehi dezakeena darabilena. Alegia, datu-base lexikalean landuta dagoenean oinarritzen dena da, EDBLn euskararen kasuan, 80.000 sarreratik gora eta

120 morfotaktika-ezaugarri dituen. Horrez gain, azaldutako erregela morfofonologikoak ditu, azala eta lexikoaren arteko aldaketak kontrolatu ahal izateko.

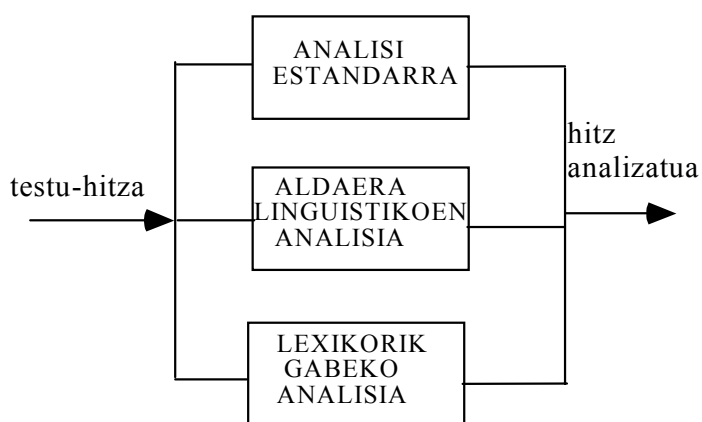
6.1.2 Itzultzaile aberastua

Aldaera linguistikoen tratamenduaz arduratzen da, hain zuzen ere, estandarrean lekuri ez dutenak detektatu eta normalizatu ahal izateko: aldaera dialektalak zein ezezagutzak eragindako akatsak. Aldaera hauek ez ditu itzultzaile estandarrak ezagutzen, baina testu errealean agertu ohi dira eta irtenbidea eman behar zaie. Beraz, sarrera edo morfema ez-estandarrek estandarrei egiten diete erreferentzia (1.500 inguru bildu dira euskaraz). Are gehiago, erregela bereziak ere sortu dira, nahiz hauek fonetikoak diren gehienetan eta, beraz, azaleko mailatik gertuago aplikatzen diren morfofonologikoak baino. Urrats honek estaldura handitu egiten du, baina gainsorkuntzak ere gora egiten du.

6.1.3 Itzultzaile orokorra

Lexikoan lemarik ez duen analizatzaile orokorra da, hizki ez-askeak lexikoan mantentzen dituen. Lexiko-sarreraren ordezkaria kategoriaz nagusiak edo irekiak (izenak, adjektiboak, aditzak, etab.) erabiltzen dira, maiztasunik handienekoak eta sistematikoenak direlako, besteak beste. Karaktereen konbinazio guztiak biltzen dituen automata orokor batek detektatzen ditu sarrera izan daitezkeenak, morfotaktika estandarra eta erregelak erabiliz. *Stemmer* batek bezala jotzen du.

Beraz, hiru urratsak egin daitezke forma bat analizatu ahal izateko, baina urrats batean analisisia lortuz gero, beste urratsetako analisisiak baztertuko dira (ikus 11. irudia).



11. irudia.- Hiru urratsetan oinarritutako analisisa.

Arkitektura honen ezaugarri nagusiak homogeneousuna, modularitatea eta berrerabilgarri izatea dira, hiru urratsak lexiko-itzultzaileetan oinarrituak baitira eta hainbat tresnatan aplikatu daitezke gainera.

Lehenengo urratsak, estandarrak, huts egiten duenean bakarrik saiateren da bigarrenarekin, eta bigarrenak erantzuten ez duenean soilik hirugarrenarekin. Oflazer-ek ere (Oflazer, 1996) halako laxatze-teknika darabiltza, nahiz modu desberdinean egiten duen. Sintaxian antzeko zerbait erabiltzen da, laxatze-murriztapen moduan.

Aurreko kapituluaren analisi estandarra ikusi dugunez, ondoko ataletan aldaera linguistikoaren tratamendua eta lexikorik gabeko analisisa ikusiko ditugu. Horrela, kapituluaren bukaeran, testuetan estaldura % 100ekoa dela esan ahal izango dugu eta, doitasuna, % 99koa.

6.2 Aldaeren tratamendua

Azpisistema morfologiko honek aldaera linguistikoak analizatzen, normalizatzen eta sortzen ditu. Prozesadore morfologikoan estaldura handitzeko erabiltzen da, gorago esan dugunez. Hain zuzen, lehenengo analizatzaile estandarrak probatzen du hitzak analizatzen eta hark ez badio emaitzarik eskaintzen, orduan aldaeren tratamendura jotzen du. Gehienetan aldaera dialektalak eta forma estandarren ezezagutzak eragindakoak dira, eta ez dira maiztasun urrikoak.

6.2.1 Lexikoa

Sarrera lexiko diren morfema ez-estandar hauek dagokien estandarri lotuta daude datu-basean, forma "egokira" bidali eta zuzendu edo normalizatu ahal izateko (*jardun[ADI]:ihardun, adjektibo[ADJ]:adjetibo, etab.*).

Gauza bera gertatzen da kategoria lagungarriekin ere, alegia, morfema ez-askeekin (deklinabideko aldaera dialektalak batez ere: *tik[ABL]:tikan[Eep]*).

Forma hauek azpilexikoekin lotzen dira eta, sistema aberasturako esportazioa burutzen denean, forma ez-estandarrak gehitzen zaizkie forma estandarrei.

Lexiko-itzultzaileen aukerei etekina atereaz, forma estandarra agertuko da ezkerraldean eta ez-estandarra eskuinaldean. Horrela, analisisa egiten denean, erregelak aplikatuko dira morfema estandar eta ez-estandarrekin, baina analisisa beti estandarra izango da. Forma ez-estandar bat analizatu ondoren sorkuntza egingo bagenu, sistema honekin

forma estandar eta ez-estandarrek lortuko genituzke; aldiz, sorkuntza sistema estandarren bidez egiten bada, dagokion forma estandarra bakarrik eskuratuko dugu. Ezaugarri hau erabiltzen da *Xuxen* zuzentzaile ortografikoan zuzenketak egiteko ere.

Adibidea: aldaeratik forma estandarrera

ihardunetikan sistema aberastuarekin analizaten bada, lortzen den emaitza *jardun+etik* da; beraz, sorkuntza estandarren ondorioz *jardunetik* lortuko da.

6.2.2 Erregela ez-estandarrek

Erregela ez-estandar berriak, paraleloan konbinatzen dira haien artean, baina sekuentzian estandarrekin. Erregela ez-estandarrek azaleko mailatik lexikokotik baino gertuago egongo dira, normalean aldaketa fonetikoak bakarrik direlako eta, ondorioz, ez dutelako informazio morfologikoaren beharrik. Lexiko ez-estandarrekin ez da gainsorkuntza gertatzen, morfema zehatzak erabiltzen dira-eta.

Erregela morfofonologiko estandarrek, gorago ikusi ditugunak, ia bere horretan balia ditzakegu hemen ere, aldaketa txiki bat bakarrik eginez: morfema-muga adierazten duen '+' karakterea ia maila lexikora iritsi arte mantendu, hainbat erabilera ez-estandar morfema-mugaren inguruan gertatzen baita: maileguetako *o/u* bukaera, esaterako.

Erregela ez-estandarrek, estandarretako osagarriak direnez, hautazkoak izango dira beti; beraz, norabide bakarreko hainbat erregela idatziko dira azaleko karaktereak eta morfema-muga konbinatuz. Erregela-multzo berri hau nahiko gainsortzailea da, idazten diren erregelak hautazkoak eta paraleloak izanik, haien arteko konbinaketa guztiak onartzen baitira.

Momentuz idatzi diren hamabost erregelak fonetikoak dira eta oso sinpleak testuinguruaren aldetik, salbuespen batekin, *a* itsatsiarena, honetan morfema-muga eta beste ezaugarri batzuk erabili behar izan direlako testuinguruan.

Zenbait erregela ez-estandar

Ondoren agertzen den erregela bakoitzak aldaera bat kudeatzen du. Hala ere, erregelak paraleloak direnez, hainbat aldaera duen hitz baten analisia egiteko gaitasuna ematen diote sistemari.

```
"h-ren galera"
h:0 => _ ;
    ! zuhaitz:zuaitz
    ! hau:au

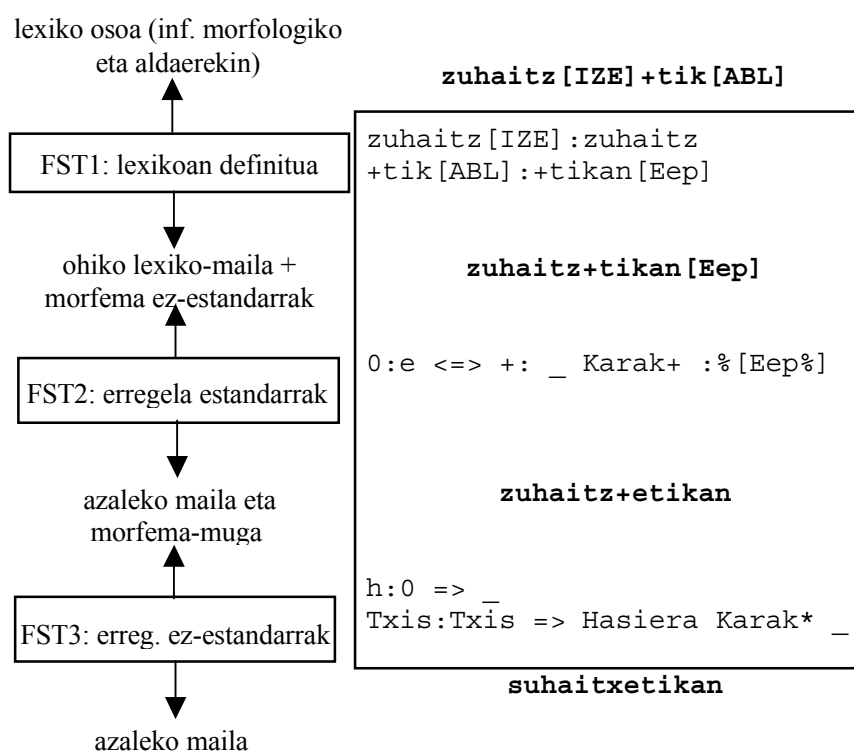
"h-ren sorrera"
0:h => [ Bokal | Hasiera ] _ Bokal ;
    ! ziur:zihur
```

```
"txistukarien artekoa"
Cx: Cy => Hasiera Karak* _ ; ! atzizkietan ez
  where Cx in (Txis)
        Cy in (Txis) ;
  ! zuhaitz: suhaitx
```

Beraz, *suhaitx* moduko hitz bat analiza daiteke, *h* baten galera, beste baten sorrera eta bi txistukariren aldaketaren konbinazioa baita. Hau interesgarria da, baina gainsortzaileegia ere izan daiteke.

6.2.3 Itzultzaileen konposaketa

Aurretik esan den bezala, hiru itzultzaile konposatu behar dira: lexikoarena, erregela estandarrena eta aldaerena. Lexikoan forma estandarrek eta ez-estandarrek daude, baina azken hauek markatuta eta dagozkien estandarrekin lotuta daude, ez-estandarrek direla jakin ahal izateko⁶⁹.



12. irudia.- Itzultzaile aberastuaren eskema

Hiru itzultzaile hauek konposatzen dira *lexc* programa erabiliz, eta, horrela, itzultzaile aberastua lortzen da. 12. irudian azaltzen da konposaketaren eskema. Bertan, behetik gorako ordenan, lau lengoia bereizten dira: azalekoa (*suhaitxetikan*), azalekoa morfema-mugekin (*zuhaitz+etikan*), lexiko-mailako forma ez-estandarrekin (*zuhaitz+tikan[Eep]*) eta lexiko-maila estandarra (*zuhaitz[IZE] + tik[ABL]*).

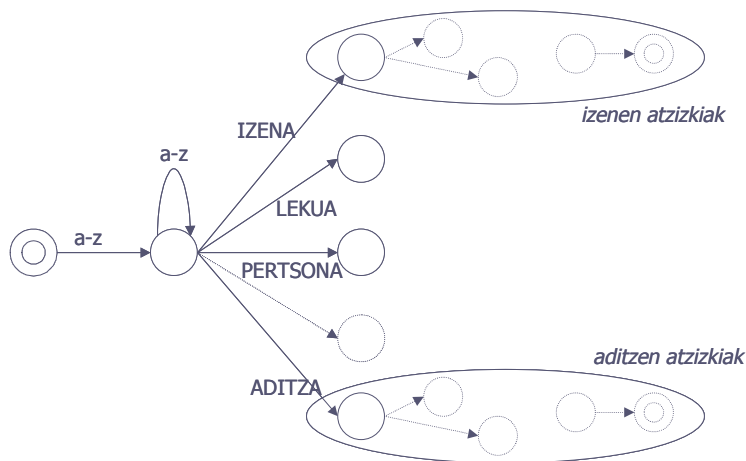
Itzultzailea gai da *zuhaitxetikan* forma ez-estandar arraroa analizatzeko, baina, ondorioz, oso gainsortzailea da, *zuhaitz+tik* morfema-sekuentziatik forma asko sor daitezke-eta: *zuhaitzetik*, *zuhaitzetikan*, *zuaitzetik*, *zuhaitxetikan*... Aztertutako hiru erregelekin gertatzen den arazo hau, hamabostekin izugarri larriagotzen da.

6.3 Lexikorik gabeko analisia

Analisi estandarrak eta aldaeren tratamenduak ez ditu hitz guztiak analizatzen, batzuetan datu-basean (hiztegian) ageri ez den lema baten flexioa baita analizatutako hitza.

Halakoetan, lexikorik gabeko analisia proposatzen da eta, horretarako, *stemmer* moduko bat eraiki beharko da lexiko-itzultzaileen teknologia erabiliz. Lortutako itzultzaileari *itzultzaile orokorra* deitzen zaio.

Lexiko txiki bat eta erregela estandarrak konposatuz lortzen da itzultzaile orokorra. Lexiko txikia kategoria irekiei (pertsona-izenak, leku-izenak, izen arruntak, adjektiboak, aditz-erroak, laburdurak eta bestelakoak) dagozkien aurrizki/atzizkietara eta lema generikoetara mugatzen da.



13. irudia.- Sistema orokorraren lexiko sinplifikatua

Lexiko txikiaren eskema sinplifikatu bat azaltzen da 13. irudian.

Beraz, lemen orokortzea egiten duen itzultzaile estandarraren azpimultzo bat da, lema eta erro guztiak, zein bestelako morfema asko, kenduta baitaude.

⁶⁹ Ondoren jaso dugun adibidea ez da kasu erreala edo gerta daitekeena: aldaerak azken muturrera eramanda ere, sistemak funtzionatzen duela erakusteko adibidea baizik.

Hala ere, analisi gehiegi⁷⁰ lortzen dira askotan eta, desanbiguazioa analizatzailearen helburutik kanpo geratzen bada ere, zenbait heuristiko⁷¹ erabili ohi dira analisi- edo lema-kopurua jaisteko. Atzizki-kopurua eta lemaren bukaera, luzera eta kategoria konbinatzen dira zenbait analisi baztertzeko. Adibidez, *bitaminiko* lema ez badago lexikoan, izen moduko analisisetan *bitamini* lema eta *ko* genitiboa (*bitamini+ko*) duena mantentzen da, baina adjektibo modukoetan, aldiz, *bitaminiko* lema mantentzen da, *iko* bukaera ohikoa baita adjektiboetan.

Dena den, modulu hau gainsortzailea da, analisi gehiegi ematen baitu forma bakoitzeko. Gainera, arazo batzuk ebatzi gabe geratzen dira. Adibidez, *Barak* analizatzean itzultzaile estandarrak *bara* maileguaren mugatu pluraltzat hartzen du, baina *Barakek* forma, aldiz, *Barak* lema ezezagunaren ergatibotzat. *Barak* lexikoan sartu gabe soluzioa zaila da, hitzez hitzeko analisisa gainditzen ez bada behintzat.

Gainsorkuntza (kasu honetan *gain-analizatze* esan beharko genuke) arrunta da estaldura eta doitasuna bildu nahi dituzten sistema morfologikoetan. Testuingurua kontuan hartzen duten tresnen lana izango da analisi horien artean desagokiak baztertzea. Baina tresna horiek txosten honen esparrutik kanpo geratzen dira.

Aipatutako hiru urratsak modu inkrementalean aplikatuz, analisi guztien emaitzan lema posible bat dago gutxienez, ez da ezer analizatu gabe geratzen (estaldura % 100 da, beraz), eta % 99an baino gehiagotan analisisa egokia da (doitasuna % 99 baino handiagoa).

⁷⁰ Kasu batzuetan 30 baino gehiago.

⁷¹ Hurbilpen moduko metodo ez-formala, akatsak egiten dituen, baina gehienetan ondo dabilena.

7 Ariketak

- 1) Pentsatu zeintzuk izango liratekeen analisi posibleak honako hitz hauetarako: *lagunen, zuen, amona*.
- 2) Gaztelaniazko “*jugar*” aditzaren erroa “*jug*” baldin bada, zer problema berezi du bere orainaldiko flexioak, alomorforik edo diakritikorik erabiltzen ez bada?
- 3) Alomorfoak eta diakritikoak erabiltzea ez da gomendagarria, baina batzuetan oso zaila da ekiditea. Azaldu halako kasuren bat euskararako.
- 4) Egin grafo bat euskarazko daten morfotaktika deskribatzeko.
- 5) 4.1 ariketa osatu honako fenomeno hauek kontuan hartzeko: genitiboaren ondoren beste atzizkiren bat ager daiteke, kasu batzuetako pluralizatzailerak (-*ta*-) morfema independente gisa adierazi behar da, eta bizidunetan deklinabide-kasu gehiago daude (-*gan*).
- 6) Interpretatu honako erregela hauek eta konpilatu automata batean:

$h:0 \Rightarrow V _ V ;$
 $y:i \Leftrightarrow V _ (0:e) s \# ;$

- 7) Norabide bakarreko erregela eskatzen duen fenomeno morfologiko bat zehaztu.
- 8) Egin erregela bat euskarazko *k:g* aldaketa kontrolatzeko. Kontuan hartu aditzez gain (*egin+ko->egingo*) toki-izenetan ere gertatzen dela (adib. *Usurbil+ko -> Usurbilgo*).
- 9) Azaldutako tresnak erabiliz eta zenbakien adibidean oinarrituz, prestatu eta probatu euskarazko daten (letraz idatzita) deskribapena/ezagutza. Zuriuneen ordeaz azpimarra sinboloa erabiliko da, sistema halako adibideak ezagutzeko gai izan dadin: *hilak_5, 2001eko_abenduaren_8an, ...* Deklinabidearen kasu arruntenak bakarrik hartu kontuan, eta ez erabili alomorforik. Eguna eta urtea zifraz bakarrik idazten direla pentsatuko da lehen bertsioan. Morfotaktikaren grafoa 4. ariketan landu da.
- 10) *lexc* eta *twolc* erabiliz, euskarazko aditz laguntzailearen orainaldiaren deskribapen morfologikoa egin. Urruneko mendekotasunak ez dira kontuan hartuko (adib. *n+a+u+t*); beraz, sistema gainsortzailea izango da. 2 atzizki bakarrik landu: *n* erlatiboa eta *la* konpletiboa, eta bi aurrizki: *ba* baldintzazkoa eta *bait* kausazkoa. Konturatu aldaketa fonologikoak gertatzen direla aurrizkiak zein atzizkiak biltzean, adib. *bait+da -> baita, dut+n -> dudan*.
- 11) Azaldutako tresnak erabiliz, prestatu euskarazko morfologiaren bi atal: izen arrunt bizigabearen eta adjektibo erregularren flexioa. Bukatutakoan pentsatu eratorpenak dakarren konplexutasuna eta arrazoitu erregela berri baten beharra.
- 12) Azaldutako tresnak erabiliz, prestatu gaztelaniazko morfologiaren bi atal: aditz jokatu erregularra eta adjektibo erregularra. Adjektiboetan txikigarria (-*ito*, -*ita*), generoa eta numeroa hartu beharko dira kontuan eta ezin da alomorforik erabili. Adib. (*raro/rarito, poca/poquita*) Aditzetan 3 jokaera erregularrenak eta bakarrik orainaldia, infinitiboa, gerundioa eta partizipioa. Hasiera batean ez barneratu aditzetan klitikoak (-*me*, -*lo*, ...). Bigarren bertsioa prestatu klitikoak kontuan hartuz. Ez hartu kontuan azentuak, ez aditzetan ez adjektiboetan.

- 13) Bibliografia kontsultatuz, aztertu nolako proposamenak egiten diren prozesadore morfologikoak modu azkarrean garatzeko. Hartu oinarritzat artikulu hau: <http://citeseer.nj.nec.com/270795.html> (*Practical Bootstrapping of Morphological Analyzer* – Oflazer eta beste, 2000).
- 14) Espresio erregularrek aukera handia eskaintzen dute morfologiaren tratamenduan. Egia esan, azaldutako lexikoak eta erregelak espresio erregularren azpimultzo bat dira. Aztertu espresio erregularren aukerak artikulu honetan oinarrituta: <http://citeseer.nj.nec.com/karttunen96regular.html> (*Regular Expressions for Language Engineering*).

8 Glosategia

alomorfoak

Bi morfema alomorfo direla esaten da informazio morfologiko bera eta forma kanoniko bera dutenean. Horrela, gaztelaniaz “jug” eta “jueg” izan daitezke sistema baten sarrera alomorfoak, biak adierazten baitute “jugar” aditzaren erroa. Ez da gomendagarria asko erabiltzea, mantentze-lana zailtzen dutelako, eta gardentasun linguistikoa galtzen, baina batzuetan haien erabilerak sistemaren garapena errazten du.

atzizkia

ik. morfema

aurrizkia

ik. morfema

baterakuntza

ik. egoera finituak

diakritikoak

Informazioa morfofonologikoa modu kriptikoan eman eta forma kanonikoa desitxuratzen duten sinboloak dira. Salbuespenen detekzioa eta erregelen aplikazioa kontrolatzea da haien helburua, baina ahal denean, lexiko-itzultzaileetan adibidez, saihestu behar dira bestelako informazio morfologikoa erabiliz.

egoera finituak

Grafoetan oinarritutako konputazio-eredu sinplea eta, ondorioz, azkarra programaren abiaduraren aldetik. Hizkuntza-teknologietan egoera finituetako eredua kontrajartzen zaio *baterakuntza-mekanismoari*, azken hau konplexuagoa eta, ondorioz, motelagoa izanik. Sintaxiaren eredu klasikoak *baterakuntzan* oinarritzen dira, egoera finituekin ezin baitira fenomeno sintaktiko guztiak adierazi. Dena den, egoera finituetan oinarritutako tresnak erabiltzen dira gaur egun tratamendu sintaktiko partziala burutzeko.

elkarketa

Elkarketa lema bat baino gehiago biltzean sortzen da eta askotan hitzaren muga gainditzen du. Ondorioz, sarri elkarketaren tratamendua ez da flexioarekin batera egiten eta morfologia konputazional klasikotik at geratzen da, lexikoan sartuz.

eratorpena

Fenomeno morfologiko ezagunena eta aztertuen *flexioa* bada ere, eratorpena ere fenomeno arrunt samarra da. *Flexioa* sintaxiak zuzenean eragina da eta ez dakar kategoria-aldaketarik. Eratorpenean, berriz, kategoria aldatzen da eta lexiko-sarrera berriak sortzeko erabili ohi da. Horrez gain, eratorpena oso irregularra da morfotaktikaren aldetik eta horregatik konplexua da konputazionalki deskribatzeko.

flexioa

ik. eratorpena

gainsorkuntza

Deskribapen morfologiko batean hizkuntzaren forma okerrak analizatzen edo sortzen direnean, deskribapena gainsortzailea dela esaten da. Oso zaila da estaldura osoa lortzea eta gainsorkuntza erabat baztertzea, batez ere eratorpenean eta elkarketan, estaldura osatzeko jeneralizazioak egin behar direlako, gainsorkuntza eraginez.

jarraitze-klasea

ik. morfotaktika

klitikoa

ik. morfema

lema

ik. morfema

morfema

Morfologian, esanahia duen unitate linguistikoa. Bi motatakoak bereizten dira: *lema* eta *erroak* batetik, hitza osatzeko ezinbestekoak direnak, lexikotan eta hiztegieta agertzen direnak, eta *hizkiak* bestetik, aurrekoen modifikatzaileak direnak. Azken hauek 3 motatakoak izan daitezke: hitzaren hasieran agertzen diren *aurrizkiak*, hitzaren bukaeran kokatu ohi diren *atzizkiak* eta hitzean zehar txertatzen diren *artizkiak*. *Morfema ez-askeak* eta *klitikoak* hizkien sinonimotzat hartzen dira.

morfofonologia

Morfemak biltzean gertatzen diren aldaketen definizioa. Adibidez, euskaraz askotan irakurtzen den erregela morfofonologiko bat hau da: $a+a = a$. Honek esan nahi duena zera da: a-z bukatutako morfema bati a-z hasitako bat lotzen zaionean, a bakarra geratzen da. Aldaketa hauen iturria morfologikoa edo fonologikoa da batzuetan, baina beste batzuetan ortografikoa ere izan daiteke.

morfotaktika

Morfemen arteko konbinazio posibleen definizioa. Behin morfemak definituta daudela zehaztu egin behar dira haien arteko konbinazio posibleak. Sistema batzuetan morfotaktikaren definizioa *paradigma* bitartez egiten da. Bi mailatako morfologian morfotaktika definitzeko jarraitze-klase bat esleitzen zaio morfema bakoitzari, jarraitze-klasea izanik ondoren joan daitekeen morfema-multzoaren identifikazioa. Adibidez, euskaraz adjektiboen jarraitze-klasean graduatzaileak egongo dira (-ago, -egi, -en) baina izenenean ez.

9 Bibliografia

Gogoratu gai honetaz ingelesezko bibliografia asko Interneten bertan dagoela. Bilatzaile arruntak erabiliz aurkitu dezakezu (adib. www.google.com), baina badago bat horretarako propio eraikita: *Citeseer* izeneko bilatzailea <http://citeseer.nj.nec.com>. Bertan idazlea, izenburua edo gako-hitzen bitartez aurkituko duzu hainbat eta hainbat artikulua. Ingelesezko liburuak, aldiz, ez dira hemengo liburu-dendetan erraz aurkitzen, nahiz unibertsitatean eskatuz edo teledenda batean erosiz lor ditzakezun (www.amazon.com, www.megadenda.com,...).

9.1 Oinarrizko bibliografia

- Alegria I. 1995. *Euskal morfologiaren tratamendu automatikorako tresnak*. Doktoretza-Tesia. UPV-EHU.
- Antworth E.L. 1990. *PC-KIMMO: A two-level processor for morphological analysis*. Occasional Publications in Academic Computing, No. 16, Dallas, Texas.
- Beesley K., Karttunen L. 2002. *Finite State Morphology: Xerox Tools and Techniques*. Cambridge University Press.
- Sproat R. 1992. *Morphology and Computation*. The MIT Press.
- Urkia M. 1997. *Euskal morfologiaren tratamendu informatikorantz*. Doktoretza-Tesia. UPV-EHU.

9.2 Bestelako bibliografia

- Abaitua J. 1994. Dependencias locales y anáforas vacías en euskara. *ASJU* (XXVIII-3), Donostia, 837-869 or.
- Aduriz I., Aldezabal I., Arriola J.M., Artola X., Urkia M. 1995. Euskararen normalizazioa eta linguistika konputazionala. *Euskera* 1994-3 (39. liburukia), Bilbo, 1579-1588. or.
- Aduriz I., Alegria I., Artola X., Ezeiza N., Sarasola K., Urkia M. 1997. A spelling corrector for Basque based on morphology. *Literary & Linguistic Computing*, Vol. 12, No. 1. Oxford University Press. Oxford.
- Aduriz I., Aldezabal I., Ansa O., Artola X., Diaz de Ilarraza A., Insausti J. M. 1998. EDBL: a Multi-Purposed Lexical Support for the Treatment of Basque. *Proceedings of the First International Conference on Language Resources and Evaluation*. Vol II. pp 821-826. Granada. May 28-30,.
- Agirre J.M. 1985. Deklinabideari buruzko oharra, *Euskeraren iker atalak* 1985-3, 60-130. or.
- Agirre E., Agirre A., Alegria I., Arregi X., Artola X., Diaz de Ilarraza A., Goenaga P., Maritxalar M., Sarasola K., Urkia M. 1991. Bi mailatako morfologiaren euskararako egokitzapena. *Elhuyar*, vol. 17, 6-14.
- Aldezabal I., Alegria I., Ansa O., Arriola J.M., Ezeiza N. 1999. Designing spelling correctors for inflected languages using lexical transducers. *Proceedings of EACL'99*, 265-266. Bergen, Norway.
- Alegria I., Artola X., Sarasola K., Urkia M. 1996. Automatic morphological analysis of Basque. *Literary & Linguistic Computing*, Vol. 11, No. 4, 193-203. Oxford University Press. Oxford.
- Alegria I., Artola X., Sarasola K. 1997. Improving a Robust Morphological Analyser using Lexical Transducers. *Recent Advances in Natural Language Processing. Current Issues in Linguistic Theory (CILT) series*. John Benjamins publisher company.. Vol. 136. pp 97-110.

- Allen J., Hunnicutt M., and Klatt D. 1987. *From text to speech: the MITalk System*. Cambridge University Press.
- Anderson S.R. 1982. Where's Morphology?, *Linguistic Inquiry* 13, 571-612. or.
1988. *Morphological theory*, Newmeyer eds., Vol. I.
1992. *A-Morphous Morphology*. Cambridge University Press, Cambridge. (1995eko edizioa).
- Anick P. eta Artemieff S. 1992. A High-level Morphological Description Language Exploiting Inflectional Paradigms, *Actes de COLING-92*, Nantes, 67-73. or.
- Antworth E.L. 1990. *PC-KIMMO: A two-level processor for morphological analysis*. Occasional Publications in Academic Computing, No. 16, Dallas, Texas.
- Aronoff M. 1976. *Word Formation in Generative Grammar*. MIT Press, Cambridge, Massachussets.
- Arregi X. eta Urkia M. 1989. *ATEF eta ROBRaren euskararako egokitzapena*. (Barne-txostena), EHU.
- Azkarate M. 1990. *Hitz elkartuak euskaraz*, Mundaiz (Euskal Filologia, 3. zk.), Deustuko Unibertsitatea, Donostia.
1996. Euskaltzaindiaren araugintza (II. Jagon Jardunaldiak, Tolosa; txosten argitaragabea).
- Azkarate M. eta Altuna P. 1992. -ko/-dun atzizkiez, *Luis Villasanteri omenaldia*, Iker-6, Donostia, 91-112.
- Azkue R.M. 1923-25. *Morfología Vasca (I, II eta III)*. (1969ko edizioa), La Gran Enciclopedia Vasca, Bilbo.
- Barton G., Berwick R., and Ristad E. 1987. *Computational Complexity and Natural Language*. MIT Press.
- Bear J. 1986. A Morphological Recogniser with Syntactic and Phonological Rules, *COLING 86*, Bonn, 272-276 or.
- Bear J. 1988. Morphology with two-level rules and negative rule features. *Proc. of 12th. COLING*, 28-31.
- Beard R. 1988. On the separation of derivation from morphology: Toward a Lexeme / Morpheme-based morphology, *Quaderni di Semantica* 9, 3-59 or.
- Bessley K., Karttunen L. 2000. Finite-State for Non-Concatenative Morphotactics. *Proc. of the ACL '2000*. Hong Kong.
- Black A., van de Plassche, J., Williams B. 1991. Analysis of Unkown Words through Morphological Descomposition. *Proc. of 5th Conference of the EACL*, vol. 1, 101-106.
- Bosque I. 1990. *Las Categorías Gramaticales*, Ed. Sintesis, Madrid, (1991ko edizioa).
- Cahill L. 1990. J. Syllable based morphology. *Proc. of 13th COLING*, vol.3, 48-53.
- Calder J. 1988. Paradigmatic morphology. *Proc. of the 4th EACL*, 58-65.
- Carroll J. eta Abaitua J. 1987. A morphological parser for Basque verbs inflection. *Conference on Artificial Intelligence*, II World Basque Congress, Donostia.
- Carter D. 1995. Rapid development of morphological descriptions for full language processing system. *Proc. of EACL '95*.
- Chanod J. P. 1994. *Finite-state composition of french verb morphology*. Xerox MLTT-005.
- De Rijk R. 1981. Euskal morfologiaren zenbait gorabehera. *Euskal Linguistika eta Literatura. Bide berriak*, Deustuko Unibertsitatea, Bilbo, 83-101 or.
1988. Basque Syntax and Universal Grammar, *Euskara Biltzarra. Congreso de la Lengua Vasca*, I. alea, Eusko Jaurlaritza, Gasteiz.
- Di Sciullo A.M. eta Williams E. 1987. *On the Definition of Word*, MIT Press, Cambridge.
- Dressler W.U. 1985. Natural Approaches to Morphology. An introduction, *Studia gramatyczne VII*, Ossolineum, 9-11. or.
- Egia G. 1985. Laburtzapena euskaraz, *Senez* 1985-3, 73-83.
- Eguzkitza A. 1991. Partikula modalez ohar pare bat, *ASJU*, XXV-3, Donostia, 961-963.
- Euskaltzaindia. 1973. *Aditz laguntzaile batua*. Euskaltzaindia, Bilbo.

- 1979a. *Euskal aditz batua*, Euskaltzaindia, Donostia.
- 1979b. Deklinabidea, *Euskera* XXIV, 633-657.
1985. *Euskal Gramatika. Lehen urratsak-I*, Euskaltzaindia, Bilbo (bigarren edizioa).
- 1987a. *Euskal Gramatika. Lehen urratsak-I (Eraskina)*, Euskaltzaindia, Bilbo.
- 1987b. *Euskal Gramatika. Lehen urratsak- II*, Euskaltzaindia, Bilbo.
- 1987c. *Hitz-elkarketa/1*, Euskaltzaindia, Bilbo.
1990. *Euskal Gramatika. Lehen urratsak-III*, Euskaltzaindia, Bilbo.
1991. *Hitz-elkarketa/3*, Euskaltzaindia, Bilbo.
- 1992a. *Hitz elkartuen osaera eta idazkera (Hitz-elkarketa/4)*, Bilbo.
1993. *Euskal Gramatika Laburra: Perpaus Bakuna*, Euskaltzaindia, Bilbo.
1994. *Euskal Gramatika: Lehen urratsak-IV*, Euskaltzaindia, Bilbo.
- 1994-. *Euskaltzaindiaren arauak*.
1999. *Euskal Gramatika: Lehen urratsak-V*, Euskaltzaindia, Bilbo.
2000. *Hitzegi Batua, Euskera XLV*.
- Goenaga P. 1980. *Gramatika bideetan*. Erein, Donostia (1985eko edizioa).
1984. *Euskal sintaxia: konplementazioa eta nominalizazioa* (doktorego-tesia). EHU, Gasteiz.
- Halle M. 1973. Prolegomena to a Theory of Word-Formation. *Linguistic Inquiry* 4, 3-16 or.
- Harris Z.S. 1946. From morpheme to utterance. *Language*-22, 142-153.
- Hopcroft J. and Ullman J. 1979. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley.
- Hurch B. 1988. Phonetics and Phonology or Phonology and Phonetics, *Z. Phon. Sprachwiss. Kommunik.forsch. (ZPSK)*, Berlin, 346-352- or.
- Hurch B. eta Oñederra L. 1987. Euskarazko fonologiaren zenbait bilakabidez, *Euskal morfosintaxia eta fonologia. Eztabaida gaiak* (P. Salaburu edit.), EHU, 7-30. or.
- Irigoien A. 1978. Deklinabideaz, *Euskera* XXIV (1978-2), 737-759. or.
1985. Euskarazko izen sintagma mugatzailerik gabekoez, *Euskera* XXX (1985-1), Bilbo, 129-139. or.
- Jacobsen W.H. 1977. The Basque Locative Suffix. *Anglo-American contributions to Basque Studies*, Reno, 163-168. or.
- Kaplan R. M. and M. Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics*, vol.20(3), 331-380.
- Karlsson F., Voutilainen A., Heikkilä J., Anttila A. 1995. *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Mouton de Gruyter.
- Karp D., Schabes Y., Zaidel M., Egedi D. 1992. A freely available wide coverage morphological analyzer for English. *Proc. of COLING'92*, Nantes, 950-954.
- Karttunen L. 1993. *Finite-State Lexicon Compiler*. Xerox ISTL-NLTT-1993-04-02.
1994. Constructing Lexical Transducers. *Proc. of COLING'94*, 406-411.
- Karttunen L., Koskeniemi K., Kaplan R. 1987. *A Compiler for Two-Level Phonological Rules in Tools for Morphological Analysis*. Center for the Study of Language and Information, Report No. CLSI-87-108.
- Karttunen L. and Beesley K.R. 1992. *Two-Level Rule Compiler*. Xerox ISTL-NLTT-1992-2.
- Kiparsky P. 1973. Elsewhere in phonology, *A Festschrift for Morris Halle* (Holt, Rinehart and Winston), Anderson S. and Kiparsky P. editors, 93-106 or.

1982. Lexical morphology and phonology, *Linguistics in the Morning Calm* Seoul (Hanshin), Yang I.-S. editor, 3-91 or.
1983. Word-formation and the lexicon, F.A. Ingeman (ed.), 3-29. or.
- Kiraz G. 1994. Multi-tape two-level morphology: a case study in semitic non-linear morphology. *Proc. of COLING'94*, 180-186.
2000. *Finite-State Morphology: Theory, Applications and Recent Developments*. Tutorial in ANLP/NAACL'2000. Seattle.
2000. *Computational Nonlinear Morphology: with emphasis on Semitic languages*. Cambridge University Press.
- Klein E. 1992. Data Types in Computational Phonology, *Actes de COLING-92*, Nantes, 149-155. or.
- Koskenniemi K. 1983. *Two-level Morphology: A general Computational Model for Word-Form Recognition and Production*. Ph.D. thesis, University of Helsinki. Publications n° 11.
1985. *Compilation of Automata from Morphological Two-level Rules*. University of Helsinki, Publication n° 15.
- Kukich K. 1992. Techniques for automatically correcting word in text. *ACM Computing Surveys*, vol.24, No. 4, 377-439.
- Kwon H-C, Karttunen L. 1994. Incremental construction of a lexical transducer for Korean. *Proc. of COLING'94*, 1262-1266.
- Lafitte P. 1944. *Grammaire Basque (Navarro-labourdin littŽraire)*. (Elkar, Donostia, 1991ko edizioa).
- Laka I. 1988. Configurational heads in inflectional morphology: the structure of the inflected forms in Basque, *ASJU XXII-2*, Donostia, 343-365. or.
- Lieber R. 1982. Allomorphy, *Linguistic Analysis* 10; 27-52. or.
- Matthews P. 1974. *Morfología. Introducción a la teoría de la estructura de la palabra*. (Paraninfo, Madrid, 1980ko edizioa).
- Mitton R. 1987. Spelling checker, spelling correctors and the misspellings of poor spellers. *Information Processing and Management*, vol 23, No.5, 495-505.
- Mitxelena K. 1961. *Fonética Histórica Vasca*, Diputación Foral de Guipúzcoa, Donostia, (1985eko edizioa).
1978. Miscelánea filológica Vasca (I), *Fontes Linguae Vasconum X*, 9, 1978, 391-397 or.
1986. Eratorbidea eta Deklinabidea, *Euskera XXXI*, 535-541. or.
- 1987-. *Orotariko Euskal Hiztegia*, Editorial Desclée de Brouwer, Ediciones Mensajero, Bilbo.
- Monachini M. eta Calzolari N. (Koord.), 1994. *Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons and Corpora. A Common Proposal and Applications to European Languages (Draft version)*, EAGLES, Pisa.
- Moreno Sandoval A. 1991. *Un modelo computacional basado en unificación para el análisis y generación de la morfología del español*. Tesis Doctoral. Universidad Autonoma de Madrid.
- Morvay K. 1988. La flexi—n vasca, *Euskara Biltzarra* (I. alea), II Euskal Mundu-Biltzarra, E.J. Argitalpen Zerbitzu Nagusia, 363-370. or.
- Mujika J.A. 1982. -(e)la eta beste konpletibatzaileak, *Euskera XXVII*, 1982-1, Bilbo, 295-327 or.
1988. Partículas modales en la flexión verbal, *ASJU, XXII*, Donostia, 463-478.
1994. Partikula modalez berriro, *ASJU, XXVIII-1*, Donostia, 333-336.
- Nida E. 1949. *Morphology. The Descriptive Analysis of Words*. University of Michigan Press (1965eko edizioa).
- Odriozola J.C. 1994. Formulazio kimikoa eta euskal deklinabidea, *Euskera 1994-3* (39. liburukia), Bilbo, 743-754. or.
- Odriozola J.C. eta Zabala I. 1994. Adjektiboen eta adberbioen arteko muga zehatzik eza, *ASJU (XXVIII-2)*, Donostia, 525-541. or..

- Oflazer K. 1994. Two-level description of Turkish morphology. *Literary and Linguistic Computing*, vol.9, No. 2, 137-148.
1996. Error-tolerant Finite State Recognition with Applications to Morphological Analysis and Spelling Correction. *Computational Linguistics*. Vol. 22, N. 1. 73-89. MIT Press.
- Oflazer K, Guzey C. 1994. Spelling Correction in Agglutinative Languages. *Proc. of ANLP-94*, Stuttgart.
- Ormazabal J. 1992. On the Structure of Complex Words: the Morphology-Syntax Interplay, *ASJU* (XXVI-3), Donostia, 725-765.
- Ortiz de Urbina J. 1986. Aspects of Basque Lexical Phonology, *Fontes Linguae Vasconum XVIII*, 1986.
- Pino M., Santalla M.P. 1996. Codificación de la anotación morfosintáctica de corpus en lenguaje SGML, *SEPLN*, 19. zk., Sevilla, 101-117. or.
- Pulman S.G., Russell G.J., Ritchie G.D. eta Black A.W. 1989. *Computational Morphology of English*, Tehnical Report No. 155, Cambridge, England.
- Rebuschi G. 1984. *Structure de l'énoncé en basque*, SELAF, Paris.
- Ritchie G. 1992. Languages generated by two-level morphological rules. *Computational Linguistics*, vol 18, No.1.
- Ritchie G., Black A.W., Russell G.J. and Pulman S.G. 1992. *Computational Morphology*. The MIT Press.
- Salaburu P. 1984. Arau fonologikoak. Hizkuntz teoria eta Baztango euskalkia: fonetika eta fonologia (I eta II), EHU, Bilbo.
- Sánchez-León F. 1997. *Análisis morfosintáctico y desambiguación en castellano*. Tesis Doctoral. Universidad Autonoma de Madrid.
- Sarasola I. 1996. *Euskal Hiztegia*, Gipuzkoa Donostia Kutxa, Donostia.
- Scalise S. 1984. *Morfología generativa*. (Alianza, Madrid, 1987ko edizioa).
- Selkirk L. 1982. *The Syntax of Words*. MIT Press, Cambridge, Massachussets..
- Shieber S.M. 1986. *An Introduction to Unification-Based Approaches to Grammar*, CSLI, Stanford.
- Spencer A. 1991. *Morphological theory (An Introduction to Word Structure in Generative Grammar)*. Cambridge University Press, Cambridge.
- Szpyra J. 1989. *The Phonology-Morphology Interface*. Routledge, Chapman & Hall Ltd., London.
- Theron P., Cloete I. 1997. Automatic Acquisition of two-level Morphological Rules. *Proc. of the ANLP'97*. 103-110. Washington.
- Trask R.L. 1985. -ko Atzizkia Euskaraz, *Euskera 30-1*, 165-173. or.
- Trost H. 1990. The application of two-level morphology to non-concatenative German morphology. *Proc. of COLING-90*, Helsinki, vol.2 371-376.
1994. Morphology with a null-interface. *Proc. of COLING'94*, 141-147.
- Tzoukermann E. and Liberman M. 1990. A finite-state morphological processor for Spanish. *Proc. of COLING-90*, Helsinki, vol.3, 277-281.
- Urkia M. 1996. Morfologia konputazionala eta euskal morfologia, *ASJU*, XXX-1, Donostia, 359-382.
- UZEI. 1982. *Hizkuntzalaritza hiztegia (I eta II)*, Elkar, Donostia.
1988. *Laburtzapenen gidaliburua. Siglak - ikurrak- laburdurak (erizpideak eta zerrendak)*, Elkar, Donostia.
- Varela S. 1990. *Fundamentos de Morfología*. Ed. Síntesis, Madrid.
- Villasante L. 1972. *La declinación del vasco literario común*, Ed. Franciscana Aránzazu, Oñati (3. edizioa).
1974. *Palabras vascas compuestas y derivadas*, Ed. Franciscana Aránzazu, Oñati.
1976. *Sintaxis de la oración compuesta*, Ed. Franciscana Aránzazu, Oñati.
1978. *Estudios de sintaxis vasca*, Ed. Franciscana Aránzazu, Oñati.

- Zabala I. 1994. Laburtzapenen deklinabidea, *Euskera* 1994-3 (39. liburukia), Bilbo, 735-742. or.
- Zajac, R. 1998. Feature Structures, Unification and Finite-State Transducers. *FSMNLP'98: International Workshop, on Finite State Methods in Natural Language Processing*.
- Zaysser L. 1996. Représentation de l'ambigüité morpho-syntaxique, *NLP+IA'96 / TAL + AI'96*, Moncton, N.B., Canada, 288-292. or.
- Zwicky A. 1990. Inflectional morphology as a (sub)component of grammar, Luschÿtzky H. (ed.), *Morphologica 1988*, Benjamins, Amsterdam.
- Zytsar J.V. 1985. Euskal deklinabidea, *Euskera XXX*, 1985-1, Bilbo, 159-163. or.

10 Aurkibide alfabetikoa

adibideak

adibide oso bat, 65
aldagaiak, 56
alfabetoa, 52
asimilazioa, 40
ber/bir aldaketa, 39
definizioak, 54
epentesia, 41
erregela ez-estandarrek, 132
erregela morfosintaktikoak (euskara), 122
erregelak, 54
erregelen eragina, 57
erregelen gatazkak, 58
ezabaketa, 41
ezaugarri morfologikoak, 45
jarraitze-klaseen definizioa, 60
lehena, 30
lexiko laburra, 61
lexiko txikia, 44
lexiko-itzultzailearen sorrera, 63
multzoak, 53
plurala gaztelaniaz, 40
r gogorra, 39
tagalog, 41
testuinguru anitz, 55
urruneko mendekotasuna, 45
zenbakien idazkera, 65

adibideak (euskara)

a organikoa, 113
a/e aldaketa, 106
adberbioa, 76, 86
adberbioaren morfotaktika, 99
aditza, 76, 85
aditzaren morfotaktika (euskara), 96
aditz-mota, 80, 91
adizki jokatueta a azalekoa, 104
adizkiak, 86
adjektiboa, 76, 85
alfabetoa, 101
analisi morfologikoa, 111
a-ren galera, 105
aspektua, 80, 91
atzizki lexikalak, 95
aurrizki lexikalak, 94
azpilexikoak, 84, 91, 92

azpisisistemaren proba, 64

definizioak, 102
deklinabidea, 88
determinatzailea, 76, 85
diakritikoak, 101, 113
e epentetikoa, 103
e/i aldaketa, 107
elipsia, 82, 94
elkarketa-marra, 82, 94
eratorpena, 81
e-ren galera, 104
erlazioa, 80
erregela taktikoak, 108, 111
esaera, 76
esklamazioa, 76
ezaugarri morfologikoak, 101
gradua, 80, 90
herskari ahostuna/ahoskabea
aldaketa, 107
hitz-elkarketa, 112
h-ren galera, 106, 113
izen arrunt bizigabeen morfotaktika, 98
izena, 75, 84
izenordaina, 75, 85
k/g aldaketa, 106
k-ren galera, 105
leku-izen bereziak, 112
lema, 86
loturazkoa, 77
marratxoa, 82, 94
marratxoaren desagertzea, 108
menderagailu askea, 77
mugatasuna/numeroa, 87
multzoak, 102
n-ren galera, 106
partikula, 77, 81
pertsona-izen bereziak, 112
r epentetikoa, 103
r-ren galera, 104
siglak, 87, 112
sorkuntza morfologikoa, 110
t/d aldaketa, 107
t-ren galera, 105

aditz-flexioa (euskara), 72

aditz-mota (euskara), 80, 91

- aldaerak, 69, 130, 131
 - erregelak, 132
 - euskara, 88
- alomorfoak, 14, 17, 27, 69, 96, 139
- aplikazioak, 14
- aspektua (euskara), 80, 91
- atzizkia, 13, 139
 - euskara, 95
- aurrizkia, 13, 139
 - euskara, 94
- azpikategorizazioa
 - euskara, 72, 74
- banaketa morfologikoa, 11
- baterakuntza, 16, 23, 116, 139
- bi mailatako morfologia, 17, 21, 72
- birdeklinabidea (euskara), 109, 117
- corpusak, 15, 109, 116
- deklinabidea (euskara), 79, 88
- desanbiguazioa, 15, 111
- diakritikoak, 23, 42, 52, 139
 - euskara, 101
- doitasuna, 68
 - euskara, 109
- EDBL, 74
- egoera finituak, 16, 23, 139
 - automatak, 36
 - itzultzaileak, 24, 36, 41
 - morfologia, 27
- elipsia (euskara), 81, 94, 109, 121
 - zero morfema, 94
- eraginkortasuna, 18
 - euskara, 116
- eransketa, 24
 - euskara, 103
- erlazioa (euskara), 80
- erregelak, 54
 - aldagaiak, 56
 - berridazketa, 27, 32
 - bi mailatako erregelak, 100
 - eraginak eta gatazkak, 57
 - euskara, 72
 - ez-estandarrak, 132
 - fonologikoak, 31
 - formatua, 34
 - konpilatzeko komandoak, 56
 - konposaketa lexikoarekin, 62
 - motak, 35
 - multzoak, 43
 - ortografikoak, 31
 - osagaiak, 34
 - paraleloak, 27
 - testuinguru anitz, 55
 - testuingurua, 35, 100
- erregelak (euskara), 100, 103
 - alfabetoa, 101
 - definizioak, 102
 - erregela morfosintaktikoak, 116
 - erregela taktikoak, 108, 112
 - ezaugarri morfologikoak, 101
 - metaerregelak, 103
 - multzoak, 102
 - ortografikoak, 107
- erro-patroia, 13
- espresio erregularrak, 24, 49
 - eragileak, 51
- estaldura, 17, 68, 69
 - euskara, 109
- etiketatze morfosintaktikoa, 15
- ezabaketa
 - euskara, 104
- ezaugarria/balioa, 74
 - euskara, 82
- fonologia, 12
 - aldaketa fonologikoak, 13
 - aldaketa morfofonologikoak, 22, 28
 - erregela fonologikoak, 31
 - morfofonologia, 140
- forma kanonikoa, 17
- gainsorkuntza, 18, 68, 139
 - euskara, 110
- generoa (euskara), 72
- gradua (euskara), 80, 90
- gramatika
 - euskara, 73, 100
- hitzaren gramatika, 13, 71, 95
 - euskara, 73
- hizketaren analisia, 15
- hizketaren sintesia, 15
- hizkuntzak
 - eranskariak, 71
 - euskara, 22, 71
 - ingelea, 22

- hizkuntzak
 - semitikoak, 22
 - suomiera, 22
 - tagalog, 41
- hiztegiak, 15
- jarraitze-klasea, 22, 28, 29, 60, 83, 95, 140
 - euskara, 73
- kategoria (euskara)
 - adberbioa, 74, 76
 - aditza, 74, 76
 - adizkiak, 78, 86, 91
 - adjektiboa, 74, 76
 - azpikategoria, 74
 - determinatzailea, 74
 - esaera, 74, 76
 - esklamazioa, 74, 76
 - hitz anitzeko unitate lexikalak, 71, 77
 - izena, 74, 75
 - izenordaina, 74, 75
 - kategoria lagungarriak, 78, 86
 - kategoria morfosintaktikoak, 78, 79, 87
 - kategoria nagusiak, 74
 - laburtzapenak, 78
 - loturazkoa, 74, 76
 - menderagailuak, 91
 - partikula, 74, 77
 - posposizioa, 74
 - siglak, 87
- klitikoa, 140
 - euskara, 78
- lema, 13, 71, 95, 140
- lematizatzailea, 15
- lexikoa, 16, 17, 27, 59
 - aldaerak, 131
 - azpilexikoak, 28, 84, 95
 - azpilexikoak (euskara), 83, 91
 - datu-basea, 68
 - karaktere anitzeko sinboloak, 61
 - konpilazioa, 62
 - konposaketa erregelekin, 62
 - lexikoan ez dauden lemak, 130, 134
 - lexikoko maila, 27
 - mantenua, 68
 - sarrerak, 28
 - sinbolo bereziak, 59
- lexiko-itzultzaileak, 33, 41, 44, 116
- konpilazioa, 63
- konposaketa, 62, 133
- orokorra, 134
- probak, 64
- lexiko-sorkuntza
 - euskara, 72
- morfema, 13, 95, 140
 - euskara, 73
 - morfema askeak (euskara), 74
 - morfema ez-askeak (euskara), 81, 87
 - morfema-muga, 87
 - morfema-zatikatzaila, 116
- MORFEUS, 72
- morfografemika, 13
- morfologia
 - analisi morfologikoa (euskara), 111
 - analizatzaile morfologikoa (euskara), 73
 - egoera finituak, 27
 - elkarketa, 12, 18, 82, 139
 - elkarketa (euskara), 72
 - eratorpena, 12, 18, 24, 139
 - eratorpena (euskara), 72, 81
 - euskara, 71
 - flexioa, 12, 15, 139
 - flexioa (euskara), 72
 - funtzioak, 12
 - zatatzaile morfologikoa (euskara), 72
- morfosintaxia, 116
- morfotaktika, 13, 16, 23, 28, 29, 84, 140
 - euskara, 73, 95
- mugatasuna/numeroa (euskara), 79, 87
- ortografia, 13
- osagaien metaketa, 116
 - euskara, 72
 - kasu-markak (euskara), 117
 - mugatasuna (euskara), 119
- paradigma, 16
 - morfologia paradigmatikoa, 24
- programak, 15
 - ATEF, 20
 - DECOMP, 19
 - lexc, 59, 100
 - LinguistX, 47, 49
 - mmorph, 46
 - PC-KIMMO, 46, 100

programak, 15

prozesadore morfologikoa, 11, 16

prozesadore morfologikoa (euskara),
72, 111

sailkapena, 24

twolc, 49, 100

sarrera lexikoak

euskara, 73

silabak, 17

sistema lexikoa

euskara, 72, 73

sorkuntza morfologikoa

euskara, 109

stemming, 11, 16, 130

trukaketa

euskara, 106

urruneko mendekotasuna, 13, 45

zuzentzaile ortografikoa, 15, 116