LEHEN PARTEA: ANALISI MORFOLOGIKOA

II. Egoera finituko morfologiaren inguruan.

Morfologiarako eredu konputazional desberdinen aurkezpena egitea eta horien barruan bi mailatako morfologiarena kokatzea eta sakontzea da bigarren kapitulu honen xede nagusia.

Formalismo morfologikoak aztertu baino lehen berauek 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 azaldu eta gero guk erabiliko dugun bi mailatako formalismoa azaltzen da zehaztasun handiagoz.

Bi mailatako morfologia izeneko formalismoaren osagaiak banan-banan aztertuz, deskribapen-ahalmenaren aldetik bere aldeko eta aurkako irizpideak zehazten dira, eta ondorioz, bere puntu ahulenaren gainean, morfotaktikarenean hain zuzen, proposatutako hobekuntzak aztertzeaz gain gure proposamenaren berri ematen da. *Jarraitze-klase hedatuak* deitu dugun mekanismo honek, jarraitze-klase arruntei leporatutako arazo guztiak konpontzen ez baditu ere, euskararako aplikazio zuzena duen funtsezko bat bideratzen du, morfemen arteko urruneko menpekotasunarena hain zuzen.

Azkenik, bi mailatako morfologiaren ereduaren konputazio-konplexutasuna aztertzen da, honek sistema errealetan erabiltzeko bideragarritasuna adierazten baitu. Gaia nahikoa polemikoa izan da eta formalismo hau euskarari egokitu izanak honetaz erakutsi diguna ere azalduko da.

Kapitulu honetan gida gisa erabili dugun liburua, R. Sproat-en Morphology and Computation (1992), gai honetan gehiago sakontzeko oso baliagarria da.

II.1 Analisi morfologikoa: sarrera gisakoa.

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

Aurretik aipatutako liburuan (Sproat, 92:17) egiten diren galderak izan daitezke kontzeptu horiek azaltzeko iturburua:

> "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 begirako morfologiaren kontzeptu garrantzitsuenak ondorioztatzen dira:

- Funtzioei begira hiru kontzeptu azaltzen dira nagusiki, *flexio-morfologia* eta *eratorpen-morfologia* eta *elkarketa*. Lehenengoa sintaxiak eskatua da, erregularra da normalean kategoriaren arabera, eta ez du funtzio sintaktikoa aldatzen. Eratorpena aldiz, sintaxiak ez du eragiten, ez da erregularra eta kategoria gramatikalaren aldaketa gerta daiteke. Elkarketa lema bat baino gehiago biltzean sortzen da eta, askotan hitzaren muga gainditzen duenez, bere tratamendua korapilatsuagoa da flexio edo eratorpenarena baino.

- Morfemen arteko loturari begira fenomeno desberdinak gerta daitezke, gure inguruko hizkuntzetako ohizko aurrizki eta atzizkiak erabiltzen dituen *kateatze sinple*tik, arabiera bezalako hizkuntzetako *erro-patroi* eredu konplexuraino. Konplexutasunaren aldetik tartean egongo liratekeen beste fenomenoak ere aipa daitezke: artizkien bidezko lotura, bikoizketa, etab.

- Loturak gertatzeko murriztapenak, askotan *morfotaktika* deritzana, inguruko morfemen funtzioa izatea da arruntena, nahiz eta hizkuntzaren arauera erregulartasun- eta hurbiltasun-gradu oso desberdinak aurkitu.

- Aldaketa fonologikoak batzuetan fonologiak zuzenean eraginda izan daitezke (suomieran adib.) eta beste batzuetan ortografiak (ingelesa adib.), horrexegatik *morfofonologikoak* deituko ditugu, bibliografian aurreko izenez gain

morfografemika ere agertu arren. Aldaketa hauen kopurua eta aldaketa gertatzeko baldintzak oso bestelakoak izan daitezke hizkuntza desberdinetan. Fenomeno honen aurrean, analisi morfologikoa egiterakoan, sistema batzuetan *alomorfoak* erabiltzen dira, hau da, morfema bera adierazteko forma bat baino gehiago erabiltzea. Aldaketa fonologiko konplexuaren adibide gisa, hizkuntza batzuetan gertatzen den bokal-armoniaren fenomenoa dugu, non puntu batean gertatzen den bokal baten aldaketak ondoko bokalen aldaketa ere eragin baitezake.

II.2 Morfologiaren eredu konputazionalak eta zenbait adibide.

Atal honetan deskribatuko dugu zein eredu erabili diren analisi zein sintesi morfologikoa ordenadorez burutu nahi izan denean. Gure helburua euskararen morfologiaren tratamendua izan denez, honen ezaugarri den morfemen kateatzeari aurre egiten dioten teknikak azaltzen dira, bestelako hizkuntzen fenomeno batzuk —hizkuntza semitikoenak adibidez— aipatzen badira ere.

II.2.1 Eredu konputazionalak: sailkapenerako irizpideak

Ingelesaren flexio-morfologia sinplearen[1] eraginaz ordenadorez egindako analisi/sintesi morfologikoari kasu handiegia ez zitzaion egiten (Winograd, 83). Programa eta ezagumendu linguistikoa nahasten zuten sistema primitiboak ziren ohizkoak orain dela urte gutxi arte. Azken urteetan aldiz, arlo honetan egindako lanak ugaritu egin dira honako arrazoiak direla medio: beste hizkuntzetarako sistema automatikoen garapena batetik, eta corpusetan oinarritutako analisirako eskaintzen duten abantaila bestetik.

Gaur egun prozesadore[2] morfologiko asko aurki daiteke bibliografian, bakoitza bere ikuspuntu eta ezaugarriekin. Beraien arteko konparaketa egin ahal izateko irizpide batzuk zehaztu behar dira aurretik. Irizpideak zehazterakoan aurreko atalean azaldutako galderetatik ere abiatuko gara honako irizpide hauek ondorioztatuz:

1) Formalismo edo ereduaren *deskribapen-ahalmena*, hau da, zein fenomeno adieraz edo analiza daitezke eredu hori erabiliz. Aztertuko ditugun adibideak, morfologikoki euskaratik urrutegi ez egotea nahi dugunez, ezaugarri honi

[1] Ingelesaren morfologia sinpletzat hartu bada ere, hau guztiz zalantzazkoa da; horrela Sproat-ek (1992:152-53) azpimarratzen du nola morfologia konplexuaren fama duten hizkuntzetan (suomiera edo turkiera, adib.) konplexutasuna luzeraren sinonimotzat hartu den, erregulartasuna eta aldaketen kasuistika kontutan hartu gabe.

[2] Analisi edota sintesi morfologikoa burutzen duen programari prozesadore morfologiko deituko diogu.

dagokionean antzekoak izango dira; flexioa, eratorpena zein hitzaren mailako elkarketa adierazteko gai izanda ere, morfemen arteko loturen konplexutasuna kateatze maila hutsean geldituko baita, erro-patroi bezalako eredu konplexuak kontuan hartu gabe. Era berean, irizpide honen barruan *analisi* eta *sorkuntza* burutzeko gaitasuna edo bietako bat bakarrik burutzekoa bereiziko dugu.

2) Morfologiari ekiteko modua. Teoria linguistikoak eraginda, eta hizkuntzaren egiturak zein sistema eraikitzeko konputazio-ikuspegiak ere, bi eredu bereizten dira:
   - lexikoan oinarritutakoak, erroa eta hizkiak[1] dira abiapuntua eta beraiek dira gainontzekoa gobernatzen dutenak.
   - paradigman oinarritutakoak, paradigma desberdinak dira sistemaren funtsa eta gainontzeko osagaiak paradigmen menpe daude (Calder, 89; Anick & Artemieff, 92). Horrela, lexikoaren osaketa egiterakoan paradigma da erabiltzen den irizpide nagusia.

   Sistema gehienak erro-hizkian oinarritzen dira, eta ondoren aztertuko ditugun adibideetan horrela suposatuko da besterik esaten ez bada behintzat.

3) Morfotaktika ebazteko modua. Aurretik aipatu den bezala morfemen arteko lotura posibleak zehazteko moduarekin dago lotuta. Morfemetan oinarritutako sistemetan bi prozesamendu-mota agertzen dira nagusiki: *egoera finituko morfotaktika* deituko duguna eta *baterakuntza-mekanismo*etan oinarritutakoak. Lehenengoetan morfemen arteko erlazioak grafo-eran ikus daitezke, korapiluneak morfemak eta arkuak onartutako kateatzeak izanik. Baterakuntza-mekanismoek sintaxian erabili ohi diren ezaugarrietan oinarritutako gramatikak aintzakotzat hartzen dituzte, eta ondorioz malguagoak dira, tratamendu morfologiko —edo morfosintaktikoa— errazten dute baina konplexuagoak dira konputazioaren ikuspuntutik. Horietan, eredu paradigmatikotik egindako hurbilketak dira askotan, objektuei zuzendutako ereduetan ohizko diren herentzia bezalako kontzeptuak erabiltzen direlarik (de Smedt, 84; Calder, 89; Anick & Artemieff, 92).

4) Aldaketa morfofonologikoak adierazteko modua. Bestelakoak badaude ere, bibliografian bi metodo gailentzen dira: orain dela urte batzuk ohizkoa zen

[1] Hizki terminoa aurrizki, atzizki eta artizkien multzotzat hartzen dugu lanean zehar.

programa bidezko metodo *ad-hoc*ak eta gaur egun oso arrakastatsu bihurtu den egoera finituko *itzultzaile*etan[1] oinarritutakoa.

5) Lexikoan gordetzen diren osagai-motak. Ohizkoa da morfemak gordetzea, sistema batzuetan erroak gordetzen ez badira ere; baina batzuetan gordetzen dena hitz-zatiak dira aldaketa morfofonologikoak adierazteko modurik ez dagoelako edo eraginkortasun-arrazoiengatik. Beste aukerak ere badira, silabekin lan egiten dutenak adibidez (Cahill, 90). Irizpide honen barruan kokatzen dira lexikoan batzuetan agertzen diren bi fenomeno:
   - alomorfoen erabilpena, hau da, morfema bera adierazteko lexikoan forma bat baino gehiago erabiltzea.
   - morfemen desitxuratzea, morfema bere forma ezagunean ez gordetzea hain zuzen; KIMMOn (Koskenniemi, 93) erabiltzen diren diakritikoak dira honen adibide.

Azken irizpidea eraginkortasunarena litzateke, dena den ez da aintzakotzat hartu irizpide-multzo honetan, bere formalizazioari garrantzia eman nahi izan diogulako eta batzuetan eraginkortasuna inplementazioaren araberakoa delako eta ez formalismoaren ezaugarri. Izan ere, beste atal batean sakonduko dugu honetaz bi mailatako morfologiaren eraginkortasuna eztabaidatzean.

Adibideak aztertzean aipaturiko irizpide horien arabera sailkatzen saiatuko gara; kasu batzuk, sailkapen ia-ia guztietan gertatzen den legez, alde batean edo bestean kokatzea oso korapilatsua bada ere.

Moreno Sandoval-ek (1991) ondoko sailkapena proposatzen du:

- hitza-paradigman oinarritutakoak
- “osagaiak eta prozesuak” motakoak (edo automatetan oinarritutakoak)
- “osagaiak eta kokapena” motakoak
- bi mailatakoa eta baterakuntza

Sailkapen hori lehenago aipatutako irizpideen arabera ere adieraz daiteke, eta gainera guk aurreko puntuetan proposatutakoa zabalagoa eta zehatzagoa delakoan gaude. Moreno Sandoval-ek sistemen arteko konparazioak egiteko beste hiru irizpide proposatzen du: aipatutako eraginkortasuna, egokitzapen formala eta gainsorkuntza.

Gainsorkuntzaren arazoa ereduaren arauera baino inplementazioaren menpe dago —hizkuntza-espezifikazioa egitean alearen tamaina da funtsezkoa— askotan, eredu batzuetan gainsorkuntza ekiditea oso zaila gerta badaiteke ere.

---

1 *Itzultzaileak*: etiketa gisa n-tuplak dituzten automatak edo arkuetan n-tuplak dituzten grafo zuzendu finitoak.

Egokitzapen formalaren inguruan berak proposatzen ditu lau eredu, formalizazio prozedurala eta erazagutzailea batetik eta inplementazio prozedurala eta erazagutzailea bestetik konbinatuz lortzen direnak hain zuzen. Sailkapen hau konparaketak egiteko erakargarria bada ere, inplementazioa formalismoari egokitzen zaion zerbait izaten da, edo izan beharko luke; eta guk nahiago izan dugu formalismoak sailkatzea eta ez inplementazioak. Azken finean, berak proposatutako inplementazio erazagutzailea formalismoak erabiltzen duen eredutik finkatuta datorren ondorioa besterik ez da. Baterakuntzan oinarritutako inplementazio erazagutzailearen alde sintaxiarekin eduki ohi duten homogenotasuna aipatzen da, baina azken urteotan indartuz joan den egoera finituko sintaxia (Karlsson *et al*., 92) erabiltzen bada, estatu finituko morfologia homogenoa da ere.

II.2.2 Adibideak

Ondoren bibliografiako zenbait prozesadore morfologiko aurkezten dugu. Egiten den aurkezpena ez da osoa, adibide adierazgarri batzuk besterik ez baitira azaltzen; hala ere sistema bakoitzarekin berarengandik gertu dauden beste batzuen bibliografia-erreferentzia ematen da. Aurpezpenean jarraitu dugun ordena kronologikoa izan da (ordezkaria hautatzeko garaian behintzat, nahiz eta aldamenean antzeko adibide berriagoak zehaztu), alde batetik kontzeptuen bilakaeraz konturatzeko, eta bestetik ezaugarrien arabera aurkeztea nahikoa konplexu suerta zatekeelako.

II.2.2.1 DECOMP

Analizatzaile hau, ondorengo bertsioak izan baditu ere, hirurogeiko hamarkadaren erdialdean garatu zen MITn, MITalk izeneko proiektuaren barruan (Allen *et al*., 87). Lehenengo analizatzaileetako bat da. Lexikoa edukitzeko tokiak zein sistemaren hedadura nahikoak bere garrantzia bazuten ere, analisia burutzeko izan zuten arrazoi nagusia ingelesez morfologia eta hizketaren artean dagoen lotura da.

DECOMPen funtsezko ezaugarriak honako hauek dira:

- Flexioa, eratorpena zein hitzaren mailako elkarketa hartzen ditu kontuan. Analisirako tresna da eta ez du sorkuntzarako aplikaziorik.
- Egoera finituko morfotaktika erabiltzen du, morfemen motetan oinarritua. Morfotaktika definitzeko erregela sinple batzuk erabiltzen dira.
- 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 kontutan hartzen sistema honen erregeletan.

- 12.000 morfemak osatzen dute lexikoa, Brown corpuseko 50.000 hitzetatik abiaturik lortu zirenak. Morfema bakoitzari kode bat egokitzen zaio —morfema-mota definitzen duena—, bere gainean erregelak nola aplika daitezkeen definitzen duena. Horietako kode batek erregela bakoitzak eragiten duen aldaketa behartu, debekatu edo aukeran utz dezake.

Morfemetan banatzeko erabiltzen den algoritmoak eskuinetik ezkerrera tratatzen du hitza, errekurtsiboa da, eta anbiguitateak ekiditeko morfotaktikari dagozkion egoera-aldaketei pisu bat esleitzen die analisi-eredu batzuk beste batzuei gailentzearren eta analisia azkartzearren. Horrela *scarcity*-ren eratorpen gisako analisia "*scarce+ity* " lortuko da eta ez "*scar+cite+y*" elkarketa. Emaitzen aldetik, analisien %95a zilegia dela diote, baina badirudi neurri hori beste moduluen lana kontutan hartuz egiten dela.

Sistema hau aspaldikoa da baina urteetan zehar hobetu dute. Oso ezaugarri interesgarriak ditu: morfotaktikaren tratamendu dotorea, desanbiguazio-mekanismoa eta eraginkortasuna. Eragozpenak ere leporatu behar zaizkio: analisirako baino balio ez izatea —bere aplikaziorako nahikoa bada ere— eta aldaketa morfofonologikoen aldetiko ahalmen eskasa —ingelesaren tratamendurako honetan nahikoa izan arren—. Azken arrazoi hauek direla eta, ez da morfologiarako eredu orokorra eta ez du jarraitzaile asko izan.

Espainierarako MARS (Meya, 87) izeneko analizatzaile morfologikoak antz handia du DECOMP sistemarekin, ezaugarri guztiak, analisia egin ahala burututako desanbiguazioa barne, pareka baitaitezke: analisirako bakarrik balio izatea, egoera finituko morfotaktika, aldaketa morfofonologikoak oso erregela sinpleen bidez —nahiz eta arlo honetan DECOMPekin 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 Analisis for Retrieval Support) datuak berreskuratzen laguntzeko sistema baten barruan erabiltzen da.

II.2.2.2 ATEF

ATEF itzulpen automatikorako ingurune baten barruan dagoen analizatzaile morfologikoa da. Ingurune hau GETA Grenobleko laborategian erabiltzen da (GETA, 82), eta 70.eko hamarkadaren bukaeran garatu izan da. Ingurune horren barruan harreman estua du ROBRA izeneko analizatzaile/sortzaile sintaktikoarekin. Hizkuntza askotarako erabilia izan da, frantsesa, alemanera, errusiera eta Asiako ekialdeko zenbait hizkuntzatarako

analizatzaileak eraiki baitira. Gure taldeak prototipo bat burutu du euskararako (Arregi & Urkia, 89).

ATEFen osagaiak honako 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 honak informazio hauek azaltzen dira: hitz-zatia, hau da, aldatzen ez den morfemaren zatirik luzeena, dagokion formatoa eta unitate lexikoa —erro amankomuna duten hitz-zatiak biltzeko erabilia— eta gainerako informazio morfologikoa.
- Formatoak: hiztegiko unitate-multzo bati dagokion informazioa biltzen duen eredua. Ohizkoa da atzizki berdinak hartzen dituzten lexikoko unitateei formato bera egokitzea.
- Gramatika (erregelak): erregelen multzoa, hiztegietan aurkitutako hitz-zatiei dagozkien formatoen arabera aktibatzen direnak eta zenbait ekintza buru daitezen eragiten dutenak. Berauetan bestelako baldintzak zehatz daitezke, ekintza garrantzitsuenak ondokoak izanik: 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 formatoen arabera aplikatzen ditu erregelak.

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

Horrez gain, beste bi eragozpen ditu sistema honek:

- Programa ezagumendu linguistikotik independente bada ere gramatikaren idazketa ez da erazagutzailea, metalengoaia agintzaile batetik gertu dagoen zerbait baizik.
- Programa ez da eskuragarria eta bere zehaztasunak ez dira ezagunak, eta gainera garaiko IBM *mainframe-tean* baino ezin zen erabili.

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

- Analisirako bakarrik balio du.
- Morfotaktika azpilexikoetan oinarritutako erregelen bidez burutzen da eta erregela hauek informazio morfologikoarekin lotutako ezaugarrien menpe jar daitezke. AM-n eratorpen-morfologia definitzeko erabiltzen da aukera hau.
- Lexikoan UD (hiztegi-unitate) izeneko hitz-zatiak gordetzen dira.
- Aldaketa morfofonologikoetarako ez dago mekanismorik.

II.2.2.3 KIMMO

Aurretik ikusitako ereduetan morfotaktikaren aldetik nahikoa ahaltsuak baziren ere aldaketa morfofonologikoetarako pobreak ziren. Horren zioa aplikazio-hizkuntzen ezaugarrietan aurki daiteke, zeren eta normalean oso flexio pobre eta aldaketa erregular gutxi duten ingelesa bezalako hizkuntzetarako egiten ziren prozesadore morfologikoak.

Koskenniemik (1983) bere tesian eredu berri bat proposatu zuen, bi mailatako morfologia deitutakoa, oso arrakastatsua gertatu dena bere ezaugarri garrantzitsuenari esker: analisi zein sintesirako aldaketa morfofonologikoak adierazteko formalismo ahaltsu, orokor eta eraginkorra[1] izatearena hain zuzen. Suomierarako gauzatu bazuen ere, berehala etorri zen KIMMO[2] izeneko ingeleserako bertsioa, Karttunen-ek (1983) eginda. Aldaketa morfofonologikoak adierazteko egoera finituko itzultzaileetan konpilatzen diren bi mailatako erregela paraleloak erabiltzen dira. Formalismo hau da euskararako oinarrizko tresnak diseinatzerakoan aukeratu duguna.

Hala ere KIMMO ez zen izan lehena aldaketa morfofonologikoak deskribatzeko erregela orokorrak diseinatzen. 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—, baina tarteko egoerekin arazoak gertatzen ziren. Ondoren *keçi* izeneko prozesadore morfologikoa egiterakoan Hankamer-ek (1986) *sortu eta egiaztatu* filosofiarekin zebiltzan erregela sekuentzialak ere proposatu zituen.

---

1 Ezaugarri hauen gainean ñabardurak egingo dira geroago.

2 KIMMO izenarekin Koskenniemik proposatutako bi mailatako morfologian oinarritutako prozesadore morfologiko guztiak izendatuko ditugu.

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

- Analisi zein sintesirako baliagarria da. Kateatze mailako fenomenoak bakarrik deskriba daitezke, baina bi mailatako ideia n mailatara zabalduz beste fenomeno konplexuagoak, hizkuntza semitikoenak adibidez, ebatz daitezke (Kay, 87) (Beesley, 90) (Kiraz, 94).

- 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 menpekotasuna deituriko fenomenoa deskribaezina bihurtuz. Beraz, mekanismoa oso sinplea da, baina batzuetan ez da nahikoa esanguratsu, eta horrexegatik proposatu dira aldaketak arlo honetan. Horrela, ondoko atalean aztertuko dugunez, Bear-ek (1986), Ritchie-ren taldeak (1987) Alvey sistemaren barruan, eta Trost-ek (1990) ezaugarri morfologikoetan oinarritutako baterakuntza-mekanismoak proposatzen dituzte honetarako, eta guk, aldiz, *jarraitze-klase hedatuak.*

- Aldaketa morfofonologikoena da aipatu den bezala gailentzen den aspektua, egoera finituko automaten ideia modu arrakastatsuz erabiltzen duelako xede honetarako.

- Lexikoan morfemak gordetzen dira eta, beharrezkoa ez bada ere, alomorfoak erabiltzea ez du baztertzen Koskenniemik. 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 hemen eredu honetaz ari diren erreferentzia garrantzitsu batzuk:

- Hobekuntzak: (Karttunen *et al*., 87), (Kay, 87), (Ritchie *et al*., 87), (Bear, 88), (Trost, 90), (Karttunen *et al*., 92), (Karttunen, 93).

- Inplementazioak (aldaketa handirik proposatu gabe): (Karttunen & Wittenburg, 83), (Karlsson, 92), (Clemenceau & Roche, 93), (Oflazer, 94), (Kim *et al*., 94), (Kiraz, 94).

- Banaketa libreko tresnak : PC-KIMMO (Antworth, 90), (Karp *et al*., 92).

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

II.2.2.4 Tzoukermann eta Liberman

Aurretik ikusitako azken bi aukerak biltzen dituen sistema baten berri ematen digute Tzoukermann-ek eta Liberman-ek (1990), analisi zein sintesirako erabil daitekeena. Sistema honetan linguistak egiten duen espezifikazioa eta programak tratatzen duena nahikoa desberdinak dira, tarteko konpilazio-prozesua dela eta. KIMMOren kasuan erregeletatik automatetara iragateko egiten den konpilazioa —eskuzkoa edo automatikoa— bazegoen, baina honek ez zituen aldatzen sistemaren ezaugarriak.

Espainierarako AT&T laborategietan egindako lan honetan, KIMMOk proposatzen zituen erregelen ideia hartzen da, baina eraginkortasunari begira lexikoarekin konpilatzen dira eta alomorfoei dagozkien hitz-zatiak sortzen dira automatikoki. Sortutako hitz-zatien gainean ez da aldaketa morfofonologikorik gertatuko eta, horrela, lexikoan egingo den bilaketa karakterez karaktere egin beharrean zatiz zati egin daiteke. Horretarako, konpiladorearen emaitza ez da izango morfema eta alomorfoari dagokien hitz-zatia bakarrik, baizik eta morfotaktika kontrolatzen duen grafoan dagokion hasiera- eta bukaera-egoera ere. Adibide gisa, bere artikuluan zehazten dituzten konpilatu ondorengo lexikoko osagai batzuk:

hasiera-egoera	**bukaera-egoera**	**hitz-zatia**	**morfema**	**informazio morfologikoa**
1	7	jug	jugar	aditza
1	8	jueg	jugar	aditza
1	9	juegu	jugar	aditza
1	10	jugu	jugar	aditza
1	300	buen	bueno	adjektiboa
1	500	mariscos[1]	mariscos	izena, mask., plur.
150	500	o		sing.,orain.,indik.,1.

[1] Izenetan generoari eta zenbakiari dagozkien flexioak ebazten ditu konpiladoreak.

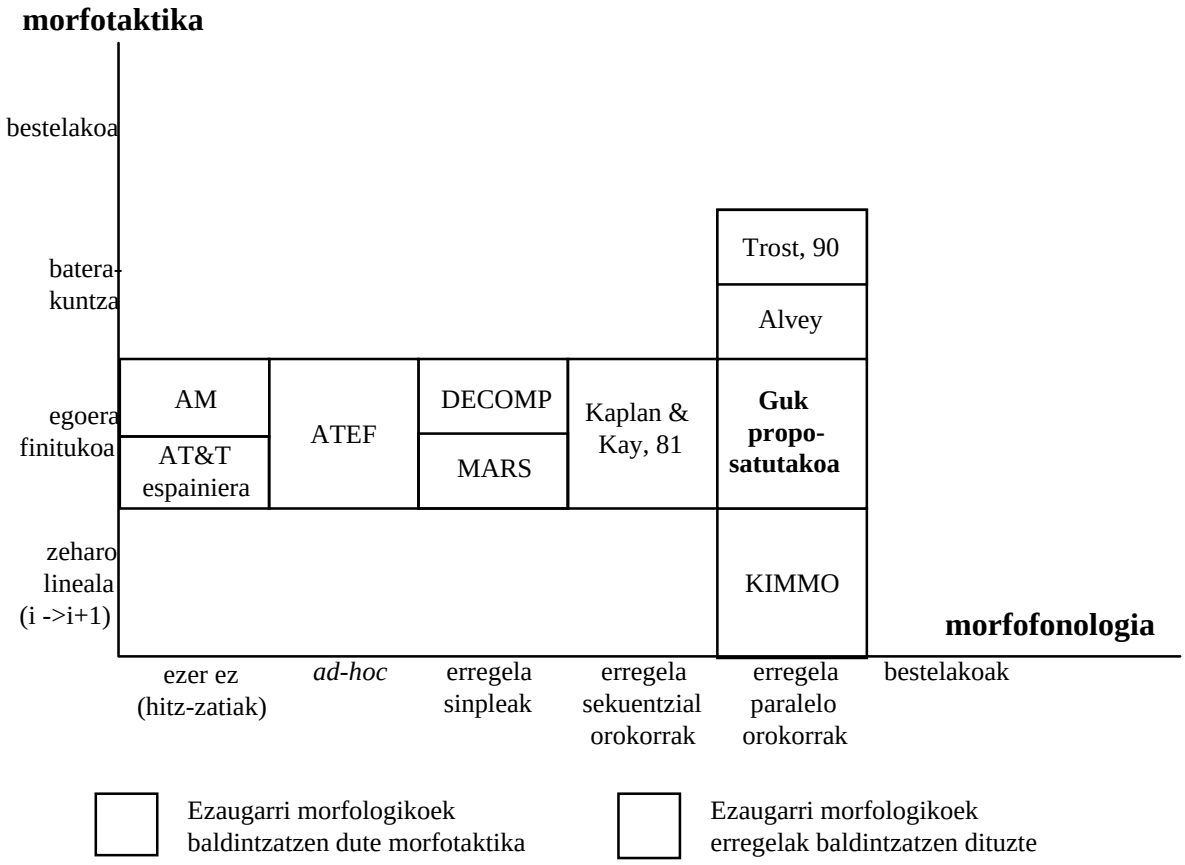
150 500 as sing.,orain.,indik.,2.

Horrela espezifikazioaren ikuspuntutik ezaugarriak KIMMO sistemarenak bezalakoak badira ere, programaren ikuspuntutik ATEF eta AM izenarekin aztertutako kasuetatik gertuago dago.

Beraiek oso interesgarri jotzen dute hau espainiera bezalako hizkuntz erromantzeetarako, eta alemanerarekin esperimentatzeko asmoa azaltzen dute. Bi mailatako morfologiaren barruan aztertuko dugun bezala, Karttunenek (1993, 1994) *lexc* izeneko konpiladorea erabiltzen duten lexiko-itzultzaileak proposatzen ditu, Koskenniemiren formalismoaren inplementazioa hobetu eta azkartzen dutenak. Karttunenen proposamen honetan grafoaren arkuetan morfemak edo hitz-zatiak egon beharrean, karaktere-bikoteak agertzen dira, bi mailatako formalismoaren ezaugarriak bere horretan mantenduz.

II.2.3 Sailkapen bat

Aurreko adibideak aztertu eta gero, II.1 irudian ikus daitekeen sailkapena ezar daiteke. Sailkapen honetan sartzeko sistemek honako ezaugarri hauek bete behar dute: morfemez osatutako lexikotan oinarriturik egotea eta kateatze-mailako fenomeno morfologikoak deskribatzea. Halako sistemetarako eskema orokorra da morfotaktika eta morfofonologia bereiztea dagoenean behintzat; eta kasu berriak aztertu ahala egunera liteke irudia. Beste idazle batzuek aipatzen duten bezala (Ritchie *et al*., 92), beste ereduekin konparaketa zehatzak egitea oso zaila gertatzen da, eta horrexegatik horiek iruditik kanpo gelditzen dira.



II.1 irudia.- Azaldutako prozesadore morfologikoen sailkapena morfotaktikaren eta morfofonologiaren tratamenduaren arabera

II.3 Bi mailatako morfologia.

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

- Eredu orokorra da, kateatze-mailako morfologian behinik behin, eta beraz, gure inguruko edozein hizkuntzari aplika dakioke.
- 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 dituzte, kontzeptualki zein konputazionalki errazago bihurtuz.

- Aurreko ezaugarriak kontutan hartuz, esan daiteke sistemaren konputazio-konplexutasuna ez dela altua, eta ondorioz, makina txikietan sistema errealak ezartzea bideratzen duela.

Ondoko pasarteetan formalismo honen osagai garrantzitsuenak diren lexiko-sistema, morfotaktika ere definitzen duena, eta erregela morfofonologikoak sakonean azaldu ondoren, sistemari egindako kritikak zerrendatuko dira, eta bukatzeko, morfotaktikari dagokionean, proposamen desberdinak eta guk egindako ekarpena azalduko dira.

II.3.1 Lexiko-sistema.

Lexiko-sistemak morfema-multzoa eta morfotaktika definitzen ditu. Hiru elementu funtsezkok osatzen dute sistema lexikoa: lexiko-sarrerak, azpilexikoak eta jarraitze-klaseak.

Lexiko-sarrera bakoitzak hiru eremu 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 hutsa egoki lekizkiekeen bitartean, hautapen-markei hutsa besterik ez zaio 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 erregulatzen dituzte jarraitze-klaseek.

- Sarrerari dagokion *informazio morfologikoa*, analisiaren emaitza gisa agertuko den informazioa alegia.

Lexikoa morfemen artean egon daitezkeen kateamenduen arabera sailkaturik dago azpilexikoen bidez. **Azpilexikoek** aurreko morfemekiko kateatzeari dagokionean ezaugarri berdineko elementu lexikoak biltzeko balio dute. Hori dela eta, morfotaktikak behartzen du azpilexikoen eraketa, linguistikoki elementu artifizial samarrak gertatuz.

Horrela, deklinabide-atzizki guztiak, adibidez, ezin dira azpilexiko berean egon, batzuk lema-motaren arabera aukeran badira.

Azpilexiko guztien definizioek egitura bera dute: identifikadorea den izena, ezaugarriak eta sarrera-multzoa. Amankomuneko informazio morfologikoa duten morfema-multzoa markatzeko erabil daitezke ezaugarriak. Horrela hitz-hasieran egon daitezkeen morfemak ezaugarri batez ezagut daitezke.

Jarraitze-klasea azpilexiko multzo bat da, morfotaktikaren aldetik unitate bat dena, eta paradigma baten osagaiekin identifika daitekeena. Identifikadore 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 bakarrak. Beraz, jarraitze-klaseak hitz batean ager daitezkeen morfemen arteko konbinaketa posibleak definitzeko mekanismoaren oinarria dira.

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

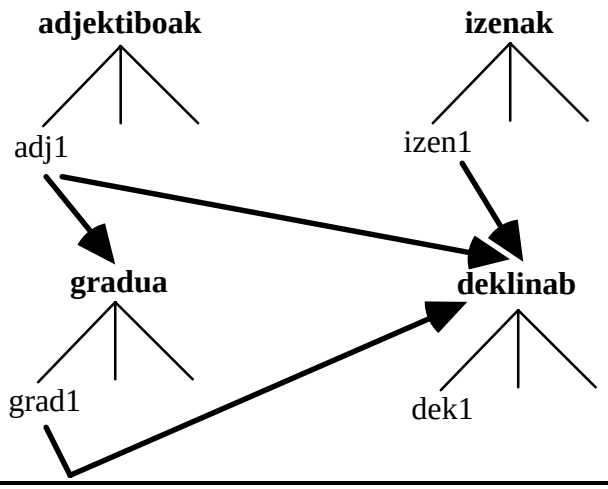
Jarraitze-klaseen eta azpilexikoen arteko bereizketa ez da funtsezkoa zeren lexiko-sistema beste modu honetaz ikus baitaiteke: estekatutako azpilexiko-multzoa. Ikuspegi honetatik morfotaktika definitzen duen grafoa ondo eratzeko elementuak besterik ez dira jarraitze-klaseak eta azpilexikoak.

Adibidez, eta sinplifikazio bat eginez, euskarazko adjektiboek eta izen arruntek deklinabide-atzizki berberak hartuko dituzte, baina lehenek baino ezin dute gradu-flexiorik hartu. II.2 irudian ikus daiteke lexiko sinplifikatu honen eraketa.

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 #[1] ...
dek2 # ...
...

JARRAITZE-KLASEA **izen_jk**
deklinab

JARRAITZE-KLASEA **adj_jk**
deklinab
gradua



II.2 irudia.- Lexiko-sistemaren erazagutzearen eta egituraren adibidea

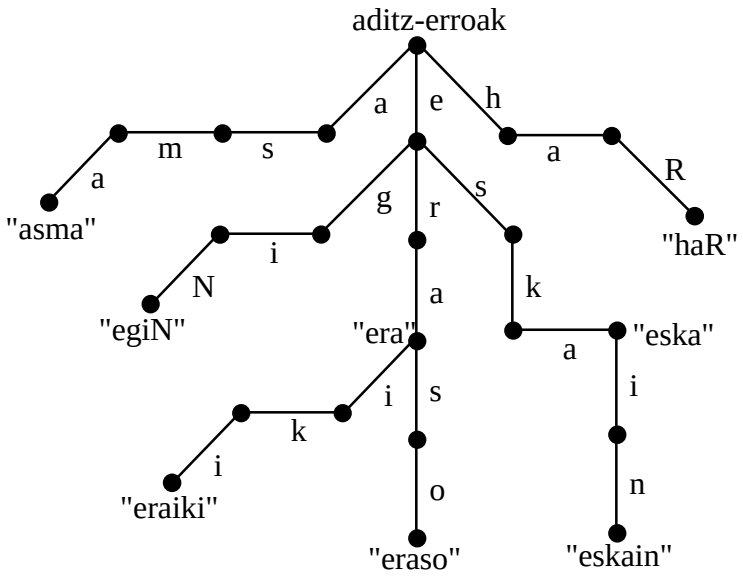
Aldaketa morfofonologikoak burutzeko erregela-multzoa dagoenez, lexikoan **alomorfo**rik definitzko beharrik ez dago. Hala ere Koskenniemik ez ditu baztertzen bi kasuta berezitan:

[1] # sinboloak jarraitze-klase hutsa adierazten du.

- azala eta lexikoaren artean aldaketa handiak edota ez-erregularrak gertatzen direnean. Horien artean daude berak erabiltzen dituen *aukera-patroiak*[1] izeneko azpilexiko txikiak.
- jarraitze-klaseak eratzeko orduan azpilexiko txiki asko sortzea lexiko-sarrerak errepikatzea baino egokiago jotzen badu ere, gehiegizko dispertsioa ekiditeko horrelakorik egin daiteke.

Aurretik aipatutako morfotaktikaren deskribapen-ahalmen apalak eragindako lexiko-sarreren bikoizketa ere bada alomorfoen iturria.

Informazio-egituraketaren aldetik lexiko-sistema azpilexiko-zuhaitz bat da eta azpilexiko bakoitza errepresentatzeko hostoetan informazio morfologikoa gordetzen duen *trie*[2] egitura erabiltzen da. ***Trie*** **egitura** grafo bat da non arku bakoitzak lexikoko karaktere bat adierazten duen eta adabegi bakoitzak horretarainoko arkuek osatutako morfemari dagozkion jarraitze-klasea eta informazio morfologikoa gorde ditzakeen. Esan bezala, morfemari dagokion jarraitze-klasea zenbait azpilexikorekin lotura gauzatzen duten arku-multzo bat bezala ikus daiteke.



II.3 irudia.- Azpilexiko sinple baten *trie* egitura

[1] *aukera-patroiak*: azpilexiko txiki hauekin zera egiten 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).

[2] *trie* terminoa re*trie*val hitzetik hartua da. Zuhaitz erako egitura bada ere ez da *tree*-rekin nahastu behar. Informazio gehiagorako ikus Knuth-en *The art of Computer Programming*, vol. 3, 481-489, 1973.

Adibidez *asma*, *egiN*[1], *era*, *eraiki*, *eraso*, *eska*, *eskain* eta *haR*[2] aditz-erroek osatutako azpilexiko baten egitura II.3 irudian ikus daiteke.

Karakterez karaktere atzitzeko eta informazioa modu trinkoan gordetzeko ahalmena aipa daitezke *trie* egituraren alde.

II.3.2 Bi mailatako erregelak.

II.3.2.1 Sarrera.

Aldaketa morfofonologikoak deskribatzeko Koskenniemik sortutako bi mailatako erregela-sistema izan zen zalantzarik gabe Koskenniemiren ekarpen handiena. Hori argi eta garbi gelditzen da bibliografian, eta idazle batzuek hori kontutan harturik eredu honen izena zalantzan jartzera iristen dira, bere ordez bi mailatako fonologia proposatuz. R. Sproat-en aipatutako liburuan (1992: 92-93) horrelako aipamena 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 Koskenniemi 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 finituko itzultzaile (FST)[3] paralelo bihurtzen dira eta karaktere-bikoteak onartuko dira baldin automata guztietan onartzen badira. Bi errepresentazioen artean, lexikokoa eta azalekoaren artean hain zuzen, ez dago tarteko egoerarik, eta hauxe da fonologia sortzailearekiko diferentzia nagusia. Beraz, hitzen analisia azaleko formari dagozkion errepresentazio lexiko onargarriak aurkitzean datza. Alderantziz gertatzen da sorkuntzan, errepresentazio lexiko ezagunetik abiatu eta berari dagozkion azaleko errepresentazioak bilatzen baitira.

Aipatu den bezala Koskenniemik 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

---

1 N karaktereak —morfofonemak, Koskenniemiren terminologiaz— gal daitekeen n karakterea adierazten du.

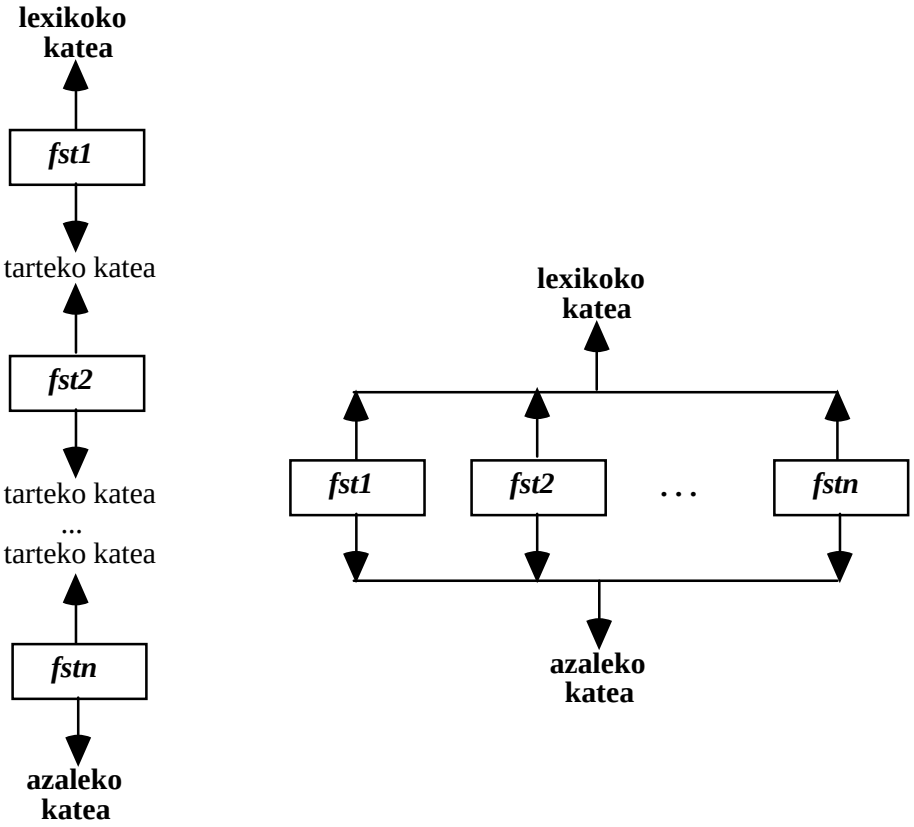
2 R karaktereak r gogorra markatzen du.

3 Egoera finituko itzultzaile (finite state transducer - FST) eta egoera finitutako automata (finite state automaton - FSA) baten artean dagoen desberdintasuna zera da: FSAren alfabetoko osagaiak sinbolo sinpleak diren bitartean FSTarenekoak bikoteak dira. Izena itzultzailea izan arren errazago ulertzen da bikoteak ezagutzeko edo ez ezagutzeko automata bezala.

egoera finituko itzultzaileetan konpilatzen ziren[1], eta ondorioz analisi zein sorkuntzarako balio du.

Aurreko eredu horiek, Kaplan eta Kay-rena, Koskenniemirenarekin konparatuz — eratorpen-fonologia eta erazagutze-fonologia izenekin bereizten ditu Karttunenek (1992)—, honako desberdintasunak zeuzkaten:

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



II.4 irudia.- Egoera finituko itzultzaile ordenatu eta paraleloen arteko konparaketa

[1] Ideia hau Jhonson-en (1972) ekarpena da.

Hala ere, Kaplan-en arabera (Kaplan, 88) (Kaplan & Kay, 94) bi ereduetan karaktere-kateen arteko erlazio erregularrak adierazten dira, erlazio hau itxia izanik 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.

II.4 irudian konputazioaren aldetik bi ereduen artean dagoen desberdintasuna azaltzen da. Azpimarratu behar da bi sistemetan erregelak egoera finituko itzultzaile bihurtu arren, hauek bi mailatako morfologian itzultzaile baino bikote-kontrolatzaileak direla.

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

II.3.2.2 Osagaiak

Esan bezala bi mailatako morfologiaren ezaugarririk garrantzitsuena lexiko-maila eta azalekoa parekatzen duten erregela-multzoan datza. Erregela hauen sintaxia zehaztu baino lehen defini dezagun beraiekin lotuta dauden zenbait kontzeptu:

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

- *azaleko alfabetoa*: analizatzeko formetan ager daitezkeen karaktere guztiak.

- *lexiko-alfabetoa*: lexikoan ager daitezkeen karaktere guztiak. Bertan azaleko karaktereez gain morfofonemak eta hautapen-markak daude.

- *karaktere-bikote konkretua*: lexikoko eta azaleko karaktere banaz osatutako bikotea. Øak ager daiteke bi mailetan, azalean morfofonema zein hautapen-markekin parekatzeko eta orokorrean azaleko karaktereren baten desagerpena edo sorrera adierazteko.

- *karaktere-multzoa*: amankomuneko ezaugarriak dituzten karaktereez osatutako multzoa, identifikadore batez ezagutzen dena. Horrela V bokalen multzoa izaten da eta C kontsonanteena. Konbentzioz `=` 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 erdiabstraktua deritze.

II.3.2.3 Erregelen formatua

Erregelen hasierako sintaxiak aldaketa batzuk izan ditu (Koskenniemi, 85; Ritchie *et al.*, 92, Karttunen, 93) deskribapen-ahalmena irabazi eta konpiladoreak inplementatu ahal izateko, eta automatetarako itzulpena eskuz egin behar ez izatea ahalbideratzen duena. Aipatutako azkena erabiliko da hemen, konpiladore hori erabil dezakegu eta.

Erregelen formatua eta osagaiak honako hauek dira:

formatua `cp op lc _ rc`

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

Eragilea (op), testuinguruaren eta korrespondentzian adierazitako bikotearen artean zer-nolako erlazioa dagoen finkatzen duena. Lau eratakoa izan daiteke: *testuinguru-murriztapena* (=>), *azalekoaren derrigortzea* (‹=), aurreko biak batera (<=>) edo *debeku-ezarpena* (/‹=). Azaleko derrigortzearena berridazketa-erregeletan erabiltzen denaren baliokidea izango litzateke. Azken mota, debekuarena hain zuzen, Bear-ek (1986) proposatu zuen salbuespenen tratamendua errazago adierazi ahal izateko.

Bakoitzaren esanahia honako hau da:

op	**adibidea**	**interpretazioa**
=>	l:a => lc_rc	*l* lexikoko karakterea azalean *a* bihurtzen da **baldin** testuingurua lc_rc bada.
<=	l:a <= lc_rc	lc_rc testuinguruan *l* **beti** gauzatzen da *a* bezala.
<=>	l:a <=> lc_rc	*l* karakterea *a* bezala gauzatzen da lc_rc testuinguruan eta ez beste inoiz.
/<=	l:a /<= lc_rc	l **ez** da **inoiz** a bezala gauzatzen lc_rc testuinguruan.

Gehien erabiltzen dena eragile konposatua da, baina aldaketa baten gauzatzea aukeran denean, testuinguruaren murriztapena ere erabili ohi da.

Testuingurua (lc_rc), korrespondentzia gertatzen deneko kasuak mugatzen dituena, aurreko eta ondorengo karaktereen arabera. _ karaktereak ezkerreko edo aurreko testuingurua (lc) eta eskuineko edo ondoko testuingurua (rc) banatzen ditu, bietako bat hutsa izan badaiteke ere.

Ezkerreko zein eskuineko testuinguruetan karaktere-bikoteen segidak adierazten dira, eta 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 erreferentziatzen dira. # sinboloak hitzaren muga adierazten du, beraz, ezkerreko testuinguruan hitz-hasiera eta eskuinekoan bukaera adieraziko du.

Testuinguruko bikoteak zenbait eragileren bidez konbina daitezke. Aukera hau urtetan zehar zabaldu da eta espresio erregularretan erabiltzen diren zenbait sinbolo hartu izan dira. Makoak **[ ]** espresioak biltzeko erabiltzen diren bitartean, parentesiek **( )** aukeran dagoena biltzeko balio dute. Hona hemen testuinguruko eragile batzuk (eragile hauek diakritiko gisa erabiltzen badira % ihes-karakterea jarriko zaie aurretik):

eragilea	**adibidea**	**interpretazioa**
kateaketa:	k:g o	k:g bikotea o:o bikoteaz jarraituta
bilketa: |	k:g | k:k	lexikoko k azaleko g edo k-rekin
errepikapena : ***** edo +	[%+:]+ [V:V]*	lexikoko + karak. behin edo gehiagotan bokala 0, 1 edo n aldiz
osagarria: \	\V	lexikoan bokala ez duen edozein bikote
kenketa: -	C-h	edozein kontsonante h izan ezik
ahaztea: /	V+ / h	bokalak h-ak kontutan hartu gabe

Posiblea da erregela bakoitzean testuinguru bat baino gehiago jartzea (**;** karakterea erabiliko da testuinguru bakoitzaren bukaeran).

Testuinguruan bikoteak zehaztea badago ere, askotan bietako bat zehaztea nahikoa da eta gainera honek erregela orokorrago bihurtzen du, gehiegi zehaztearen akatsari aurre eginez. Adibidez, testuinguru batean lexikoko k zehazteko `k:k` bikotea edo `k` zehaztea gehiegizkoa izan daiteke `k:g` bikotea ere zilegia delako; beraz, kasu horretan `k:` zehaztea da egokiena. Aldaketa fonologikoak gobernatzen dituzten erregeletako testuinguruan azaleko karaktereak izango dira nagusi, aldaketa morfologikoei dagokienetan aldiz, lexikoko karaktereak.

Adibideak hirugarren kapituluan azalduko dira. Izan ere oso komenigarria da Anworth-en liburuaren (1990) seigarren kapitulua kontsultatzea fenomeno morfologiko desberdinei dagozkien erregelen adibide zehatzak baitaude bertan.

II.3.2.4 Erregelatik automatara

Erregelatik egoera finituko automatara iragan ahal izateko konpiladoreak egon badaude baina hala ere interesgarria da eskuz nola egiten den jakitea, horrela erregela-mota desberdinen esanahia hobeto ulertuko baitugu.

l:i bikote bat bada, a:b ezkerreko testuingurua eta c:d eskuinekoa ikus ditzagun lau erregela motak eta dagozkien egoera finituko itzultzaileak. Kontutan hartu ondoko konbentzioak: bikoteetan lehen karakterea lexikokoa da eta bigarrena azalekoa, automataren 0 egoerak ezinezko bidea adierazten du, egoeraren ondoko bi puntu sinboloak bukaera-egoera eta puntu bakarrak ez-bukaerakoa. Karaktere bakarreko testuingurua suposatu dugu luzeagoa denean orokortzea oso erraza baita[1].

- testuinguru-murriztapena: **l:i => a:b _ c:d**

  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.

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

- azalekoaren derrigortzea: **l:i <= a:b _ c:d**

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

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

[1] Ezkerreko zein eskuineko testuinguru luzekoa azpiautomata bat bezala ikus daiteke eta.

- aurreko bion konposaketa: **l:i <=> a:b _ c:d**

```
    a l l 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 /<= a:b _ c:d**

  testuingurua egiaztatuz joaten da 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
```

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

Eskuzko konpilazioa lagungarria da ondo ulertzeko erregelen sintaxia, baina, oso zaila ez bada ere, erregela-kopurua eta testuinguruen konplexutasuna handitu ahala lana neketsu bihurtzen da eta akatsak aurkitu eta zuzentzeko oso zorriketa nekagarria burutu behar da. Gainera eskuzko konpilazioaren ondoren gerta liteke erregelen esanahia eta automatena bat ez etortzea. Hau guztia ekiditeko zenbait konpiladore sortu dira, haien artean Koskenniemik (Koskenniemi, 85), Ritchie-ren taldeak (Ritchie *et al.*, 92) eta Xerox-eko Karttunen eta Beesley-ek proposatutako (1992)*Twolc* . Gainera erregelen arteko gatazkak detekta daitezke eta kasu batzuetan automatikoki konpondu ere. Gure proiektuan PC-KIMMOn bezala eskuzko konpilazioa egin badugu ere, gaur egun Xeroxeko Twolc tresna dugu erabilgarri.

Pena merezi du, hala ere, halako konpiladore baten nondik-norakoak zehazteak. *Computational Morphology* (Ritchie *et al.*, 92) liburuaren 7.3 kapituluan honetaz aipatzen dena oso baliagarria da zeregin honetan. Konpiladoreak burutu behar dituen urratsak honako hauek lirateke:

- Multzoak hedatu eta aldagaiak onartzen badira hauek instantziatu.
- Eragilearen motaren arabera erregela bakoitza automata bihurtu. Ezkerreko zein eskuineko testuinguruak bi automata bihurtu ondoren, erregela-mota kontutan hartuz ondokoa egiten da:

a) testuinguru-murriztapena (=>) bada korrespondentzian zehaztutako bikotearen bidez lotzen dira aipatutako automatak,
b) azalekoaren derrigortzea (<=) bada korrespondentziarena ez den baina lexikoko karaktere bera duen edozein bikoteren bidez lotzen dira, automata berria baztertze-automata gisa birdefinituz.
c) biak batera (<=>) direnean eskuineko testuinguruaren automata bikoiztu ondoren aipatutako loturak gauzatzen dira.
d) debeku-ezarpenaren (/‹=) kasuan derrigortzerenean egindakoa errepikatzen da baina lotura korrespondentzian zehaztutako bikotearekin eginez.

- Sortutako automatak ez-determinista direnez determinista bihurtu.
- Automatak bildu eta minimizatu.

Aurreko guztia eginkizun sinpletzat har daiteke kontutan hartzen ez bada helburua ez dela lengoaia bat ezagutzen duen automata bat egitea, momentuero berrabiatu behar den bat baizik. Gainera automaten artean gatazkak sor daitezke eta konpiladoreak, honetaz konturatzeaz gain, ebatz ditzake kasu askotan (Karttunen & Beesley, 92: 22-25).

II.3.3 Programa eta exekuzio-eredua.

Koskenniemik bere tesiaren laugarren kapituluan programaren zehaztasun guztiak ematen ditu. Zehaztasun handiegitan sartu gabe, berak proposatutakotik oso gertu dagoen gure inplementaziokoari buruz hitz egitean sakontasun handiagoz azalduko baita ondoko kapituluan, azpimarra ditzagun puntu garrantzitsuenak:

- Lexikoa eta erregela-sistema modulu independenteetatik kontrolatzen dira.
- Analisian, ilararen aukera bakoitzeko, azalaren arabera lexikoko karaktere parekagarriak bilatuz doa, ondoren bikotearen bideragarritasuna testuinguruan egiaztatzen da automatak mugituz, eta bideragarriak direnean analisi-ilaran kokatzen dira. Lexikoan morfema baten bukaera aurkitzean jarraitze-klaseari dagozkion azpilexiko guztion aukerak ilararatzen dira. Beraz, *backtracking*-ean oinarritutako prozedura da. II.5 irudian algoritmoa aurkezten da.
- Sorkuntzan —hasierako Koskenniemiren proposamenean lexikoko maila osoa duen sarrera suposatzen da— lexikoa ez da erabiltzen, beraz, automaten arabera sortzen dira azaleko aukera desberdinak, ilaran kokatzen direnak eta ondoren bazter daitezkeenak. Kasu honetan morfotaktika ez da egiaztatzen.

- Morfema (edo morfema-segida) batetik abiatuta sor daitezkeen forma flexionatu zein eratorri guztiak lortzeko, analisiaren algoritmoan aldaketa txiki pare bat egin behar dira: azalak gobernatu beharrean prozesua lexikoak gobernatzen du sarrera-morfema aurkitu arte, eta ondoren analisiaren prozedura jarraitzen da baina azaleko murriztapenik gabe.

```
algoritmoa analizatu
hasiera
Ilaratu_Hasierako_Lexikoak
bitartean Ilara_Ez_Hutsa
    hasiera
    egoera = IlarakoLehena
    baldin AzalaBukatuta eta BukaerakoMorfema eta AutomatakBukaeran
        orduan IrteeraAnalisia(egoera)
    baldin JarraitzekoEgoera orduan
        hasiera
        LexLortuUmeakEtaJarraitzeLexikoak
        egin UmeBakoitzeko
            hasiera
            baldin LexKaraktereEzabatzeaPosible
                orduan AutomatakMugituEtaIlaratu
            baldin LexKaraktereEtaAzalaPosible
                orduan AutomatakMugituEtaIlaratu
            amaia
        egin JarraitzeLexikoBakoitzeko
            Ilaratu
        baldin Azaleko_elipsia orduan Ilaratu
        amaia       (* aurreko egoera tratatua *)
    amaia           (* aukera guztiak *)
amaia       (* algoritmoa *)
```

II.5 irudia.- Analizatzeko sasialgoritmoa

Hobeto ulertu ahal izateko ikus dezagun *egingo* hitza analizatzen ari denean gertatzen den kasu bat. Azaleko *egi* ezagutu izan denean aukera desberdinak daude: lexikoan, besteen artean, *egiA*, *egiN* eta *eginbehaR* agertzen dira, Aren parekatze-karaktere posibleak A:a eta A:0 eta Nrenak N:n eta N:0 izanik lau bide irekitzen dira *egiN*-en kasuan aukera bikoitza baitago. *egiN:egin* aukera izango da arrakasta izango duen bakarra gainontzekoak antzuak dira eta: *egiA:egi* aukera morfotaktikak baztertuko du aurretik

bazterturik gelditzen ez bada erregela morfologikoen bidez, *egiN:egi* aukera *N:0* aldaketa gobernatzen duen erregelak eragozten du, eta *egin:egin* aukerak ondorengo karaktereetan egingo luke porrot lexikoan bikote onargarriak ez direlako aurkitzen.

II.3.4 Sistemaren gaineko kritikak eta proposamenak.

Bi mailatako morfologiaren gainean lan handia eta publikazio asko egin da 1983an egindako lehen proposamenetik. Garapen handi honen ondorioz bi mailatako formalismoen artean "kutsu" desberdinak bereizi arren, bere funtsa, morfofonologia deskribatzeko erregelak hain zuzen, bere horretan mantentzen da. Garapen horretan, aurretik aipatutako hobekuntzez aparte —erregela-mota berria (/<=), sintaxi esanguratsuagoa eta konpiladore automatikoa—, gertatutako kritikak eta proposamenak bilduko dira ondoko multzoetan:

- deskribapen-ahalmena
- diakritikoen beharra
- morfotaktika

II.3.4.1 Deskribapen-ahalmena.

Koskenniemiren proposamenaren gaineko kritika garrantzitsuenak morfotaktikari buruz izan badira ere, badaude morfofonologiaz —morfografemika, batzuen definizioan— aritzen diren aldaketa-proposamenak. Black-ek eta zenbait lankidek (Black *et al*., 87) ondokoa kritikatzen dute erregelen gainean: bikote bakarra egotea korrespondentzian eta erregela berriak idatzi ahala gatazkak eta nahasketak sortzea. Arazo hauek ebazteko beste erregela-eredu bat proposatzen da 1993an zehazten dena (Pullman and Hepple, 93). Eredu berri honen abantaila bakarra fenomeno morfologiko konplexuak adierazteko malgutasuna da, baina gure kasuan ez dugu egokitzat jo gutxitan agertu baitzaizkigu lehen artikuluan aipatutako arazoak.

Ritchie-ren (Ritchie *et al*., 92:181) taldeak ezaugarrietan oinarritutako erregelak erabiltzeko komenientziaz hitz egiten du baina ikerketa-lan handiagoaren beharra ikusten du horretarako. Aipaturiko lan hauetan oinarriturik Carter-ek (1995) *Core Language Engine* (Alshawi, 92) delakoarentzat egindako ekarpen berrian, beste berrikuntzez gain, erregelen formatoaren aldaketa proposatzen du, bertan ezaugarrien erabilera proposatuz. Hala ere erregeletako korrespondentzian karaktere bat baino gehiago adierazteko aukera eman arren, karaktere bakar batera murriztea gomendatzen du.

Beste aldetik, kateatze-morfologia ebazteko diseinatua izan zen hasiera batean bi mailatako morfologia. Hala ere formalismo honetan oinarrituta erantzuna eman nahi izan zaio kateatze-morfologiatik kanpo geratzen diren zenbait arazori: artizkiak, bikoiztea eta hizkuntza semitikoen *erro-patroi* motako fenomenoak. Euskararen kasuan aurkitzen ez badira ere oso labur aipatuko ditugu lerro honetan egindako lanak.

Lehen kasurako Antworth-ek (1990:156) tagalogeraz erabiltzen den *in* artizkiaren arazoa ebazteko —lehen kontsonantearen atzean kokatzen dena— ondokoa proposatzen du:

- marka (X) batez ezagutzen da artizki hau lexikoan, eta aurrizki bezala adierazten da morfotaktikaren aldetik.
- erregela baten bidez 0:i eta 0:n (in-en sorrera) bikoteak onartzen dira ezkerreko testuinguruan X marka badago.

Beste kasuetan halako erregela(k) asmatzea zailagoa da ez-naturalak baitira, eta gainera konputazioaren ikuspuntutik, erregela hauek konplexutasuna igotzen dute.

Bikoiztearen arazoan PC-KIMMOren egileak (157-159 orr.) antzeko zerbait proposatzen du tagalogeraz bikoizten den lehen kontsonante-bokal silabaren bikoizketaren kasuan. Bi morfofonemaren bidez adierazten da edozein bikoizte lexikoan eta erregela pare batez kontrolatzen da morfofonema hauen gauzatzea azaleko mailan. Hala eta guztiz ere adibide honetan sinplea dena beste kasuetan askoz konplexuagoa gerta daiteke. Adibidez, walpirieraren bikoizteetarako hamalau mila egoera inguruko automata bat aurrikusten du Sproat-ek.

Hizkuntza semitikoetarako ere zenbait proposamen egin dira bi mailatako morfologian oinarriturik. Inportanteenak izan dira Kay-k 1987an proposatutako 4 mailatako eredua eta 1990ean Beesley-k egindako “desbiderapena”, lexiko anitzen arteko bi mailatako prozesua aplikatzen duena. Kay-ren proposamenean lau maila edo zinta proposatzen dira: sarrerakoa edo azalekoa lehena, kontsonante-erroena bigarrena, patroiena hirugarrena eta bokal-morfemena laugarrena. Egoera finituko itzultzaileak bi zintatatik irakurri beharrean lautatik irakurtzen du, baina zintaren batean ez aurreratzea adieraz daiteke kode batez. Teoria honetan oinarriturik eta lehentxeago aipatutako Pullman-en erregela-eredu berria erabiliz Kiraz-ek (1994) inplementazio bat proposatzen du. Beesley-renean aldiz, ohiko 2 mailak erabiltzen dira baina bi lexiko bereizten dira erroena eta bokalekin konpilatzen den patroiena, biak *trie* egiturarekin. Patroienak agintzen du baina kontsonanteei dagokien hutsuneak aurkitzean lexiko-trukea gertatzen da, horrela bi lexikoak ireki eta ixten

direlarik[1] . Bokalizazioak sortzeko lexikoko bokalen desagerpena onartzen da erregelen bidez, horrela bokalizazio partzialak ere onartzen direla.

II.3.4.2 Hautapen-markak edo diakritikoak.

Lexikoko karaktere hauek erregelen aplikazioa kontrolatzeko erabiltzen dira. Bide eskas eta nahasgarritzat hartu izan den honek abantaila bat dakar: erregelen sintaxia oso sinplea da bere osagai bakarrak karaktere-bikoteak eta eragileak baitira.

Horren truke lexikoan agertzen diren osagai ez-naturalak dira karaktere hauek, hizkuntzalariaren lana narrasten dutenak eta datuen ulergarritasuna galarazi. Honen aurrean zenbait proposamen agertzen dira bibliografian: Bear-ena eta Trost-ena.

Bietan ideia nagusia bera da, lexikoko elementuek dituzten ezaugarri morfologikoen bidez erregelen aplikazioa baldintzatzea; baina lehen proposamenean (Bear, 1988) erregelak anulatzeko ezaugarri berezi bat eransten den bitartean, bigarrena ahalmentsuagoa da (Trost, 91), gainontzeko ezaugarri morfologikoek baldintza baitezakete erregelen aplikazioa.

Gure proiektuan markak mantendu ditugu bi arrazoi nagusirengatik: batetik morfotaktika egoera finituko mekanismoen bidez burutzen delako —kontuan hartu behar baita aurreko bietan morfotaktika burutzeko ezaugarri morfologikoetan oinarritutako baterakuntza-mekanismoak proposatzen direla—, eta bestetik eskuragarri dauden tresnetan (PC-KIMMO, Alvey, Twolc, etab.) halakorik ezin delako erabili.

II.3.4.3 Morfotaktika: jarraitze-klaseak vs. baterakuntza-mekanismoak.

Koskenniemik proposatutako formalismoari aldaketa morfofonologikoen trataeratik datorkio arrakasta eta, horregatik, hasierako izena hori izan gabe, idazle askok *bi mailatako fonologia*ren izena egokitu diote. Horren ondoan, proposatutako morfotaktika —hitzaren gramatika edo sintaxia beste batzuen terminologian— oso pobrea da, zeharo lineala baita, morfema bakoitzean ondoren etor daitezkeenen multzoa zehaztea izanik morfotaktika adierazteko aukera bakarra. Bide honi jarraitu diote zenbait inplementazio eta tresna: Karttunen-en inplementazioa (1983), PC-KIMMO (Antworth, 1989), Xerox-eko Lexc (Karttunen, 1993).

Morfotaktikaren aldetik aipatutako pobrezia horrek duen arazo nagusia *urruneko menpekotasun*ari dagokiona da, hau da, morfema baten agerpenak ondo-ondokoa ez den beste morfemen agerpena baldintzatzen dueneko kasua. Adibidez, ingelesez *en*, *joy* eta

---

[1] Gogoratu behar da ideia hau ATEF izeneko prozesatzailean agertzen zela.

able morfemak zilegiak dira eta hirurak jarraian joan badaitezke ere, azken biak bakarrik ezin dira lotu. Alegia, *en*-ek *able*-en agerpena baldintzatzen du, edo beste era batean esanda, *en* eta *able*-ren artean urruneko menpekotasuna dago.

Eredua aldatu gabe arazo honen ebazpena bihurria da. Bi modutan egin daiteke, erregela baten bidez edo morfotaktikaren bidez buru daiteke ondoko prozeduraretako bat aukeratuz:

- urruneko menpekotasun-erlazioa duten morfemak markatu, baldintzatzailea hautapen-marka batez (bukaeran) eta baldintzatua beste batez, eta erregela batek bigarrenaren gauzatzea kontrolatuko du ezkerreko testuinguruaren arabera.
- tartean egon daitezkeen morfemek osatzen duten azpilexikoa(k) bikoiztu, bat aurrizkia onartzeko eta bestea ez onartzeko, atzikien eraketarako bakoitzari jarraitze-klase desberdin bat emanez. Aztertutako adibidean *en* har dezaketen morfemen jarraitze-klaseei morfema baldintzatua (*able*) egokituko zaio. Bigarren irtenbide hau ez da gomendagarria kasu hoentan alomorfo asko sortu behar baita.

Beste proiektu batzuetan morfotaktikaren eredua zeharo aldatzea proposatu da, bi mailatako morfologiaren jatorrizkoa oso pobrea dela eta. Proposamen ezagunenak Bear-ek (1986), Trost-ek (1990, 1994), eta Alvey-n (Ritchie et al., 87; Ritchie et al., 92) azaldutakoak izanik. Hiruretan baterakuntza-mekanismoak proposatzen dira; Bear PATR formalismoan oinarritzen den bitartean, Ritchie-ren taldean GPSG[1] aukeratzen dute. Alvey-ren kasuan baterakuntzak morfotaktika burutzea baino helburu handiagoa du, eratorpenak sortutako kategoria aldaketa zein elkarketen eta informazio morfosintaktikoaren tratamendua bideratzen baitu[2] —aipatutako liburuaren (1992) hirugarren, laugarren eta bosgarren kapituluak oso gomendagarriak dira—. Honetaz arituko gara luzeago III.4 pasartean.

Trost-en kasuan azpimarratzekoa da alemaneraren umlaut izeneko fenomenoa ebazteko egindako lana. Carter-ek (1995) hauekin bat dator aipatu den lanean.

Baterakuntza-mekanismoen alde aipatzen diren abantailak hauek dira: potentzia, malgutasuna eta sintaxiarekin bat etortzea. Moreno Sandoval (1991) EUROTRA proiektuaren barruan, eta bi mailatako morfologiari jarraitu gabe, eredu hauen alde

---

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

2 Tratamendu honi *morfosintaktikoa* deituko diogu eta ez da harritzekoa kasu honetan halako tratamendua burutzen *hitzaren gramatika* terminoa erabiltzea eta ez morfotaktika.

agertzen da beste aldeko irizpide bat aipatuz, inplementazio erazagutzailea. Izan ere baterakuntzak aurkako irizpide bat du, eraginkortasunarena hain zuzen. Eraginkortasun-galera horrez gain aldaketa honekin morfofonologia eta morfotaktika prozesu banatu eta sekuentzial bihurtzen dira, eraginkortasunaren aldetik kaltegarria izan daitekeena, konputazio-konplexutasuna aztertzean azalduko dugun bezala.

II.3.5 Ekarpen bat: jarraitze-klase hedatuak.

II.3.5.1 Deskripzioa

Gure proiektuan morfotaktikari ekiteko garaian tarteko irtenbide bat hartu dugu honako filosofia honi jarraituz: hasierako eredua ahalik eta gutxien aldatuz urruneko menpekotasuna adierazi ahal izatea. Horretarako jarraitze-klase hedatuak proposatzen ditugu (Agirre, Alegria, *et al*. 92). Hitzaren gramatikak dotoreagoak eta ahaltsuagoak badira ere, proposatutako mekanismoaren alde arazoa ebazteko nahikoa izanik oso mekanismo sinplea dela argudia daiteke. Hala eta guztiz ere oso interesgarritzat jotzen dugu Alvey-n proposatutako bidea morfotaktikarengatik baino tratamendu sintaktikoarengatik.

Euskararen morfotaktika nahikoa sinple eta lineala izanda ere urruneko menpekotasuna izenaz deskribatu dugun fenomenoa agertzen da. Ondoko kapituluan honi berrekiteko tokia egongo bada ere, bi kasu azalduko dugu orain, proposatutako morfotaktika-sistemaren aplikazioa adibide hauekin erabiltzearren:

- Aditz jokatuaren hizkiak. Aditz hauei zenbait aurrizki eta atzizki lot dakieke. Aurrizkiak *ba* baldintzazko morfema, *ba* baieztapenekoa eta *bait* kausala dira. Atzizkien artean *la* konpletiboa eta *n* erlatiboa ditugu. Aditzak morfema bakar bat hartzen duenean ez dago arazorik baina aurrizki batzuek debekatzen dituzte beste atzizki batzuk. Horrela *ba* indarrezkoa *la*-rekin konbinatzea zilegia den bitartean, *ba* baldintzazkoa eta *bait* ezin dira harekin konbinatu.

- *Nor-nori-nork nor-nork* eta *nor-nori* motako aditz laguntzaile eta trinkoen eraketa. Aditz hauetan pertsona berari dagozkion morfema konbizazio batzuk ez dira zilegiak—singularreko zein pluraleko lehen eta bigarrenekoak—, eta erroa tartean egon daitekeenez urruneko menpekotasunaren kasuaren aurrean gaude. Horrela, *nor-nork* laguntzaileak orainaldian honako patroi honi jarraitzen dio: *nor-morfema+erroa+nork-morfema* baina *naut* —*na*(nor-1.perts-sing) + *u*(ukan

laguntz-orainaldia) + *t* (nork-1.perts-sing)— ezinezkoa da[1]. Gauza bera gertatzen da *hauk* (2.perts-sing / 2.perts-sing-mask), *haun* (2.perts-sing / 2.perts-sing-mask), *naugu* (1.perts-sing / 1.perts-sing-plu), *gaitut* (1.perts-plur /1.perts-sing) eta beste batzuekin.

Adibide hauek azaldu ondoren, jarraitze-klase hedatuak zertan dautzan aztertzeko momentua iritsi da. Izenak dioen bezala, jarraitze-klaseen hedapen bat da eta hedapen honetan deskribapen-ahalmen aldetik bi posibilitate gehiago eskaintzen dira: debekuak eta jarraitze-klaseen zuhaitzak.

Debekuak jarraitze-klaseari eransten zaizkio eta morfema honen atzetik agertu ezin duten —debekatuta daudela esango dugu— azpilexiko-multzoak zehazten dira beraietan. Debekuak bereizteko ken sinboloa erabiltzen da aurrizki gisa. Horrela lehen adibidean agertzen zen *bait* aurrizkiak lexikoan duen definizioa honako hau da:

bait (ADITZ_JOK - LA - N)[2]

Honekin adierazten dena zera da: aurrizkiaren ondoren ADITZ_JOK jarraitze-klase arruntari dagozkion azpilexikoetako morfemak etor daitezke baina atzerago etor litezkeen morfemak edo atzizkiak ezin dira LA edo N jarraitze-klaseei dagozkien lexikoetako morfemak izan. Beraz, debekuek jarraitze-klase arrunt batean urruneko murriztapenak ezartzen dituzte.

Jarraitze-klaseen zuhaitz batek zuhaitz-moduko “jarraitze-bideak” zehazten ditu parentesien bidezko espresio batean zehaztuta. Ondo-ondoko morfemari dagozkion azpilexikoak bakarrik zehaztu beharrean ondoko maila desberdinetakoak zehatz daitezke, maila bakoitza sekuentziako morfema bati dagokiolarik. Zuhaitza zerrenda bat bezala adierazten da non maila berri bat adierazteko parentesi bat irekitzen den eta maila berean jarraitze-klase arrunt bat baino gehiago bereizteko koma karakterea erabiltzen den.

Nor-nork egiturarako definizio batzuk hauek lirateke:

na (ERROA (NORK23))

ha (ERROA (NORK13))[3]

---

1 Egungo inplementazioan aditzaren formak oso-osorik daude lexikoan arrazoi praktikoak direla eta; hala ere aipatutakoa jasotzen duen eredu teorikoa landuta dago.

2 Multzo hau etiketa batez adierazten da beste edozein jarraitze-klase bezala.

3 Hau lortzeko pertsoneei dagozkien azpilexikoa beste txikiago batzuetan banatzera behartuta gaude.

Honekin adierazten dena argi da erroaren ondoren etor daitekeena pertsona batzuei dagokien nork kasua dela. Adibide honetan ondoko morfemen murriztapenerako erabili bada ere —konturatu honetarako debekuak erabil zitezkeela—, zuhaitz hauekin aukera dago ondoko morfemen esparrua zabaltzeko edo murrizteko. Horrela, ingelesezko aztertutako adibidearen kasuan, *en-joy-able* , irtenbide bat eskaintzen digu guk proposatutako hobekuntza honek. Honako hau egin beharko litzateke urruneko menpekotasunaren arazo hau ebazteko:

joy JK_ADITZ

en (EN_ADITZAK (JK_ADITZ, ABLE))

EN_ADITZAK jarraitze-klasean *joy* aditza duen azpilexikoa badago eta ABLE-n *able* atzizkia duena aipatutako arazoa konponduta dago.

II.3.5.2 Sintaxia

Jarraitze-klase hedatuen sintaxia honako hau litzateke:

```
<jarraitze-klase hedatua> ::=    <jarraitze-klasea>
                                 | <jarraitze-klaseen arbola>
                                 | <jarraitze-klase murriztua>
<jarraitze-klaseen arbola> ::=   “(“<jarraitze-klaseen zerrenda>
                                   [<jarraitze-klaseen arbola>]”)”
<jarraitze-klase murriztua> ::=  “(“<jarraitze-klasea>
                                   <jarraitze-klase ezeztatua>*”)”
<jarraitze-klaseen zerrenda> ::= <jarraitze-klasea>
                                   [“,”<jarraitze-klasea>]*
<jarraitze-klase ezeztatua> ::=  “-”<jarraitze-klasea>
```

II.3.5.3 Semantika

Semantikaren aldetik ondoko definizioak proposatzen ditugu:

Izan bedi W hitza, $W=m_1+m_2+...+m_n$, non m_i $(1 \leq i \leq n)$ morfemak diren.

- **Ohiko jarraitze-klaseekin** morfema bakoitzari jarraitze-klase bat egokitzen zaio (m_i -> jk_i) eta jarraitze-klase horrek lexiko multzo bat definitzen du ($lex(jk_i)$).

  Zera egiaztatu behar da:

$\forall i\ (1 \leq i < n : m_{i+1} \{ lex(jk_i))$
eta m_n bukaerako morfema da.

- **Jarraitze-klase hedatuekin** morfema bakoitzari jarraitze-klase hedatu bat egokitzen zaio (m_i -> jkh_i), eta jarraitze-klase hori arrunta (jk), arbola (jka) edo murriztua (jkm) izan daiteke. Izan bitez
  $jkm_i = jk_i$ - deb_{i1} - ... - deb_{ip}
  $jka_i = jk_i$ **(**azp_{i1} **(** ... **(**azp_{ip}**))**
  non azp_{ij} eta deb_{ij} jarraitze-klase arruntak diren.

  Zera egiaztatu behar da:
  $\forall i\ (1 \leq i < n : m_{i+1} \{ lex\_hed_i)$
  non
  $lex\_hed_i = (lex(aurre_i « jk_i)) - lex(deb_i)$
  a«b = baldin a•Ø a, bestela b
  $aurre_i = azp_{jk} : j+k=i$ ñ $j = max_j \mid j<i$
  $deb_i = :(deb_{jk}) : j<i$
  eta m_n bukaerako morfema da.

II.4 Bi mailatako ereduaren konputazio-konplexutasuna eta azkartzeko bideak.

Koskenniemik, bi mailatako morfologia enuntziatu zuenean, egokitu zion ezaugarrietako bat eraginkortasuna izan zen. Hasiera batean honetaz asko eztabaidatu ez bazen ere ondoren oso gai polemikoa izan da, eta horren froga ondoko lanak dira: (Barton, 86), (Barton *et al.*, 87), (Koskenniemi & Church, 88), (Sproat, 92: 3.5).

II.4.1 Eraginkortasunaren aldetiko arazoak.

Bi mailatako morfologian konplexutasun-iturri nagusia erregeletatik dator. Lexikoko zein azaleko karaktere bat beste mailako batekin baino gehiagorekin egokitu ahal izatean, analisian zein sorkuntzan bide bati baino gehiagori jarraitu beharko zaio zenbait momentutan, eta honen ondorioz *backtracking*-a ekidin ezinezkoa izanik (ikus kapitulu honetan emandako algoritmoa). Bide honetatik eta intuitiboki honako ondorio honetara heltzen gara: zenbat eta aldaketa gehiago onartu, eta zenbat eta testuinguru murriztugabeagoa izan aldaketa horietan, orduan eta eraginkortasun-galera handiago gertatuko da aztertzeko bideak gehiago dira eta. Gainera, lexikoko karaktereak desagertzea dagoenean, ezabapenak onartzen direnean alegia, bide-kopurua handitzen da

ikaragarri, posizio bakoitzean ezabatzen den edozein lexikoko karaktereren agerpena kontutan hartu behar baita.

Erregelak konplexutasun-iturri nagusia badira ere —ondoko pasartean beraiei dagokien konplexutasuna baino ez dugu aztertuko— lexikoa ere konplexutasun-iturri bada. Hona hemen lexikoarekin lotutako zenbait puntu eraginkortasunarekin zerikusia dutenak:

- Jarraitze-klase bati dagozkion lexiko anitzak. Honek zera esan nahi du: analisia egitean bide asko jorratu beharko dira, haien artean gehienak antzuak direla.
- Hasierako azpilexiko anitz egoteak aurretik aipatutako ondorio berbera dakar. Nahiz eta bitxia irudi ahal izan, hasierako lexiko anitz egotea arruntzat jo behar da, egoera finituko morfotaktika mantentzen bada gutxienez, aurrizkiek baldintza baititzakete ondoren doazen lemak.
- Analisi-prozesuan erregelek sortzen duten konplexutasuna lexikoak murritz dezake, posible diren aukeretako bat lexikoan ez dagoenean. Sorkuntzan aldiz ez dago halako murriztapenik, beraz bide guztiak jorratuko dira.

II.4.2 Konputazio-konplexutasuna zehaztuz.

Eraginkor izatearen hasierako uste hura honetan oinarritzen zen: egoera finituko makinak oso sinpleak eta azkarrak dira eta egoera- zein arku-kopuruak ez du eragin handia abiaduran, zenbait inplementaziotan frogatu ahal izan zen bezala.

Barton izan zen gai honi sakonean eta formalki ekin zion lehena eta eztabaidaren sortzailea. Bere argudiotan sakondu gabe —honetaz sakontzeko 1987an publikatutako liburua (Barton *et al.*, 87) kontsulta daiteke— bere arrazonamendua eta ondorioak ondoko puntu hauetan labur daitezke:

- Ereduaren konputazio-konplexutasuna kalkulatzeko konplexutasun ezaguneko problema baliokide bat bilatzen du, teknika honi *laburtzea* esaten zaio.
- Bi mailatako morfologia erabiliz egindako sorkuntza *Boolean satisfiability* (SAT hemendik aurrera) problemara laburgarria dela aurkitzen du.
- Ondorioz, sorkuntza NP-gogorra dela esan dezake.
- Antzeko bidetik ezagutza edo analisiaren konplexutasuna aztertzen du eta konplexutasun berekoa dela dio mugarik gabeko ezabapenak ez badira onartzen, zeren eta kasu horretan konplexutasuna handiagoa baitagokio, PSPACE-gogorra hain zuzen.

Ondorioz esan daiteke bi mailatako eredua ez du zertan eraginkorra izan behar, beraren bidez oso problema konplexuak kode baitaitezke. Barton urrutiagora doa eta desegokitzat jotzen du, ez baitu bereizten lengoaia naturalaren problema bat —morfologiarena— konplexuago diren besteetatik.

Aurrekoa ikusita, Koskenniemik eta Church-ek (1988) erantzuten diote praktikan oinarrituz eta honako ideia hauek azpimarratuz:

- Barton-en frogapena ondo dagoela baina laburtzeko aukeratutako problema dute desegokitzat.
- SAT motako problemetan denbora modu esponentzialean hazten den bitartean, hizkuntza desberdinetarako bi mailatako morfologiaren inplementazioetan lineala dela frogatutzat ematen dute.
- Aurkitutako kasu konplexuenean, urruneko murriztapenenean eta horren barruan bokalen armoniarenean, abiadura ez da esponentzialki hazten; gainera hau ez da ohiko fenomenoa.

Laburbilduz, beren ustez bi mailatako eredua egokia da, hizkuntza anitz desberdinetako morfologia modu eraginkorrean deskribatu da eta.

II.4.3 Proposatutako hobekuntzak.

Aurretik esandakoaz honako gomendio hauek luza daitezke bi mailatako prozesadore morfologiko eraginkorrak egiteko:

- Erregelen aldetik ahal den neurrian murriztapenak ezkerreko testuinguruan jartzen saiatzea aukerak lehenago murriz daitezen, eta ahalik eta ezabapen gutxien zehaztea, maiztasun handiko karaktereen ezabapena ekidituz.
- Lexikoaren eraginaz konplexutasuna haz ez dezan, lexiko bakarreko edo gutxiko jarraitze-klaseak hobestea. Izan ere azken irizpide honek alomorfoen erabilera bultzatzen du eta Koskenniemik aipatutako beste baten kontra doa, morfemak errepikatzea baino lexiko txiki anitzen erabilera bultzatzen zuen eta.

Gomendio taktiko hauez gain zenbait aldaketa proposatu dira mota honetako prozesadore morfologikoak azkartzearren; haien artean bi hauek azpimarratu daitezke[1]: lexiko anitzei aurre egiteko Bartonek proposatutako lexikoen fusioa eta Karttunen-ek proposatutako lexiko-itzultzaileak.

---

1 Kasu batzuk ezin dira azaldu arrazoi komertzialak direla eta dokumentaturik ez daudelako.

II.4.3.1 Lexikoen fusioa.

Oinarrizko ideia oso sinplea da, lexiko guztiak bakar batean biltzen dira —edo logikoen izan daitekeena, hirutan: aurrizkiak, erroak eta atzizkiak— horrela bilaketa lexiko bakar batez burutzen da, bide-kopurua murriztuz eta hiztegia trinkotuz —*trie* egitura horrela trinkoagoa baita—.

Arazo bat sortzen da ordea, morfotaktikarena. Lexikoaren egitura aldatzen ez bada behintzat, morfotaktikari buruzko informazioa —azpilexikoak eta jarraitze-klasea— zuhaitzaren hostoetan dago, beraz bide desegokiak aukera daitezke jorratzen jarraitzeko morfofonologiaren aldetiko murriztapenik ez dagoen bitartean, baita alperrik jorratu ordea, morfotaktikak bide horiek debekatzen dituenean. Hauxe bera gertatzen da morfofonologia eta morfotaktika prozedura independiente gisa inplementatzen direnean.

Beraz, abiaduraren ikuspuntutik fusioaren interesa zalantzazkoa da, alde batetik irabaz daitekeena bestetik gal daitekeelako. Espazio kontuan bere egokitasuna zalantzarik gabekoa da. Euskararekin guk egindako fusioko probetan, ondoko kapituluan zehaztuko diren datuetan oinarriturik, abiaduraren aldetik hobekuntza aipagarririk ez dela lortzen ondorioztatzen da. Bere momentuan luzatuko gara honetaz.

II.4.3.2 Lexiko-itzultzaileak.

Karttunen-en proposamena (Karttunen *et al.*, 92), (Karttunen, 93), (Karttunen, 94) harantzago doa. Berak proposatutako *lexiko-itzultzaile*ek konplexutasunaren aldetik dakarten iraultzaz gain, bestelako abantailak ere dauzkate formalismo morfologikoaren funtsaren ikuspuntutik: batetik lexikoko adierazpide arbitrarioak ekiditen dira forma kanonikoa bultzatuz eta lexiko mailan karaktere berezirik egon beharrean informazio morfologikoa egonda; bestetik tarteko adierazpideak eta erregela-multzo anitzak konbinatuz deskripzio morfologikoa erraz daiteke eta deskribapen-ahalmena handitu[1].

Horretarako forma flexionatuak forma kanonikoekin ezkontzen ditu —adibidez *better* eta *good*—, nahiz eta horretarako lexiko eta azalaren arteko distantzia handitu. Distantzi handitze honi aurre egiteko tarteko egoerak onartzen dira lexikoko eta azaleko mailen artean, Kaplan & Kay-ren ideia zaharrak berreskuratuz.

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

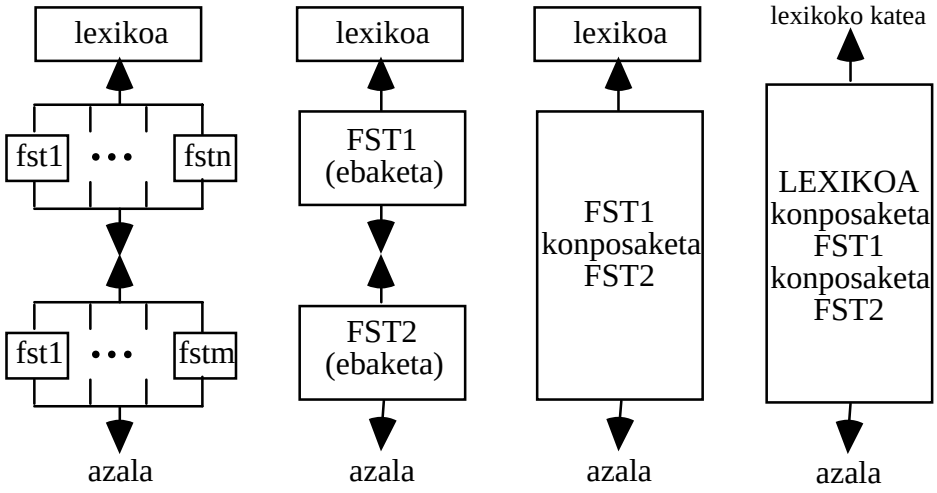
---

[1] Kaplan eta Kay-ren (1994) ideietan oinarriturik ere, Carter-ek (1995) *Core Language Engine* delakoproiekturako egindako lanean antzeko ideia proposatzen du, baina konpilazioan hizkiak sartzen baditu ere erroak kanpoan gelditzen dira sistema malguagoa izan dadin, horretarako eraginkortasuna galtzen duen arren.

- Normalean tarteko maila bat bereizten da, ohiko bi mailatako ereduan lexikokoa zena (adierazpide arbitrarioak, diakritikoak eta guzti). Lexikoko mailan informazio morfologikoa sartzen da baina ez hostoetan, arkuetan baizik.

- Ondoan dauden mailen arteko aldaketak bi mailatako morfologiari dagozkion egoera finituko itzultzaileen bidez gobernatzen dira, guztiak bakar batean konpila daitezkeenak —horretarako *twolc* konpiladorea dagoela.

- Tarteko egoeren arazoaren aurrean, sekuentzia ekartzen duena, itzultzaileen arteko konposaketa[1] proposatzen du, itzultzaile bakar bat lortuz —honetaz *lexc* konpiladorea arduratzen da.



II.6 irudia.- Itzultzaileen arteko ebaketa eta konposaketa lexiko-itzultzaile bat lortzeko (Karttunen *et al.*, 92)

- Lexikoko maila eta ondoko 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 —*be+Pres+Sg+P3+Verb : is* da horren adibide bat— bi maila horien arteko parekatzea definitzen dutenak —*lexc* konpiladorea arduratzen da bihurketa hauei dagozkien grafoen eraketaz.

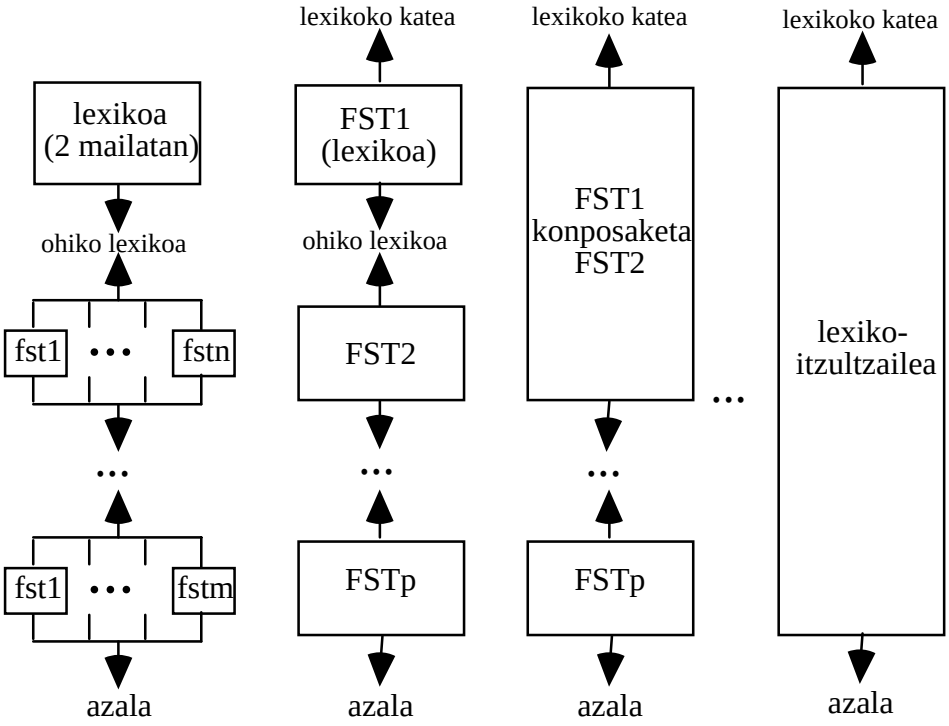
- Lexikoko mailaren eta aurretik zegoen itzultzailearen arteko konposaketa itzultzaile bakar batean sistema osoa edukiz burutzen da. Hau da funtsezkoena eraginkortasunari begira.

---

[1] Hau sinplifikazio txiki bat da. Errealitatean gertatzen dena ez da ondoko maila bakoitzeko ebakidura lehen eta ondoren konposaketa, baizik eta Karttunen-ek (1994) deitzen duen ebakitze-konposaketa (*intersecting composition*) prozesua.

Diseinu hau Kaplan-ek eta Kay-k (Kaplan, 88) (Kaplan & Kay, 94) egindako ekarpen teorikoan oinarritzen da, II.3.2.1 pasartean azaltzen dena.

Diseinu hori praktikan jartzean espero zutena baina askoz ere emaitza hobeak lortu zituzten. Erregelen ebaketa eta konposaketaren ondorioz lortzen den egoera-kopuruaren magnitude-ordena lau edo bost zifra hamartarrekoa da, teoriaz izan zitezkeen hamabi zifretakoetatik oso urrun. Beste aldetik, lexikoarekin konposaketa burutu ondoren egoera zein arku kopurua laburtu egiten da hasiera batean asko handitzea espero zitekeenean; nonbait lexikoak abstrazioa murrizten du eta.



II.7 irudia.- Lexiko-itzultzaile orokor bat lortzeko urratsak (Karttunen 94)-n oinarritua.

Lortutako emaitzak ikaragarriak dira, frantseserako honako datu hauek ematen dituztelarik: 50 K-egoera eta 100 K-arku inguru, denak Mega bat baino gutxiago hartzen duena[1] eta zenbait mila hitz/segundo abiadura-ordena analisian. PC-KIMMO baina ehundaka aldiz azkarrago.

Bukatzeko hausnarketa bat: lexiko-itzultzaile hauek bi mailatako morfologiaren hobekuntza dira edo eredu berri bat? Kapituluaren hasieran egiten genuen sailkapenera eta kasu-errebisiora itzuliz Tzoukermann-ek eta Liberman-ek proposatutakoarekin du zerbait

[1] Honetan kodetzeko teknikak ere badu bere garrantzia. Honetaz gehiago sakontzeko (Karttunen, 90) erreferentzia da lagungarri.

amankomunean: erregelak desagertu dira exekutatzen den prozesadore morfologikotik. Honen ondorioz, exekuzioaren ikuspuntutik eredu berri baten aurrean gaudela esan badaiteke ere, lengoaia baten morfologia deskribatzeko garaian bi mailatako eredutik oso gertu dago erregelak bi mailatakoak baitira.

Lexiko-itzultzailearen kasuan eredu berri baten aurrean gaudela argiagoa da, erregela-sistema desberdinen konposaketak eskaintzen duen deskribapen-ahalmena batetik, eta deskripziorako erraztasuna bestetik, bide berriak irekitzen baititu (ikus II.7 irudia). Beraz, bi mailatako eredu batetik askoz orokorrago eta ahalmentsuago den maila anitzeko beste batera pasa gara, aplikazio-esparruk asko zabaltzen delarik. Aplikazioez sakontzeko interesgarriak dira Chanod-ek (1994) egiten duen frantses-aditzaren deskribapena eta Kwon-ek eta Karttunen-ek (1994) egiten duten korearraren deskribapen inkrementala.

Hurrengo kapituan, euskararen gaineko aplikazioan hain zuzen ere, lexiko-itzultzaile hauen ahalmenaz eta dagozkien tresnen erabileraz sakontzeko aukera egongo da.