

Caprate

Lengoaia Naturalez idatzitako problemen interpretaziorako sistema

Egilea: **KEPA SARASOLA GABIOLA**

Urtea: 1988

Zuzendaria: M.F. VERDEJO

Unibertsitatea: UPV-EHU

ISBN: 978-84-8428-266-9

Hitzaurrea

Tesi hau “hizkuntzaren tratamendu automatikoa” arloak Euskal Herrian egin zuen hasiera-hasierako urratsetako bat izan zen. Felisa Verdejo eta Paco Garijo ikerlariak, gai horretan doktore berri zirenak, 1981.ean Parisetik etorri ziren EHUko Informatika Fakultate berrira irakasle modura aritzera. CAPRA izeneko proiektua martxan jarri zuten. Proiektu horretan, konputagailu bidez konputagailu-programak idazten irakatsi nahi zen. Proiektuaren barruan bi doktorego-tesi gorpuztu ziren hizkuntzaren tratamenduaren inguruan, 1988.ean eta 1990.ean aurkeztu zirenak. Kepa Sarasolaren tesian, gazteleraz idatzitako programazio-problemen enuntziatuak automatikoki ulertzeko sistema bat eraiki zen. Beste tesian, Arantza Díaz de Ilarrazarenean, gaztelerazko elkarrizketak kudeatzen ziren, programak idazten irakasten zuen tutore automatikoa eta ikaslearen arteko elkarrizketak, hain zuzen ere. CAPRA sistema gaztelaniarako egin bazen ere, euskararako ere egokitu ziren zenbait modulu .

Urte bertsuetan Joseba Abaitua hizkuntzalariak Manchester-en aurkeztu zuen bere tesia. Euskararen morfologia eta syntaxirako gramatika lexiko-funtzional bat proposatu zuen. Hiru tesi horiek izan ziren aitzindariak “hizkuntzaren tratamendu automatikoa” arloan Euskal Herrian.

CAPRA sistemaren fruitua prototipo ausart, berritzaile eta sendo bat izan zen, baina geroago ez zen inoiz bilakatu produktu publiko bat, produktu erabilgarri bat. Bere ekarpen handiena heziketa, ikerlarien trebakuntza izan zen, CAPRA proiektuan trebatutako doktore berri haiek lxa taldea sortu baitzuten geroxeago. 1990.ean hasi ziren euskararen tratamendu automatikoaren bidea urratu nahian. Hasierako urratsak itzulpen automatikora bideratu baziren ere, laster konturatu ziren alferrikakoak izango zirela ahalegin haiek, lehenago ez bazuten sortzen euskararen lexikoa eta morfologia lantzeko oinarria. Beste hizkuntzetan morfologiaren tratamendua oso erraza zen, baina euskararen kasuan hori ez zela egia nabaritu zuten berehala. Horregatik itzulpengintza automatikoaren gaia eta esaldiak sakonki ulertzeko proiektua, biak ala biak, aparkatuta utzi zituzten, eta hasierako indar haiek ganorazko analizatzaile morfologiko bat sortze aldera inbertitzea erabaki zuten. Garai hartan Finlandian antzeko arazoak zituzten beren hizkuntzarekin, baina Kimmo Koskeniemi irakasle finlandiarrak asmatutako *Bi Mailatako Morfologia* metodologia erabiliz morfologia konputazionalki landu zuten arrakastaz. Hori ikusita guk ere metodologia bera aplikatu genuen euskararekin, eta arrakastaz aplikatu ere. Geroago, 1994.ean, Donostiako Informatika-Fakultatean 7 partaideko talde bat ari zen euskararen inguruko ikerketan, eta *Xuxen* zuzentzaile ortografikoaren lehen bertsioa plazaratzen genuen. Argi dago CAPRATE tesian eta CAPRA proiektuan egindako testu-azterketek oinarritzko ezagutza eta metodologia eman zigutela, oinarri hori funtsezkoa izan baitzen geroago hizkuntzaren teknologian euskararen lehen urratsak ondo bideratu ahal izateko.

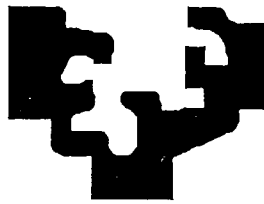
1988. urtean aurkeztutako tesi-lana ausarta, berritzailea eta sendoa izan zen. Konputagailu-programazioko enuntziatuen lau-bost esaldiko enuntziatuak ulertzea zen helburu. Gaurko perspektibatik ondotxo dakigu orduko helburuak oso altuak zirela, handiegiak ere agian. Gaur egun ondo dakigu jakin inongo aplikazio errealetan ez dela saiatzen hainbat esaldi ulertzen, sakonki dena

ulertzea zailegia da eta. Baina gure CAPRA proiektuan aplikazio-eremua nahiko murrizta eta zehatza zen, egokia zen orduan lortu genituen baleko emaitzak lortu ahal izateko. Geroago, euskarari heldu genionean, helburu apalagoak, errealistagoak, jarri genituen, emaitza aplikagarriagoak lortzearren. Beste hizkuntzetan ere joera bera nagusitu izan da urte guzti horietan. Baina oso litekeena da epe ertainean nazioarteko zientzia-komunitateak CAPRATE bezalako sistemei berriro ekitea. Ez orain dela 20 urteko erreminta apalekin, ez; oraingo analizatzaile sintaktikoak askoz ahaltsuagoak dira; oraingo baliabide semantikoak (Wordnet eta Euskal Wordnet adibidez) eskura izanda esaldien esanahia lortzeko xedea ez da hain “zaila”. Epe erdian ikusiko ditugu horrelakorik, eta orduan tesi honetan egin ziren ekarpen batzuk berriro izango dira baliagarriak, besteak beste, esaldien tipifikazioa, enuntziatuen esangura errepresentatzeko modua eta enuntziatuak interpretazioa lortzeko ikuspuntua.

Kepa Sarasola
EHUko *Informatika Fakultatea*
2009

**INFORMATIKA FAKULTATEA
FACULTAD DE INFORMATICA**

eman ta zabal zazu



EUSKAL HERRIKO UNIBERTSITATEA - UNIVERSIDAD DEL PAIS VASCO

**CAPRATE: LENGOAIA NATURALEZ IDATZITAKO
PROBLEMEN INTERPRETAZIORAKO SISTEMA**

Kepa Sarasola Gabiola-k

Informatikan Doktore-Gradua

lortzeko aurkezten duen

TXOSTENA

Donostia, 1988



Gizonen lana jakintza dugu
ezagutuz aldatzea
naturarekin bat izan eta
harremanetan sartzea

Xabier Lete

Eskerrak eman nahi dizkiet

lan honen motibatzaileei (Jose Ramon, Josu, Felisa),

garapenaldiko lagunei (CAPRA mintegikideak eta beren eritzi kritikoa eman didatenak),

programatze eta idazte orduko kolaboratzaileei (oro har Fakultateko lankideak, eta batik bat Luis Marco, Julen, Marisa, Julio, Josemi, Arantza, Iñaki eta Xabi),

euskarazko bertsio hau posible egin duten lagunei (Eduarne, Guadalupe, Jesús, Jose Ramón, Montserrat, Patxi, Pedro, Olatz eta Xabier),

Arantza Diaz-i bide osoko "aholkulari tekniko"-ari

eta batez ere eskerrak lan honen garrantzia desmitifikatuz, berau unibertsitate eta bizitarako projektu sakonagoetan kokatzen lagundu didazuen adiskide horiei.

AURKIBIDEA

1. SARRERA OROKORRA.....	1
2. FUNTZIONAMENDUAREN ADIBIDE ZEHATZA.....	9
2.1. SARRERA.....	11
2.2. ANALISI LEXIKO-MORFOLOGIKOAREN FASEA.....	12
2.3. ANALISI SINTAKTIKO-SEMANTIKOAREN FASEA.....	14
2.3.1. Sarrera.....	14
2.3.2. Lehen perpausa.....	20
2.3.3. Bigarren perpausa.....	29
2.3.4. Hirugarren perpausa.....	34
2.4. ESPEZIFIKAZIOAREN SORKUNTZAREN FASEA.....	37
2.4.1. Enuntziatuaren errepresentazioaren berregituratze eta osotasun-egiaztapena.....	37
2.4.2. Sorkuntza.....	40
3. AZTERKETA ETA KOKAPEN BIBLIOGRAFIKOA.....	41
3.1. SARRERA.....	43
3.2. SARRERA HISTORIKO LABURRA ETA EGUNGO LENGOAIA NATURALAREN ANALISATZAILEEN JOERAK.....	44
3.3. ATNetan OINARRITUTAKO ANALISATZAILEEN AZKEN GARAPENAK.....	46
3.3.1. Ezagumendu semantikoaren erabilpena.....	47
3.3.2. "Barrutik-kanporanzko" analisi-estrategia.....	48
3.3.3. ATNak turrustan.....	49
3.3.4. Konjuntzioen tratamendua.....	52
3.4. "IRAKASKUNTZA" ETA "PROBLEMEN ENUNTZIATUEN ULERMENA" DIRELAKO EREMUETAN ERABILITAKO LENGOAIA NATURALAREN ULERMEN-SISTEMAK.....	56
3.4.1. Ezaugarriak.....	56
3.4.2. SOPHIE.....	57
3.4.3. Mellish-en analisatzailea.....	61
3.4.4. Gomez-en programazioari buruzko problemen ebazpen-sistema.....	63
3.5. KOKAPEN BIBLIOGRAFIKOA.....	65

4. DISEINU OROKORRA ETA BESTE APLIKAZIOTARAKO TRESNA ERABILGARRIAK.....	69
4.1. SISTEMAREN EGITURA ETA FUNTZIONAMENDUA.....	71
4.2. BESTE APLIKAZIOTARAKO TRESNA ERABILGARRIAK.....	74
4.2.1. ATNen interpretatzailea.....	74
4.2.1.1. Ezaugarri orokorrak.....	74
4.2.1.2. Arkuak.....	74
4.2.1.3. Ekintzak.....	76
4.2.1.4. Testak.....	79
4.2.1.5. Forma ebaluagarriak.....	82
4.2.1.6. Analisirako estrategia	85
4.2.2. Hiztegiaren egitura.....	86
5. BEHARREZKO EZAGUMENDUAK ETA BEREN ERABILPENA...	93
5.1. APLIKAZIOAREN EREMUARI DAGOKION EZAGUMENDU EZ-LINGUISTIKOA.....	95
5.1.1. Errepresentazio-lengoaia.....	95
5.1.2. Azken errepresentazioaren osotasunari buruzko ezagumendua.....	99
5.1.3. Identifikadoreen tratamendua.....	102
5.2. EZAGUMENDU LINGUISTIKOAREKIN LOTUTAKO EREMUARI BURUZKO EZAGUMENDUA.....	102
5.2.1. Hiztegia.....	102
5.2.2. Analizatzaile lexikoa.....	106
5.2.3. ATN sintaktiko-semantikoak.....	107
5.2.3.1. Enuntziatuak interpretatzeko ikuspuntua....	107
5.2.3.2. Aukeratutako kategoria sintaktiko eta semantikoak.....	108
5.2.3.3. Sortutako ATNen deskribapena.....	109

6. LORTUTAKO EMAITZAK.....	135
6.1. TRATAMENDU SINTAKTIKOAREN EBALUAZIOA.....	137
6.2. TRATAMENDU SEMANTIKOAREN EBALUAZIOA.....	144
6.3. TRATAMENDU PRAGMATIKOAREN EBALUAZIOA.....	149
6.4. EBALUAZIO OROKORRA.....	153
6.4.1. CAPRA projektuan azaldutako helburuak.....	153
6.4.2. Enuntziatu-sailkapena.....	154
6.4.3. Ebaluazioa.....	156
ONDORIOAK.....	159
BIBLIOGRAFIA.....	163
ERANSKINAK.....	173
A ERANSKINA : Inplementazioaren ezaugarriak.....	175
B ERANSKINA : Objektuen errepresentazioan erabilitako predikatuak.....	177
C ERANSKINA : Problema ebazkarrien enuntziatuak.....	181

1. SARRERA OROKORRA.

1. SARRERA OROKORRA.

CAPRATE sistemaren helburua Lengoaia Naturelez idatzitako Programazioko problemen ulermen automatikoa da. Problema bakoitzerako lortu nahi den emaitza, espezifikazio-lengoaia batez problemari dagokion errepresentazio formalak da. Lan hau Ordenagailuz Lagunduriko Irakaskuntza Adimentsuaren gaineko proiektu orokorrago batetan sartuta dago. CAPRA izeneko proiektu hau oinarriko programazioaren irakaskuntzarako diseinatu da, ondoko hiru muduluok dauzkalarik:

- Lengoaia Naturelezko interface-a duen sistema tutorea [Fernandez et al. 88].
- Programa-eraikuntzari buruzko sistema aditua [Garijo et al. 86]
- Ikaslearen ezagumenduen eta jokaeraren ereduak erabiltzen dituen ezagumendu-basea.

Irakasteko metodologiak programak diseinatzeko estrategia klasiko baten pausoak burutzen ditu: Lehenengo pausoa ikasleak espezifikazioa idazten du, bigarrenetan algoritmo abstraktoa, eta amaitzeko zenbait birfinketa egin eta gero programa lortzen du. Modulu Tutoreak sailtasun gorakorrez ordenatuta dauden problema-bilduma bat aurkezten dio ikasleari, ebazteko pauso bakoitzean laguntza eskaintzen diolarik.

Bi helburu lortu nahi izan dira CAPRATE sistema diseinatzeko. Alde batetik, zailtasun-maila bereko testuak prozesatzeko beharrezko diren ezagumenduak aztertzea. Beste aldetik, CAPRA sistemari modulu berri bat eranstea Lengoaia Naturelez idatzitako enuntziatu berriak eta bere espezifikazio formalak onartu ahal ditzen.

Zailtasun-maila bereko testuen artean bi eremu-multzo posible ikusten ditugu: zenbait objektu eta beren arteko erlazioak deskribatzen dituzten problema-enuntziatuak, eta matxuretarako konponbideak deskribatzen dituzten eskuliburuak.

Batetik linguistikari eta bestetik aplikazio-eremuari dagozkien ezagumenduak bereiztea diseinurako printzipio bat izan da. Zenbait sistemetan [Brown and Burton 75], [Gómez 81] ezagumendu-iturri bion arteko harreman estu-estuak definituz emaitza harrigarriak lortu dira oso eremu konkretuetarako, baina beste aplikazio-eremuetarako aplikagarritasuna guztiz urria izaten da horrelako sistemetan. Ezagumendu-mota biak modu independientez adierazten badira, ezagumendu linguistikorako egindako azterketa beste aplikaziotan ere erabilgarria izango da, eta gainera, ikuspuntu linguistiko huts batetik ere interesgarria gerta liteke.

Ezagumenduen bereizketa honen ondorioz, ezaugarri bereak dituen beste aplikazio batetarako egokitze-lana errazagoa litzateke, aplikazio-eremuari dagozkion ezagumenduak baino ez bailirateke aldatu beharko. Era berean, buruturiko tresna linguistikoak enuntziatuen ulermenaz gain beste xedeetarako ere balio lezake, hain zuzen ere, ikasle eta modulu tutorearen arteko interface-a eraikitzeke

Ezagumendu-mota bien arteko bereizketa lortu nahi denez, ezin daiteke hauta analisi sintaktiko eta semantiko sakonak egiten ez dituen analisatzaileak, ezta ulertzeko prozesua aplikazioaren informazio hutsen bidez gidatzen duen analisatzaileak ere. Azken urteotan analisatzaile-mota desberdin anitz garatu dira. Gaur egungo ikerkuntzaren joera batek analisi semantikoen optimizazioa eta beronek analisi sintaktikoarekin eduki lituzkeen harreman sakonak aztertzen ditu. CAPRATE-ren tratamendu semantikoa diseinatzeko jarri diren helburuak ondokoak dira:

- Tratamendu semantikoa ez da exekutatu behar sintaktikoa bukatu ondoren, elkarlanean tartekatu behar dira biak. Honela izanda, sintaktikoki bai baina semantikoki zuzen ez diren analisi-aukera asko, lehenbailehen baztertu ahal izango da, analisi sintaktiko osoa bukatu arte itxaron gabe.
- Tratamendu sintaktiko eta semantikoen tartekaketak beste aukera erakargarri bat eskaintzen du: esanguraren errepresentazioa ez da atzeratu behar testu osoa edo esaldi bakoitza guztiz analisatu arte. Testuko osagai sintaktiko esanguratsu bat agertzen den bezain laster bere errepresentazioa eraikitzen hasi ahal da. Semantikaren

erabilpen goiztiar hau izen-sintagmaren mailan eramaten badugu [Mellish 85], bi abaintail aterako ditugu: alde batetik, eraginkortasunez irabazpen bat lortzen da, eta bestetik, laguntza berri bat eskaintzen da erreferentziak ebazteko, informazio sintaktikoaz gain aurretik eraikitako errepresentazioaren informazio semantikoa erabili ahal izango baita.

- Aukeratutako aplikazio-eremuan maiz agertzen diren elipsi eta erreferentzi-ebazpenaren problemak gainditzeko gaitasuna izan behar du tratamendu semantikoak.
- Objektuen errepresentazioa gehikuntzaz eraiki behar da, hau da, objektuen deskripzioa ez da bukatutzat emango objektua aurkezten duen izen-sintagma amaitzen denean, testuko ondorengo hitzak analisatzerakoan ere objektuaren kategoria semantikoa, izena, eta propietateen definizioak osatzeko aukera egongo da.

Hiru fasetan banatzen da itzulpena. Lehenengoan analisi lexiko-morfologikoa burutzen da. Bigarrenengan analisi sintaktiko-semantikoaren bitartez enuntziatuko objektuen errepresentazioa lortzen da. Eta hirugarrenean, azken fasean, objektuen errepresentazioa osatu ondoren enuntziatuari dagokion espezifikazioa eraikitzen da.

Analisi sintaktiko-semantikoaren fasea tratamendu semantikoaz aberastu ditugun ATNen bidez implementatuko da. Tresna hau erabiliz ezagumendu linguistikoak eta aplikazio-eremuari dagokion ezagumendua ondoko mailetan banatu ahal izan dira beren arteko loturaren garrantziaren arauera: ATNen interpretatzailea (eremuarekiko guztiz independentea dena), ATN-multzoa (eremuarekiko menpetasun liraina duena, aplikazio-eremuko azpilengoaia mugatzen baitu), eta hiztegia (aplikazio-eremuaren menpekoa sarrera bakoitzari bere esangura dakarrelako, baina linguistikaren menpekoa ere bere kategoria sintaktiko eta semantikoak dauzkalako).

Aplikazio-eremuari dagokion ezagumendua funtsezkoa da espezifikazioa lortzeko azken fasean, bertan aztertzen baitira tinkotasunari buruzko baldintza orokorrak.

Aurkezten den txosten honen edukia ondokoa da:

Bigarren kapituluak sistemaren funtzionamenduaren adibide zehatz bat aurkezten du. Bertan CAPRATE sistemaren osagaiei buruzko gainerako deskripzioak ematen dira. Sistemaren ahalmen garrantzitsuenak erakusten dituen enuntziatu adibide bat aukeratu da, bere prozesamenduko fase guztiak azaltzeko.

Hirugarren kapituluak azterketa bibliografikoa eta lan honen kokapena biltzen dira. Lehenengo sailean Lengoia Naturean barruan egun dauden iker-lerro nagusiak aurkezten dira. Bigarren sailean ATNen formalismoaren historia azaltzen da, azken urte hauetan Bates-en eta Winograd-en eredu klasikoetarako ([Bates 78] eta [Winograd 83]) proposatu diren hobekuntzak erakusten dira. Hirugarren sailean CAPRATE-ren aplikazio-eremu berberetan mugitzen diren hiru sistema errepresentagarri deskribatzen dira. "Ordenagailuz Lagunduriko Irakaskuntza Adimentsua" eta "Problemen enuntziatuen ulermena" dira aztertutako eremuak.

Laugarren kapituluak sistemaren diseinu orokorra eta bere hiru osagai nagusiak azalduz hasten da, eta beste aplikazioetarako erabilgarri diren tresnen azalpenarekin amaitzen da. Tresna hauen bidez, aplikazio-eremuarekin zerikusirik ez duten ezagumendu linguistikoak adierazten dira.

Bostgarren kapituluak beharrezko diren bestelako ezagumendu guztiak agertzen dira. Banatu egiten dira alde batetik aplikazio-eremuari dagokion ezagumendu ez-linguistikoa, eta bestetik, ezagumendu linguistikoarekin lotutako eremuaren ezagumendua.

Bostgarren kapitulu honetako lehenengo saila beste bi apartatuetan banatzen da. Lehengoan enuntziatuen esangura errepresentatzeko definitu den lengoia aurkitzen da, eta bigarrenean enuntziatuen osotasuna egiaztatzeko erabiltzen den ezagumendua.

Bostgarren kapituluko bigarren sailean hiru tresna azaltzen dira: hiztegia, analisatzaile lexiko-morfologikoa eta espainieraz idatzitako enuntziatuak analisatzeko definitu den ATN-multzoa.

Seigarren kapituluan lortutako emaitzen ebaluazioa egiten da. Analisiaren fase bakoitzerako eta itzupen-prozesu orokorrerako egiten da. Sistemaren etorkizuneko garapenerako zenbait proposamen aurkezten dira.

Sistemaren erakarpen nabarmenenak bereizten eta laburtzen dituzten ondorioek bukaera ematen die txostenari.

Hiru eranskin gehitzen dira: inplementazioaren ezaugarriak, objektuen errepresentazioan erabilitako predikatuen zerrenda eta ebazitako problemak.

2. FUNTZIONAMENDUAREN ADIBIDE ZEHATZA.

2.1. SARRERA.

2.2. ANALISI LEXIKO-MORFOLOGIKOAREN FASEA.

2.3. ANALISI SINTAKTIKO-SEMANTIKOAREN FASEA.

2.3.1. Sarrera.

2.3.2. Lehen perpausa.

2.3.3. Bigarren perpausa.

2.3.4. Hirugarren perpausa.

2.4. ESPEZIFIKAZIOAREN SORKUNTZARAKO FASEA.

2.4.1 Enuntziatuaren errepresentazioaren berregituratze eta osotasun-egiaztapena.

2.4.2. Sorkuntza.

2. FUNTZIONAMENDUAREN ADIBIDE ZEHATZA.

2.1. SARRERA.

Kapitulu honetan, CAPRATE sistemak tratatzen dituen problemen enuntziatuak ulertzeko beharrezkoa den konplexutasun-maila azaltzen da, eta honekin batera analisi-prozesuaren funtzionamendua erakutsi. Aukeratutako enuntziatuan biltzen da gure sistema ebazteko gai den eragozpen sintaktiko, semantiko eta eremuari dagozkionen atal esanguratsu bat; beste enuntziatu adibideetan agertzen diren eragozpeneei buruzko eztabaida 4., 5. eta 6. kapituluetan agertzen da.

Erabiliko dugun enuntziatu adibidea honako hau da:

Dada una secuencia S de 20 números enteros, obtener otra secuencia con los elementos pares de S. Determinar su longitud.

CAPRATE sistemak, enuntziatu bakoitzerako sortzen du espezifikazio bat, CAPRA projektuan definitutako espezifikazio-lengoaiaz adierazten dena. Aukeratutako adibidean emaitza ondokoa da:

ESPECIFICACION

ENTRADA: Sea S secuencia_de ent definida

LONG(S)=20

SALIDA: Calcular sek77 SUBSEC_DE S formado_por_los I1 de S tal_que I1 mod 2 = 0

Calcular elem70 ent tal_que

LONG(sek77)=elem70

FINESP

Espezifikazioa 3 fase kontsekutiboetan eraikitzen da

1) Analisi lexiko-morfologikoa. Problemaren testutik abiatuz enuntziatua osatzen duten lexiko-unitateen zerrenda sortzen du; unitate bakoitzari zein hiztegiko sarrera dagokion, zein kategoria sintaktikoa eta behar denean morfologia arloko bestelako informazioak zehaztuz.

2) Analisi sintaktiko-semantikoa. Aurreko fasearen emaitzatik abiatuz, enuntziatuan agertutako objektuen deskripzioa adierazten duen asertzio-multzoa sortzen du; denbora berean analisi sintaktikoaren arbola lortzen da informazio semantikoz osatuta dagoena.

3) Espezifikazioaren sorkuntza. Asertzio-multzoa datu gisa hartuz espezifikazioa sortzen du, CAPRATE sistemaren azken emaitza hain zuzen.

Aukeratutako enuntziaturako fase bakoitzeko bilakaera azaltzen da ondoan.

2.2. ANALISI LEXIKO-MORFOLOGIKOAREN FASEA.

Problemaren testutik abiatuz enuntziatua osatzen duten lexiko-unitateen zerrenda sortzen du; unitate bakoitzari zein hiztegiko sarrera dagokion, zein kategoria sintaktikoa eta behar denean, "numero", "genero", "persona", "tiempo", "modo" eta morfologia arloko honelako informazioak zehaztuz.

Jeneralean, hiztegiko sarrera batek bere zenbaki, genero, pertsona, denbora edo moduaren arauerako aldakuntza guztiak adieraziko ditu. Adibidez, "blanco" sarrerak adieraziko ditu "blanca", "blancos" eta "blancas" hitzak. Guztiak amankomunean duten informazioa hiztegiko sarreran agertuko da; bereizten dituen informazioa (generoa eta zenbakia kasu honetan) analisatzaile lexiko-morfologikoak burutuko du. Salbuespenez, aldakuntz bakoitzak oso erabilpen-maiztasuna handia duenean, guztiak izango dira hiztegiko sarrera bereiztuak: adibidez, "un, una, unos, unas, el, la, los, las" artikulua hiztegiko sarrera desberdinak dira.

Unitate lexiko batek analisi bat baino gehiago ekar dezakeela kontutan hartuz, horietako bakoitza analisi posible guztien zerrenda batez errepresentatuko da.

Aukera desberdin bakoitza zerrenda batez ere errepresentatzen da. Bere lehen osagaia, hiztegiaren atzipenari dagokion aukera-mota bereizten duen hitza da. "Palabra" eta "fverbal"

direlakoek hiztegiko sarrera eta gainerako informazio morfologikoa duten posibilitateak sartzen dituzte. "Signop", "ident" eta "exp" direlakoek puntuazio-zeinuak, identifikadorek eta espresioak sartzen dituzte, hiztegiko sarrerarik zehaztu gabe; testutik jasotako karaktere-taldea besterik edukitzen ez dutelarik.

Aukera bakoitzeko bigarren osagaia dagokion hiztegiko sarrera da; ala identifikadorea, espresioa edo puntuazio-zeinua horrela balitz.

Gainontzeko osagaiak errepresentatzen dituzte hiztegian agertzen ez diren baina analisi lexikoan lortu diren informazioak. Adibidez: *[palabra elemento [num pl] [sufijos 0]]* zerrendak hiztegian "elemento" hitzaz eskuragarri dagoen informazioari "numero plural" ezaugarria gehitzen dio; generoari buruz ezer gehitzen ez diolarik, zeren "elemento" beti maskulinoa bait da eta horrexegatik hiztegian argi agertuko da. Azkenik, *[sufijos 0]* espresioak ez du ezer gehitzen gaztelanizko testuak aztertzean, baina erroa (hiztegiko sarrera) modifikatzen duten atzikiak azaltzeko erabiliko da euskarazko testuen analisisan, kasu sintaktikoa eta antzeko informazioak edukiko duelarik.

Enuntziatu adibiderako definituriko emaitza ondokoa da:

```
[[[palabra dar [cat verbo] [impersonal] [participio] [gen fem]
      [num sing] ] ]
```

```
[[palabra una [cat art] [sufijos 0]]
```

```
 [palabra una [cat pron] [sufijos 0]] ]
```

```
[[palabra secuencia [cat nom] [num sing] [sufijos 0]]]
```

```
[[ident S [sufijos 0]]]
```

```
[[palabra de [cat prep] [sufijos 0]]]
```

```
[[palabra 20 [cat adj] [sufijos 0]]]
```

```
[[palabra numero_entero [cat nom] [num pl][sufijos 0]]]
```

```
[[signop , ]]
```

```
[[palabra obtener [cat verbo] [impersonal] [infinitivo]
      [sufijos 0]]]
```

```
[[palabra otro [cat pron] [num sing] [gen fem] [sufijos 0]]
```

```
 [palabra otro [cat det9] [num sing] [gen fem] [sufijos 0]] ]
```

```
[[palabra secuencia [cat nom] [num sing] [sufijos 0]]]
```

```
[[palabra con [cat prep] [sufijos 0]]]
```

```

[[palabra los [cat pron] [sufijos 0]]
 [palabra los [cat art] [sufijos 0]] ]
[[palabra elemento [cat nom] [num pl] [sufijos 0]]]
[[palabra par [cat nom] [num pl] [sufijos 0]]
 [palabra par [cat adj] [num pl] [sufijos 0]] ]
[[palabra de [cat prep] [sufijos 0]]]
[[ident S [sufijos 0]]]
[[signop . ]]

[[palabra determinar [cat nom] [impersonal] [infinitivo]
 [sufijos 0]]]
[[palabra su [cat adj] [num sing] [sufijos 0]]]
[[palabra longitud [cat nom] [num sing] [sufijos 0]]]
[[signop . ]]
]

```

Kontutan hartzekoa da zerrendako elementu bakoitza zerrenda ere dela, bertan lexiko-unitate bakoitzerako posibleak diren analisiak daudelarik. Adibidez: "una" hitzari bi morfologi-analisi posible dagokio, lehena artikulua bezala eta bigarrena izenorde bezala. Analisatzaile sintaktikoak banan bana miatuko ditu aukera desberdinak aipatutako unitate lexikoa tratatzean.

Kontutan hartzekoa da ere "numero-entero" terminoa bi hitzek osatzen dutela eta analisatzaileak horrela ezagutzen duela.

2.3. ANALISI SINTAKTIKO-SEMANTIKOAREN FASEA.

2.3.1. Sarrera

Analisi fase hau ATNen bidez burutzen da tratamendu semantikoaz eta eremuari dagokionaz osaturik.

ATNak definitu ditugu testu bat komaz edo puntuz bereizturiko perpausa-multzo bat bezala analisatzeko; izen-sintagma, preposiziodun talde, adberbio-talde eta aditz motako osagaiak dituzten perpausak analisatzeko; eta determinatzaile, izen, izenorde, adjektibo, adjektibo-talde, erlatibo eta adberbio-taldeetan banaturiko izen-sintagmak ere analisatzeko.

Tratamendu semantikoa analisi sintaktikoarekin paraleloan burutzen da; perpaus, izen-sintagma, adjektibo, adjektibo-talde eta erlatiboazko menpekoei dagozkien ATNetan txertatzen delarik. Tratamendu semantiko honek 4 eginkizun nagusi ditu:

- a) Analisi sintaktikoak eskeinitako analisi-aukera batzuen bazterketa, semantikoki onartezinak direlako (ez bait dute esangura onargarririk ematen)
- b) Definizio semantikoa osatzeko beharrezko diren elementuen bilaketa gidatzen dituzten eskakizunen sorrera.
- c) Enuntziatuaren esangura errapresentatzen duen asertzio-multzoaren sorrera.
- d) Elipsi eta erreferentzien ebazpena.

Eginkizun hauek burutzeko asmoz unitate sintaktiko-semantiko bakoitzari nukleo bat definitzen zaio ondoko moduan:

<u>Unitate sintaktiko-semantikoa</u>	<u>nukleoa</u>
Perpausa	aditza
Izen-sintagma	izena (ez balego: izenordea, identifikadorea edo espresioa)
Adjektiboa	adj
Adjektibo-taldea	adj
Erlatiboa	aditza

Unitate sintaktiko-semantikoaren nukleoa izan daitezkeen hitzek ondoko informazioa dute elkaturik hiztegian:

- "Definizio semantikoa" osatzeko unitatea sintaktiko-semantikoan beharrezkoak diren gainontzeko osagaien kategoria semantikoa eta funtzio sintaktikoa. Balio dute analisiaren bukaeran bete beharko diren eskakizunak sortzeko nahiz semantikoki onartezinak diren analisi sintaktikoaren aukerak baztertzeko.
- Betebeharreko osagarrien eskakizun guztiak betetzen direnean burutuko diren ekintza semantikoak. Eskakizun

guztiak betetzen ez badira posibilitate hori bazterturik geratzen da.

- Beste osagarri gehigarri (betebeharreko ez direnak) posibleen kategoria semantikoa eta funtzio sintaktikoa; eta horiek erakartzen dituzten ekintza semantikoak.

Perpaus, izen-sintagma, adjektiboa, adjektibo-taldea, edo erlatibozko esaldi baten analisiak arrakasta duen bakoitzean asertzio berriak gehitzen dizkiote elkartutako ekintza semantikoek agertutako objektuen zerrendari. Informazio hori globala da, eta ondorioz, ATN guztiek atzi dezakete.

Objektu berriak sortzen dituzten ekintza semantikoak izen-sintagmak analisatu ondoren baino ez dira agertzen. Bere exekuzioa desberdina da analisaturiko izen-sintagmaren pluraltasun eta determinazioaren arauera.

- izen-sintagma indeterminatu singularra bada, objektu berria inolako egiaztapenik gabe sortzen da. Adibidez:

"Dado un numero primo N ..."

Adibide honetan N lehenengo aldiz agertzen da enuntziatuan, objektu konkretua da eta ez da izango aurretik sortutako objektu baten erreferentzia.

- izen-sintagma determinatu singularra bada, objektu berri bat sortu baino lehen objektu errepresentatuen artean bilatzen da analisatzen ari den izen-sintagma erreferentziazeko balio dezakeen bat. Erreferentzia lortzen bada, ez da objektu berririk sortuko, bestela objektuaren sorkuntza burutu egingo da. Erreferentzia zilegitzat hartzen da biek identifikadore berbera baldin badute edo mota berekoa nahiz propietate bateragarrikoak badira. Adibidez:

"Dada la secuencia S de 20 enteros
... la longitud de la secuencia S ..."

Kasu honetan, "la secuencia S" delakoaren lehen agerpenak objektuaren errepresentazioa ekarriko du, baina bigarrenak ez du ekarri behar zeren aurretik determinatutako objektu baten erreferentzi bezala erabili baita da.

- izen-sintagma indeterminatu plurala bada objektu-sorkuntzaren ekintza aldatu egiten da, objektuari [sekdesk] kategoria semantikoa asignatuz, hori geroago objektu egituratu baten deskribatzaile edo objektu egituratu bezala hartuko delarik. Normalean, objektu egituratu hori sekuentzia batean gauzatuko da, baina batzuetan multzotzat har liteke. Adibideak:

"... una secuencia de 20 numeros ..."
"Dados 20 números ..."

Lehenengo adibidean "20 numeros" delakoa objektu egituratu baten deskribatzailea da (sekuentzia bat kasu honetan). Bigarrenean, aldiz, 20 zenbakik osatzen duten objektu egituratua adierazten du harek.

- Izen-sintagma determinatu eta plurala bada objektu-sorkuntzaren ekintza aldatu egiten da, objektuari [sekdeskdetpl] kategoria asignatuz, hori indeterminatu pluralek sortutako kategoriaren [sekdesk] baliokide bezala interpretatu ahal izango delarik. Adibidez:

"... obtener otra secuencia con los elementos pares de S ..."

"Los elementos pares de S" izen-sintagmak objektu egituratu baten deskribatzailea adierazten du (sekuentzia batena kasu honetan).

Enuntziatu-adibidearen analisi-prozesua modu progresiboan burutzen da, komaz edo puntuz banaturik agertzen diren 3 perpausetan zatiturik hain zuzen ere.

Analisi sintaktiko-semantikoaren fasearen sarrera da aurretik egindako analisi lexiko-morfologikoaren irteera. Enuntziatuan agertutako objektuak errepresentatzen dituen asertzio-multzoak osatzen du irteera:

```

[desalida obj14 [ ]]
[longitud_inv obj14 [obj12]]
[tipo obj14 [elem ent]]
[tipo obj13 [borrado]]
[desalida obj12 [ ]]
[t_q_elem obj12 [obj13]]
[prop_par obj13 [ ]]
[contenido obj13 [obj5]]
[tipo obj13 [sekdeskdetpl]]
[tipo obj12 [sek]]
[deentrada obj5 [ ]]
[nombre obj5 S]
[tipo obj4 [borrado]]
[longitud obj5 [obj3]]
[tipo obj5 [sek ent]]
[tipo obj5 [sek]]
[tipo obj4 [sekdeskr ent]]
[valdef obj3 [20]]
[tipo obj3 [elem nat]]

```

Errepresentazio hau nola lortu den ondoko pasarteetan adierazten da urratsez urrats. Erabilitako predikatu bakoitzaren esangura II. eranskinean definitzen da. Hona hemen asertzio-multzo horren interpretazio labur bat:

- obj3 20 zenbaki osoa da.
- obj5 S izeneko 20 luzeera duen osozko sekuentzia bat da, enuntziatuan sarrera-datu gisa definiturik dagoena.
- obj4 ez da kontutan hartuko, azken finean "borrado" gisa definitu bait da.
- obj12 irteera-sekuentzia bat da, obj13 deskribatzaileak definitutakoa
- obj13 delakoak S-ren osagai bikoitiak adierazten du.
- obj14 irteerako zenbaki oso bat da, obj12 sekuentziaren luzeera hain zuzen.

Analisi-sintaktikoaren zuhaitza fase sintaktiko-semantiko honen bigarren emaitza da. Zuhaitz sintaktikoaren izen-sintagmek erreferentzia bat daukate, asertzio-multzoan errepresentatzen duten objektuari dagokiona

```
[[orac [verbo dar]
      [objdir [gnom [ref obj5]
                [art una]
                [nom secuencia]
                [gprep [prep de]
                        [gnom [ref obj4]
                                [[card [20 [ref obj3]]]
                                [nom numero_entero]]]]]]]]
[signo_punt ,]
```

```
[orac [verbo obtener]
      [objdir [gnom [ref obj12]
                    [det9 otro]
                    [nom secuencia]]]
      [gprep [prep con]
              [gnom [ref obj13]
                    [art los]
                    [nom elemento]
                    [[adj par]]
                    [gprep[prep de]
                        [gnom [ref obj5] [ident S]]]]]]]
[signo_punt .]
```

```
[orac [verbo determinar]
      [objdir [gnom [ref obj14]
                    [det9 su]
                    [nom longitud]]]]]
[signo_punt .]
```

Testu osoari begira, analisia etengabe burutzen bada ere, analisiaren garapena argituko dugu perpaus bakoitzaren bukaerako analisi-egoera guztiak azalduz.

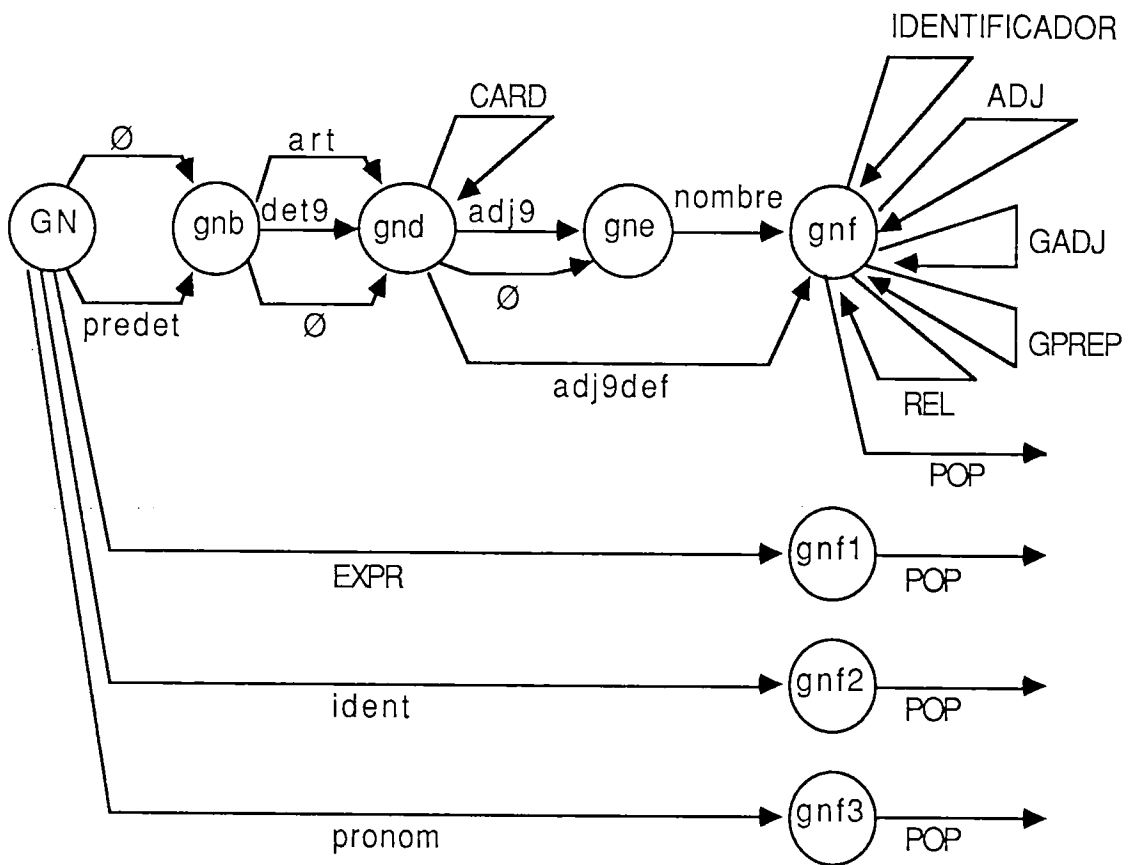
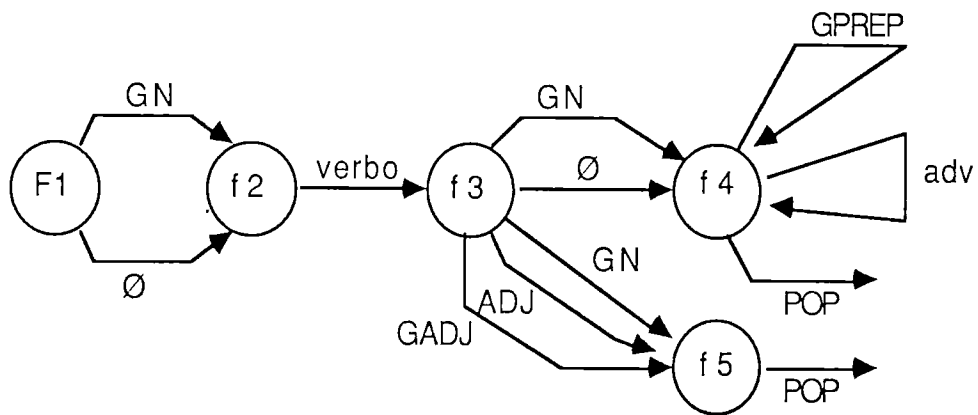
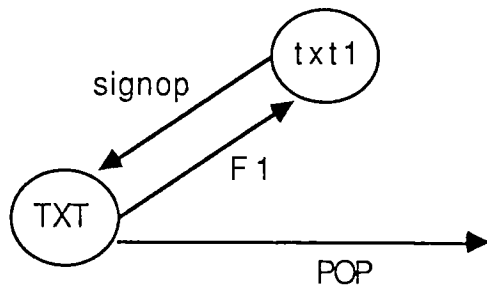
2.3.2. Lehen perpausa.

Lehen perpausa hau da:

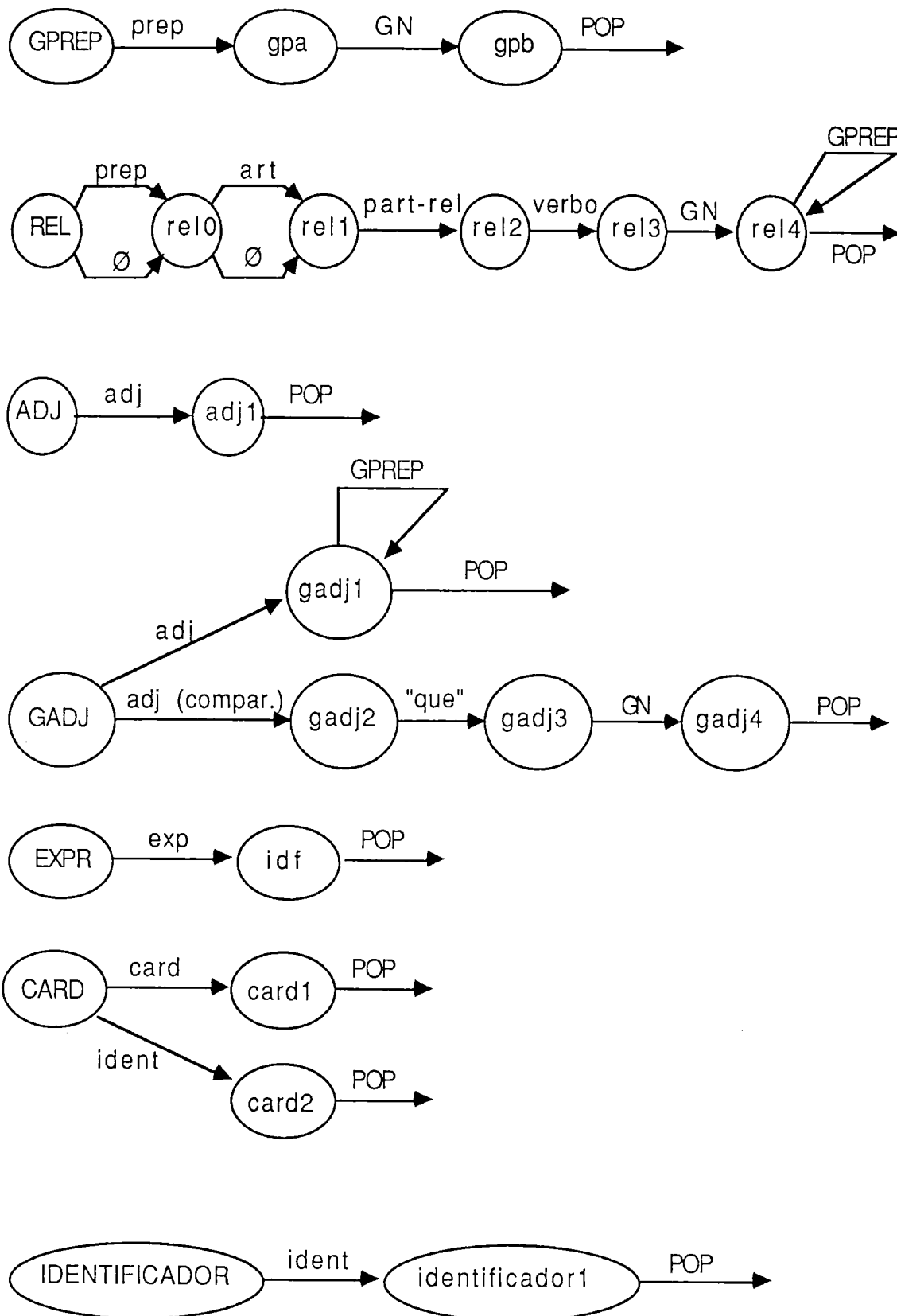
"Dada una secuencia S de números enteros"

Analisi sintaktiko-semantikoan hitzak hartzeko erabiliko den iturburu ez da perpaus hori, baizik eta analisi lexikoaren emaitz zerrenda. Hala ere, enuntziatu honetan bion arteko diferentzia eskasa denez ikusita eta notazioa sinplifikatzearen, hemendik aurrera testu-perpausaren hitzak aipatuko ditugu, zehaztasun osoz unitate lexikoak aipatu beharko bagenituzke ere.

Analisirako emandako urratsak ondoan adierazten dira 2.1.a eta 2.1.b irudiak erabiliz. Irudi hauetan sistemak erabilitako ATNen multzoa biltzen da. Horietako bakoitza 5.2.3.3 pasartean deskribatuta dago sakonki.



2.1.a. irudia. Definitutako ATNak.



2.1.b. irudia. Definitutako ATNak (jarr.).

Emandako urratsak honako hauek dira:

1) Analisia hasten da "TXT" ATNari dei eginez, enuntziatua aztertzeko. Halaber, ATN honek "F1" ATNa deitzen du perpausa analisatzearen. Iragaite hau ondokoa da: TXT F1 ?

2) "F1" ATNak subjektu sintaktikoa den izen-sintagma bilatu egiten du baina ezinezkoa gertatzen zaio, zeren "dada" aditzetik hasitako izen-sintagmarik ez bait dago. Subjekturik ez dagoela kontutan hartuz f2 egoerara jauzi egiten da. "dada" perpausaren nukleoa den aditz-forma bezala bereiztean aktibatzen dira "dar" aditzari dagozkion eskakizun semantikoak. Ondorioz, [obj] kategoriako (orokorra) objektu zuzen baten eskakizuna aktibatu egiten da. Aurkitzen bada ekintza sintaktikoek "de entrada" propietatea gehituko diote objektuari, enuntziatuaren datu bezala identifikatuko duena. f3 egoeratik analisiari jarraitzen zaio objektu zuzentzat erabiliko den izen-sintagma horren zai. Dagozkien iragaiteak ondokoak dira:

TXT F1 ?

F1 ∅ f2 verbo "dada" f3 GN ?

3) "GN" ATNan "gnb" egoerara jauzten da, testutik "una" artikulua hartu eta "gne" egoerara jauzi. Momentu honetan "sekuentzia" izen-sintagmaren nukleo bezala hartzen da, eta ondorioz, beren eskakizun semantikoak sortzen dira. "sekuentzia" izeneko definizio semantikoan ez da betebeharreko osagarri agertzen. Lotutako ekintza semantikoek sortzen dute [sek] motako objektu bat, sekuentzia bat adierazten duena, beren osagaiak enuntziatuan oraindik espezifikatu gabe dagoen motakoak direlarik. "sekuentzia"-ren definizio semantikoan osagarri gehigarri gisa sartzen da preposiziodun talde bat [sekdeskr elem] kategoriako "de" preposizioa duena ("de enteros" "de 20 caracteres"), sekuentziaren oinarri-mota eta luzeera (beharrezkoa denean) duen deskribatzailea adierazten duelarik.

Eskakizun horiek sorturik eta "gnf" egoeratik abiatuz hartzen da lehenengo arkua, IDENTIFICADOR izeneko, honek sarrera-testutik "S" kontsumitzen du eta ondorioz, analisia arrakastaz bukatzekotan, S izena lotuko zaio sortuko den objektuari.

"gnf" egoerara itzuliz, ezin da identifikadore gehiagorik hartu, ez bait dago adjektiborik, ezta adjektibo-talderik ere, baina bai preposiziodun talde bat.

TXT F1 ?

F1 ∅ f2 verbo "dada" f3 GN ?

GN ∅ gnb art "una" gnd ∅ gne
gne nombre "secuencia" gnf
gnf identificador "S" gnf GPREP?

4) "GPREP" ATNak "de" preposizioa kontsumitzen du eta "GN" ATNa deitzen du.

GPREP prep "de" gpa GN ?

5) izen-sintagma bat analisatu nahieran egoerak jauzten joaten da "gnd" egoeraraino iritsi arte, non ATNa deitzen duen arkuak "20" kardinala kontsumitzen bait du, errepresentaziorako ondoko bi asertzio sortuko direlarik:

[*tipo obj3 [elem nat]*] obj3 objektua elementu naturala da
[*valdef obj3 [20]*] obj3 objektuak balio definitua du (20)
eta konstantea da.

"gnd" egoeratik "gne"-ra jauzten da non "numero-entero" elementu konposatua kontsumitzen bait da, eta beren definizio semantikoak betebeharreko osagarririk ez duenez gero, asertzio bat sortuko du, [*elem ent*] kategoria definituz izen-sintagman adierazitako objektuarentzat.

"gnf" egoeratik POP arkuarekin saiatzen da. Arkuaren analisia arrakastatsua izan dadin, nukleoak sortutako eskakizun semantikoek buruturik egon behar dute. Horrela gertatzen da adibidean, "numero entero"-k ez bait zuen eskatzen betebeharreko osagarririk. Exekutatu behar den ekintza semantikoak [*elem ent*] kategoriako objektu baten sorrera ekarriko luke, baina "numeros enteros" izen-sintagma plurala eta indeterminatua denez

[*sekdeskr ent*] kategoriako objektu berri bat sortzen da, ondoko asertzioa gehituko delarik:

[tipo obj4 [*sekdeskr ent*]]

Mota honetako objektu bat objektu egituratu baten deskribatzaile bezala erabili ahal izango da, baita mota horretako objektu bat errepresentatzeko ere. Objektu egituratu bat sekuentzia batez errepresenta daiteke jeneralean. Hautaketa hori espezifikazioen sortzaileak egin beharko luke, inplementazio honetan beti sekuentzia bezala tratatzen bada ere.

GN ∅ gnb ∅ gnd CARD "20" gnd ∅ gne
gne nombre "números-enteros" gnf POP

6) "20 numeros enteros" izen-sintagma bezala analisatu ondoren, "de 20 numeros enteros" testu-zatia preposiziodun talde bezala prozesatzen da besterik gabe.

TXT F1 ?

F1 ∅ f2 verbo "dada" f3 GN ?

GN ∅ gnb art "una" gnd ∅ gne
gne nombre "secuencia" gnf
gnf identificador "S" gnf GPREP?

GPREP prep "de" gpa GN "20 números enteros" gpb POP

7) GPREP deia arrakastaz bukaturik, utzitako "GN" ATNaren prozesua ondo burutzen da "una secuencia de enteros" izen-sintagma bezala hartzen denean.

TXT F1 ?

F1 ∅ f2 verbo "dada" f3 GN ?

GN ∅ gnb art "una" gnd ∅ gne
gne nombre "secuencia" gnf
gnf identificador "S" gnf
gnf GPREP "de 20 números enteros" gnf POP

POP arkuak prozesatzeko berriro aztertzen da ea bete diren sortutako eskakizun semantikoak. Kasu honetan "secuencia" zen nukleoa, betebeharreko osagarriak eskatzen ez zuena eta [sekdeskri elem] kategoriako preposiziodun talde bat gehitzeko aukera zuena.

Betebeharreko eskakizunak (batere ez) buruturik daudenez, ekintza semantikoak exekutatu da: [sek] kategoriako objektu bat sortzea. Kasu honetan izen-sintagma indeterminatua eta singularra da, beraz objektua sortzeko ekintza aldaketarik gabe burutzen da.

[tipo obj5 [sek]]

Ondoren, "de 20 numeros enteros" aipatutako osagarri gehigarriari lotutako ekintzak burutzen dira:

[tipo obj5 [sek ent]]
[longitud obj5 [obj3]]
[tipo obj4 [borrado]]

Horrek obj5 sekuentzia motaren definizioa osatzea dakar, bere luzeera obj3 (hau da: 20) dela ezagutaraztea, eta azkenik, deskribatzaile zaharraren mota birdefinitzea "borrado" bezala, horrela sekuentzia bezalako ondorengo erabilpen posiblea baztertua gelditzen delarik.

Azkenik, sekuentziaren izen identifikadoreak eraikitako osagarri gehigarriari lotutako ekintza burutzen da.

[nombre obj5 S]

8) "una secuencia de 20 numeros enteros" izen-sintagmaren prozesaketa "dada" aditz-formak eskatutako objektu zuzen bezala bukatu ondoren, POP arkuak "f4" egoeratik ebaluatzen du positiboki erabilpen semantiko osoaren baldintza, zeren bete egin bait da zegoen eskakizun bakarra ([obj] kategoria orokorreko objektu zuzen batena, 2. urratsean aipatu den bezala). Lotutako ekintza semantikoak "de entrada" propietatea eransten dio aipaturiko objektuari, objektua problemaren datu bat dela adieraziz. Ondorioz asertzio hau gehitzen da.

[deentrada obj5 []]

Lotutako iragaiteak ondokoak dira:

TXT F1?

F1 ø f2 verbo "dada" f3
f3 GN "una secuencia de 20 numeros enteros" f4 POP

9) "F1" ATNak arrakastaz prozesatu du lehen perpausa, horrela "TXT" ATNra itzultzen delarik "txt1" egoeran geldituz. Geroago, "," unitate lexikoaren kontsumoak "TXT" egoerarako iragaitea baimentzen du, puntuazio-zeinua bait da, eta "F1" ATNa deituz, perpaus berri baten analisiarekin saiatzen da.

TXT F1 "dada una secuencia de 20 números enteros" txt1
txt1 signop "." TXT F1?

POP arkuaren prozesaketa ez da kasu honetan arrakastatsua izango, zeren eta ATN nagusiaren analisia bukatutzat hartzeko ATNen interpretatzaileak exigitzen bait du sarrera-testua guztiz agortuta egon dadin.

Lehen perpausaren analisia bukatu ondoren, kontsumitzeko lehen hitza "obtener" denean, analisi aukera bakarra dago, ondoko asertzioak sortzen dituenak:

```
[deentrada obj5 [ ]]  
[nombre obj5 S]  
[tipo obj4 [borrado]]  
[longitud obj5 [obj3]]  
[tipo obj5 [sek ent]]  
[tipo obj5 [sek]]  
[tipo obj4 [sekdeskr ent]]  
[valdef obj3 [20]]  
[tipo obj3 [elem nat]]
```

obj3 eta obj5 objektuak agertu dira. obj3 20 zenbaki osoa da eta obj5 osozko sekuentzia bat 20 luzeerakoa, S izenekoa eta enuntziatuan sarrera-datu gisa definiturik dagoena.

obj4 objektua ez da kontutan hartuko *[borrado]* motakoa bait da. Hasiera batean *[sekdeskr ent]* motako definiturik zegoen, sekuentzia bezalako interpretazio posiblea aurrakustearren; baina *[tipo obj4 [borrado]]* asertzioak (geroago, aukera hori baztertutakoan, erantsi dena) hasieran definitutako asertzioaren esangura ezabatu du.

Sortutako egitura sintaktikoa ondokoa da:

```
[[[orac [verbo dar]
      [objdir [gnom[ref obj5]
              [art una]
              [nom secuencia]
              [gprep [prep de]
                    [gnom [ref obj4]
                          [[card [20 [ref obj3]]]]
                          [nom numero_entero]]]]]]]]
 [signo_punt ,,]]
```

Egitura sintaktikoan errepresentatutako objektuen erreferentzia egiten dela ohartzekoa da: [ref obj5], [ref obj3]

Lehen perpausari dagokion analisi posible bakarra da hori. Analisia sintaktikoa soilik izan balitz ondokoa ere agertuko litzateke:

```
[[[orac [verbo dar]
      [objdir [gnom [ref obj5]
                  [art una]
                  [nom secuencia]
              [gprep [prep de]
                    [gnom [ref obj4]
                          [[card [20 [ref obj3]]]]
                          [nom numero_entero]]]]]]]]
 [signo_punt ,,]]
```

Aukera hau ez du gure analisatzaileak kontutan hartzen, zeren "dar" aditzak ez bait du "de 20 numeros enteros" preposiziodun taldea onartzen, ez betebeharreko osagarri bezala ezta gehigarri bezala ere. Hau dela eta, analisi posible hau baztertu egin da.

"una" hitzari dagokion analisi morfologikoaren bigarren aukerak, "una" izenordetzat hartzen duenak, ez dakar analisi-posibilitaterik. Analisiaren egoera ondoan erakusten dena denean, ezin dira GPREP eta "adv" arkuak hartu, POP arkuak besterik ez lezake har, baina ondoan behar-beharrezkoa litzateke puntuazio-zeinu bat, existitzen ez duena.

TXT F1 ?

F1 ∅ f2 verbo "dada" f3 GN "una" f4

Bestelako analisi sintaktiko partzialak baztertu dira erizpide sintaktiko soilei jarraituz, baina ez dugu uste haien azalpen zehatzagorik ematea beharrezkoa denik.

2.3.3. Bigarren perpausa.

Bigarren perpausa hau da:

"Obtener otra secuencia con los elementos pares de S"

Bere analisisian emandako urrats esanguratsuenak ondokoak dira:

1) Lehen perpausaren analisisia bukatu den "TXT" egoeratik "F1" ATNa deitzen da perpausa tratatzeko asmoz.

TXT F1 ?

2) Aurreko perpausan bezala "F1" ATNak subjektua den izen-sintagma bat bilatzen du arrakastarik gabe. "f2" egoerara jauzten da, non "obtener" aditzak (perpausaren nukleoa) eskakizunak sortzen bait ditu definizio semantikorako. Definizio honek sekuentzi bat adierazten duen izen-sintagma objektu zuzena (*[sek]* kategoriakoa) eskatzen du eta gainera sekuentzia-deskribatzailea adierazten duen preposiziodun talde bat ere. Horrelako osagarriak aurkitzen badira objektu zuzenari gehitzen zaizkio ondoko propietateak: "de salida" izatearena eta bere elementuak sekuentzi-deskribatzaileaz bereizturik egotearena. Beraz, "f3" egoeratik lehen arkuak hartzen da, "GN" ATNa deitzen duena. Azken

hiru arkuak ezingo dira gainditu, agertutako aditza ez bait dagokio "ser" aditzari, ezta bestelako aditz deklarati boei ere.

TXT F1 ?

F1 ∅ f2 verbo "obtener" f3 GN ?

3) "GN" ATNan "gnb" egoerara jauzten da. "otra" izenorde bezala hartuz izen-sintagma bukatutzat jotzen da, baina arku bat ere ezingo da geroago aukeratu "F1" ATNan analisiari jarraitzeko. Analisi-aukera hori bazterturik, "otra"-ren beste posibilitate morfologikoa (det9 determinantea izatearena) aztertzen da "gnd" egoerara pasatuz. "gne"-ra jauzten da "secuencia" izen-sintagmaren nukleotzat hartzen delarik. Bere eskakizun semantikoak abiatzen dira (lehen perpausan erabilitakoak)

"gnf" egoeratik IDENTIFICADOR eta ADJ prozesatzen saiatzen da, posiblea izan gabe. GPREP-aren analisiak "con los elementos pares" prozesatzeko aukera izango du, baita "con los elementos pares de S" ere, baina horiek ezin dute betebeharreko osagarriena egin, (sekuentziak ez bait du batere eskatzen) ezta osagarri gehigarriarena ere ("de" preposizioz hasitako preposiziodun taldeak ez bait dira). Hau dela eta, ezin da GPREP arku prozesatu. Ezta REL arku ere. POParkua arrakastaz gainditzen da "secuencia"-k ez bait du betebeharreko osagarriarik eskatzen. Lotutako ekintza semantikoak asertzio hau eransten du:

[tipo obj12 [sek]]

TXT F1 ?

F1 ∅ f2 verbo "obtener" f3 GN ?

GN ∅ gnb det9 "otra" gnd ∅ gne
gne nombre "secuencia" gnf POP .

4) GNren prozesua bukatuz, perpausaren ATNra itzultzen da "f4" egoeran. Iragaita hori burutu baino lehen, aipaturiko GNa "obtener" aditzak abiatutako definizio semantikoaren barruan egokitzen dela egiaztatu da. Noski, izen-sintagmaren kategoria semantikoa [sek] zen eta objektu zuzenarena egiten du.

"f4" egoeratik "GPREP" ATNa deitzen da.

TXT F1 ?

F1 ∅ f2 verbo "obtener" f3 GN "otra secuencia" f4
f4 GPREP ?

5) "GPREP" ATNan "con" preposizioa hartzen da eta "GN" ATNa deitzen da.

GPREP prep "con" gpa GN ?

6) "GN" ATNan "gnb" egoerara jauzten da, "los" artikulua hartu eta "gnd"-ra iragan. Gero "gne"-ra jauzten da, hemen "elementos" izena hartzen delarik. "elemento"-aren definizio semantikoak "de" preposizioa duen [sek] kategoriako preposiziodun talde bat eskatzen du eta eskakizuna betetzen bada, sortuko du [elem] elementu bat, sekuentziari dagokion "contenido" propietatea izango duena. Definizio semantiko hori abiaturik, analisiari jarraitzen zaio "gnf" egoeratik.

GN ∅ gnb "los" art gnd ∅ gne
gne "elementos" izena gnf

7) Puntu honetan ADJ arkuak "pares" adjektiboa kontsumitzen du, "elemento" aldatu dezakeela egiaztatu ondoren. Geroago, GPREP arkuak "de S" kontsumitzen du "elemento"-aren osagarri gehigarritzat hartzen duelarik. Kasu honetan izen-sintagma osoa identifikadore hutsa baino ez da.

GN ∅ gnb art "los" gnd ∅ gne
gne nombre "elementos" gnf
gnf ADJ "pares" gnf
gnf GPREP "de S" gnf POP

"los elementos pares de S" izen-sintagmaren analisia gertatu ondoren POP arkuak eragindako ekintza semantikoak honetan dautza: sekuentzia-deskribatzaile motako objektu baten sorreran ([elem] motakoa ez da sortzen izen-sintagma plurala delako), S-k erreferentziatutako sekuentziaren barruan egotearen propietatea

duena (obj5-en barruan gure kasuan). "pares" osagarri gehigarriari lotutako ekintza semantikoak propietate hori erazagutzen duen asertzio bat eransten du.

```
[tipo obj13 [sekdeskdetpl]]  
[contenido obj13 [obj5]]  
[prop-par obj13 [ ]]
```

8) Preposiziodun talde hori analisatu eta gero, "F1" ATNko "f4" arkura itzultzen da. "obtener" nukleoak abiatutako betebeharreko eskakizunak burutu direla egiaztatzen du POP arkuak (objektu zuzena den sekuentzia bat eta guztiz definituta dagoen sekuentzi-deskribatzaile bat "con" moduko preposiziodun talde batean)

TXT F1 ?

F1 ø f2 verbo "obtener" f3 GN "otra secuencia" f4
f4 GPREP "con los elementos pares de S" f4 POP

Sortutako asertzioak ondokoak dira:

```
[t_q_elem obj12 [obj13]]  
[desalida obj12 [ ]]  
[tipo obj13 [borrado]]
```

Hau da, obj12 da obj13 sekuentzi-deskribatzaileak definitzen duen sekuentzia bat eta gainera kalkulatzeko datu bat da enuntziatuan.

obj13 sekuentzi-deskribatzailearen mota "borrado" bezala berdefinitzen da, geroagoko sekuentzia bezalako erabilpen-possibilitatea ekiditzeko asmoz.

9) "*obtener otra secuencia con los elementos pares de S*", perpausa analisatuz gero, "txt1" arkura iragaten da 1. urratsari jarraituz. Ondoren, "." hitzak "TXT" egoerarako iragaitea egin daiteke, "F1" ATNa deituz perpaus berri bat analisatzeko prest geldituko delarik.

TXT F1 "obtener otra secuencia con los elementos pares de S"
txt1 signop "." TXT POP ?

Tratatzeko hitzik oraindik geratzen direnez, "POP" arkuak ezin da prozesatu.

Bigarren perpausaren analisiaren bukaeran, lehen hitz kontsumigarria "determinar" denean, analisi bakar bat dago eskuragarri, ondoko asertzioak sortu dituenak:

```
[tipo obj13 [borrado]]
[desalida obj12 [ ]]
[t_q_elem obj12 [obj13]]
[prop_par obj13 [ ]]
[contenido obj13 [obj5]]
[tipo obj13 [sekdeskdetpl]]
[tipo obj12 [sek]]
```

Bi objektu agertu dira: obj12 eta obj13. obj12 irteera-sekuentzia bat da, obj13 deskribatzaileak definitutakoa. obj13ak S-ren elementu bikoitiak adierazten ditu.

Sortutako egitura sintaktikoa ondoko hau da:

```
[[orac [verbo obtener]
  [objdir [gnom [ref obj12]
    [det9 otro]
    [nom secuencia]]]
  [gprep [prep con]
    [gnom [ref obj13]
      [art los]
      [nom elemento]
      [[adj par]]
      [gprep [prep de]
        [gnom [ref obj5] [ident S]]]]]]]]
[signo_punt .]
```

Arrazoi sintaktikoak direla medio ondoko analisi sintaktiko posibleak baztertu egin dira:

- 1) *Obtener otra secuencia con los elementos pares de S*

ez bait dago "obtener" delakoaren definizio semantikorik, analisi sintaktiko posible honetan agertutako hiru preposiziodun taldeak onartzen dituen.

2) *Obtener otra secuencia*
con los elementos pares
de S

Arrazoi berarengatik eta "*con los elementos pares*" ez bait da sekuentziaren betebeharreko osagarririk ezta gehigarririk ere.

3) *Obtener otra secuencia*
con los elementos pares
de S

"*con los elementos pares*" ez bait da sekuentziaren betebeharreko osagarririk ezta gehigarririk ere.

4) *Obtener otra secuencia*
con los elementos
pares
de S

"*pares*" delakoak ez bait dezake "*secuencia*" alda.

5) *Obtener otra secuencia*
con los elementos
pares
de S

aurrekoaren arrazoi berarengatik

2.3.4. Hirugarren perpausa.

Hirugarren perpausa hau da: "*Determinar su longitud*"

Bere analisisian emandako urrats esanguratsuenak ondokoak dira:

1) Bigarren perpausaren analisia bukatu den "TXT" egoeratik "F1" ATNa deitzen da perpaus bat tratatzeko asmoz.

2) Perpaus honetan (aurrekoetan bezalaxe) ez dago subjektua den izen-sintagmarik. "f2" egoerara jauzten da, hemen "determinar" aditzak (perpausaren nukleoa) bere definizio semantikoa abiatzen duelarik. [obj] kategoriako objektu zuzena den izen-sintagma bat betebeharrekoa da. Horrelako osagarria aurkitzen bada, bere errepresentazioari "de salida" propietatea erantsiko zaio, enuntziatuko datu bezala definituko duelarik. Ondorioz, "f3" egoeratik "GN" ATNa deitzen da.

3) "su longitud" izen-sintagma bezala analisatzean oztopo bat topatzen da, nukleoaren izenak ez bait du osatzen [sek] kategoria eta "de" preposizioa duen preposiziodun talde baten betebeharreko eskakizuna. Izen-sintagma determinatua denez eta elipsi edo erreferentzi bat izan daitekeenez ("su" adjektiboaren bidezko S sekuentziaren erreferentzia, gure kasuan), analisi hau baztertu baino lehen, jadanik errepresentatutako objektuen artean bilatzen da bat, ageri ez den betebeharreko osagarriaren propietateekin bat etorri behar duelarik. Kasu zehatz honetan [sek] kategoriako agerturiko objektuak bilatzen dira. Bi dira kategoria horretako objektuak: obj5 eta obj12. obj12 hartzen da sortutako berriena delako. Horrela definizio semantikoa betetzat jotzen da eta dagozkion ekintzak burutzen dira.

GN ø gnb det9 "su" gnd ø gne nombre "longitud" gnf POP

Sortutako asertzioak hauek izango dira:

[tipo obj14 [elem ent]]

[longitud obj14 [obj12]]

4) "su longitud" izen-sintagma analisaturik, perpausaren ATNko "f4" egoerara (2. urratsa) itzultzen da, horretan GPREP eta "adv" arkuen prozesuak alferrik saiatzen direlarik. POP arkuak egiaztatzen du "determinar" nukleoa abiatutako eskakizunak ea burutuak izan diren ([obj] kategoriako objektu zuzen bat). Ondorengo ekintza semantikoek sortutako asertzioa honako hau da:

[desalida obj14 []]

5) "determinar su longitud" perpausa analisatu ondoren, "txt1" arkura iragaten da 1. urratsari jarraituz. Ondoren "." hitzak "TXT" egoerarako iragaitea uzten du, "F1" ATNa deituz perpausa berri bat analisatzeko prest geldituko delarik. Hori alferrik izango da, sarrera-testuan ez bait da hitzik gelditzen.

Tratatzeko beste hitzik ez dagoenez, POP arkuaren prozesua arrakastatsua da oraingoan.

Azken POP arku honek itzultako balioa (eta beraz, analisi osoaren emaitza) zerrenda bat da, analisi sintaktiko osoa eta problemaren enuntziatuaren espezifikazioa deskribatzen duen arbola dauzkana. Espezifikazio hori lortzeko [gen-espec] forma ebaluatzen da, izen bereko funtzioa deitzen duena, horrela asertzio-zerrendari dagokion CAPRAko lengoaiatzko espezifikazioa sortzen delarik. Sorkuntzarako garatzen diren fase desberdinak hurrengo pasartean azaltzen dira. Hirugarren perpausaren analisisa bukatutakoan, sortutako egitura sintaktikoa honako hau da:

```
[[orac [verbo determinar]
      [objdir [gnom [ref obj14]
                [det9 su]
                [nom longitud]]]]]
[signo_punt .]]
```

Hirugarren perpausa honen analisisian erantsitako asertzioak ondokoak dira:

```
[desalida obj14 [ ]]
[longitud_inv obj14 [obj12]]
[tipo obj14 [elem ent]]
```

2.4. ESPEZIFIKAZIOAREN SORKUNTZAREN FASEA.

2.4.1. Enuntziatuaren errepresentazioaren berregituratze eta osotasun-egiaztapena.

Enuntziatuan ageriko objektuak adierazten dituen asertzio-multzotik hasita eta helburu-espezifikazioaren erabateko sorkuntzaren alde aurreko urrats bezala, aurreko ezagumenduaren berregituratzea egiten da, definizioa zuzena eta osoa dela ziurtatzeko, eta gainera ondorengo maneia errazteko.

Tratamendu hori burutzen da erregela desberdinak aplikatuz asertzio bakoitzari edo objektu bakoitzarekin lotutako asertzio-multzoari. Ondoko lau azpitratamendu hauek definitzen dira:

1) Objektu bakoitzari dagozkion asertzio guztien bilketa eta "borrado" kategoriari dagozkion ezabaketa. Azpitratamendu hau burutu ondoren lortutako emaitza hau da:

```
[[obj3 [tipo [elem nat]]
      [nombre [ ]]
      [valdef obj3 [20]]]
[obj5 [tipo [sek ent]]
      [nombre S]
      [longitud obj5 [obj3]]
      [deentrada obj5 [ ]]]
[obj13 [tipo [sekdeskdetpl]]
       [nombre [ ]]
       [contenido obj13 [obj5]]
       [prop_par obj13 [ ]]]
[obj12 [tipo [sek]]
       [nombre [ ]]
       [t_q_elem obj12 [obj13]]
       [desalida obj12 [ ]]]
[obj14 [tipo [elem ent]]
       [nombre [ ]]
       [longitud_inv obj14 [obj12]]
       [desalida obj14 [ ]]]]
```

2) Zenbait predikaturen definizioa osotasun eta zuzentasunaren denaren egiaztapena. Gure adibidean, azpitratamendu honek zerean gauzatzen da: bi asertzio bilzen dira esangura osoago duen asertzio berri bakar bat sortzeko, aurreko biak ordezkatzeko dutelarik.

"t-q-elem" predikatuak bi argumentu du: lehenengoak sekuentzia errepresentatzen du eta bigarrenak sekuentziaren elementuak definitzen dituen deskribatzailea.

"t-q-elem-sec" predikatuak hiru argumentu du: lehenak eta hirugarrenak sekuentziak errepresentatzen dituzte, lehenengoa hirugarren argumentuak adierazitako sekuentziaren azpisekuentzia delarik. Bigarren argumentuak sekuentzi deskribatzaile bat errepresentatzen du, non definitzen bait da lehen argumentuak definitutako sekuentziaren gainean ze murrizte-erizpide erabili den hirugarren argumentuari dagokion sekuentziaren osagaiak lortzeko.

[t_q_elem sec1 [desc1]] motako asertzio bat eta [contenido desc2 [sec2]] motako beste baten arteko bilkura ezartzen duen erregela bat dago, bietako deskribatzaileak berdinak badira [t_q_elem_sec sec1 [desc1 sec2]] asertzio lortzen delarik. Gure kasuan ondoko bi asertzioen arteko bilkura:

```
[t_q_elem obj12 [obj13 ]  
[contenido obj13 [obj5 ] ]
```

honako asertzio hau izango da:

```
[t_q_elem_sec obj12 [obj13 obj5]]
```

Honetan, obj12, obj5-en osagai bikoitiek osatutako azpisekuentzia dela adierazten da (kontutan hartu obj13 deskribatzaileari bikoitia izatearen propietatea lotzen dion asertzioa: [prop-par obj13 []]).

Modu berean, definizioaren azterketa hau beste predikatuen gainean egin lezake:

Azpitratamendu honen ondoren lortutako emaitza ondokoa da:

```

[[obj3 [tipo [elem nat]]
      [nombre [ ]]
      [valdef obj3 [20]]]
[obj5 [tipo [sek ent]]
      [nombre S]
      [longitud obj5 [obj3]]
      [deentrada obj5 [ ]]]
[obj13 [tipo [sekdeskdetpl]]
       [nombre [ ]]
       [prop_par obj13 [ ]]]
[obj12 [tipo [sek]]
       [nombre [ ]]
       [t_q_elem_sec obj12 [obj13 obj5]]
       [desalida obj12 [ ]]]
[obj14 [tipo [elem ent]]
       [nombre [ ]]
       [longitud_inv obj14 [obj12]]
       [desalida obj14 [ ]]]]

```

3) Objektuen sailkapena ondoko hiru kategorien artean: entrada, salida eta noentradanosalida.

```

obj3 "noentradanosalida" kategoriakoa, balioa definituta
     duen konstantea bait da
obj5 "entrada" kategoriakoa, "de entrada" propietatea
     bait du
obj13 "noentradanosalida" kategoriakoa, deskribatzailea
      bait da
obj12 "salida" kategoriakoa, "de salida" propietatea bait
      du
obj14 "salida" kategoriakoa, arrazoi berarengatik

```

4) Izenen asignazioa. Izen bat lotzen zaie lehenago definiturik ez duten objektuei. Izena sortzeko identifikadore bat sortzen da, motako lehen hitzaren ondoan zenbaki arrunt bat jarriz. Eraitza hau da:

```

[[obj3 [tipo [elem nat]]
      [nombre [ ]] [valdef obj3 [20]]]
[obj5 [tipo [sek ent]]
      [nombre S]

```

```

[longitud obj5 [obj3]]
[deentrada obj5 [ ]]
[obj13 [tipo [sekdeskdetpl]]
[nombre sekdeskdetpl20]
[prop_par obj13 [ ]]]
[obj12 [tipo [sek]]
[nombre sek77]
[t_q_elem_sec obj12 [obj13 obj5]]
[desalida obj12 [ ]]]
[obj14 [tipo [elem ent]]
[nombre elem70]
[longitud_inv obj14 [obj12]]
[desalida obj14 [ ]]]

```

2.4.2. Sorkuntza.

Azken fase honetan, enuntziatuan ageritako objektu-multzorako lortutako azken errepresentazioa itzultzen da espezifikazio-lengoiara.

Objektu bakoitza bere izena, mota eta berari dagozkion propietateez espezifikatuko da, espezifikazio-lengoiaren sintaxiari jarraituz. "entrada" propietatea duten objektuak espezifikatuko dira lehen, "salida"koak gero. "noentradanosalida" propietatedunak ez dira zuzenean itzuliko, adierazten dituzten objektuen itzulpena lortzeko balio badute ere.

Azken fase honetan lortutako emaitza hau da:

ESPECIFICACION

ENTRADA: Sea S secuencia_de ent definida

LONG(S)=20

SALIDA: Calcular sek77 SUBSEC_DE S formado_por_los I1 de S

tal_que I1 mod 2 = 0

Calcular elem70 ent tal_que

LONG(sek77)=elem70

FINESP

3. AZTERKETA ETA KOKAPEN BIBLIOGRAFIKOA

3.1. SARRERA.

3.2. SARRERA HISTORIKO LABURRA ETA EGUNGO LENGOIA NATURALAREN ANALISATZAILEEN JOERAK.

3.3. ATNetan OINARRITUTAKO ANALISATZAILEEN AZKEN GARAPENAK.

3.3.1. Ezagumendu semantikoaren erabilpena.

3.3.2. "Barrutik-kanporanzko" analisi-estrategia.

3.3.3. ATNak turrustan.

3.3.4. Konjuntzioen tratamendua

3.4. "IRAKASKUNTZA" ETA "PROBLEMEN ENUNTZIATUEN ULERMENA" DIRELAKO EREMUETAN ERABILITAKO LENGOIA NATURALAREN ULERMEN-SISTEMAK

3.4.1. Ezaugarriak

3.4.2. SOPHIE.

3.4.3. Mellish-en analisatzailea

3.4.4. Gomez-en programazioari buruzko
problemen ebazpen-sistema

3.5. KOKAPEN BIBLIOGRAFIKOA

3. AZTERKETA ETA KOKAPEN BIBLIOGRAFIKOA

3.1. SARRERA

Bibliografiari eskeinitako kapitulu hau gramatiken definizio eta Lengoia Naturalaren analisatzaileen arloan, gaur egun dauden ikerkuntz adar nagusien deskribapenarekin hasten da. Ikerkuntz eremu honen sorrerari buruzko irazkin labur baten ondoren, batetik bere adar nagusiak azpimarratzen dira eta, bestetik, sorrera eragin zuten arrazoi historikoak.

Lengoiaren analisia inplementatzerakoan Iragaite-Sare Hedatuen (ATNen) formalismoa erabili da. Honez gain, informazio semantikoaren maneiurako ATNen hedapen bat ere garatu da, izan ere, ugariak bait dira ebatzi beharreko problema linguistikoak, besteak beste, elipsi eta erreferentzien determinazioak (maiz gertatzen direnak darabilgun eremuan, Programaziozko problemen enuntziatuetan). Diseinatutako ulermen-modulua Konputagailuz Lagunduriko Irakaskuntza Adimentsurako sistema baten zatia da.

Ondoko bi ataletan (3 eta 4) CAPRATE sisteman aukeratutako analisatzaile-mota eta aplikazio-eremuen azken aurrerapenak sakonago aztertzen dira.

ATNetan oinarritutako analisatzaileen azken garapenek aurre egiten diote zenbait problemari: ezagumendu semantikoaren ezarpena, "ezker-eskuin" estrategia klasikoa ez bezalako beste estrategien inplementazioa, ezagumendu-maila deberdinei lotutako ATN desberdinak turrustan koordinatzea eta konjuntzioen tratamendua.

Laugarren atalean aplikazio-domeinu honetan erabat adierazgarri diren hiru sistema aurkezten dira. SOPHIE sistema Konputagailuz Lagunduriko Irakaskuntza Adimentsuaren adibide klasikoa da, non sistema eta ikaslearen arteko elkarrizketetan elipsiak eta erreferentzia osagabeak etengabeki erabiltzen bait dira, garrantzi berezia emanez ezagumendu semantikoari analisisan eta ebazpenean. Mellish-en sistemak mekanikari buruzko problema-enuntziatuak analisatzen ditu, ezagumendu semantikoa ahalik eta azkarren erabiliz eta deskribatutako objektuen

gehikuntzazko adierazpena hautatuz. Gomez-en programazio-problemen ebazpen-sistema ez da enuntziatuen ulerkuntzara mugatzen, ebazpenerako programa bat sortzen bait du. Domeinua CAPRATE-n erabilitakoaren berdintsua da, baina mota jakin bateko gestio-fitxategien tratamendura mugatzen da.

3.2. SARRERA HISTORIKO LABURRA ETA EGUNGO LENGOAIA NATURALAREN ANALISATZAILEEN JOERAK.

Analizatzaileen lehen belaunaldia Itzulpen Automatikorako lehenengo sistemekin batera jaio zen. Oinarrizko sistema haiek sintaktikoak ziren nagusiki. Perpausen azaleko egitura aztertzen zuten, lexiko-sarrera batzuentzat bereziki eraturako interpretazio-prozedurak abiarazteko. Garai hartan teoria linguistikoak urriak eta elementalak ziren.

Chomsky-k [Chomsky 65] sortutako gramatika transformazionalak aurrerapausu nabarmena eragin zuten linguistikaren arloan. Gramatika hauek lengoaia bat formalki definitzeko aukera eskaintzen zuten, zuzenak ziren eta ez ziren perpausak metodikoki bereizi ahal izateko. Halaber, perpausen azaleko eta sakoneko egitura bereizterakoan eta bien arteko transformazio-erlazioak zehazterakoan, esanahiaren ulerkuntzarako bidea erakusten zuten. Baina problema latza ageri zen: gramatika hauek ez ziren konputagarriak.

Une hartan Woods-ek [Woods 70] ATN kontzeptua sortu zuen, inongo teoria linguistikotan oinarritu gabeko formalismoa, baina perpausen egitura eta erlazio estrukturalak analisatzeko gai zena. Horrez gain, ATNen ahalmen deskriptiboa eta gramatika transformazionalena maila berekoak ziren. ATNen abantailak hauek ziren: 1) Interpretatzaile eta gramatikaren arteko bereizketa zela eta, gramatikaren egileak ez zituen ezagutu behar inplementazioren ezaugarriak. 2) Oso ulergarriak ziren (batez ere diagrama moduan), azaleko egitura azaltzen bait zuten. 3) Erraztasun haundiak daude sakoneko egitura edo beste edozein adierazpen sortzeko. 4) Orokortasun eta erregulartasunak biltzeko ahalmena. 5) Eraginkortasuna egikaritzekoan. ATNetan oinarritutako sistemen gaur egungo joerak sakonean deskribatzen dira kapitulu honen

hurrengo atalean.

Geroago Marcus-ek [Marcus 80] gramatika transformazionalerako ezerrera bat lortu zuen (gramatikok eusten zituen "Teoria Standard"-en hedapen bat egin ondoren). Sortutako analisatzaileak estrategia determinista zerabilen. Analisia, inguru sintaktikoan, tarteko egitura okerrik eraiki gabe egin zitekeela iradokitzeraino iritsi zen Marcus, hau da, backtracking-aren beharrik gabe. Marcus-ek berak onartu zuen geroxeago zenbait kasutan nolabaiteko determinismoa lortu ahal izateko beharrezkoa zela ezagumendu semantikoa. Aitorpen honek baretu egin zituen laguntza semantikoa eduki arren sintaktikoak ziren sistemen eta analisi semantiko erabat deterministaren defendatzaileak.

Gaur egun erabiltzen den beste gramatika-mota bat GPSG-ak (Generalised Phrase Structure Grammars, Perpaus-egiturazko Gramatika Orokortuak) dira, zeinak aurreko urteetan Chomsky berak gutxieztutako PSG-etan (Perpaus-egiturazko Gramatikak) oinarritzen bait dira. Gazdar-ek sortutako formalismo honek perpaus-egituretan (izen sintagmak, aditz sintagmak, preposiziodun taldeak) murriztapenak definitzea ahalbidetzen du, era honetan ezagumendua azaleko edo sakoneko mailan indiferenteki erabiltzeko aukera eskeiniz. Gramatika transformazionalekin alderatuz, erantsitako murriztapenak sistema konputagarriak sortu dituzte, baita aldi berean, ezagumendu semantikoa konbinatzeko aukera eskaini ere.

Lengoaia Naturalaren analisian, bada beste ikerkuntz adarra, ezagumendu semantikoari garrantzi gehiago ematen dioten analisatzaileetara jotzen duena. Adar honetan zera defendatzen da, prozesaketa semantikoa ez dela sintaktikotik isolatu behar, ezagumendu semantikoak sintaktikoak baino garrantzi gehiago eduki lezakeela eta, puntako kasuetan, semantikak zuzendu behar duela analisia. Ikuspuntu honetatik erregulartasun sintaktikoak beste modu batetara adierazten dira; esate baterako, ez da ziurtatzen adjetiboak izenen atzetik doazenik, baizik eta aldarazleen kokapena objektuak adierazten dituzten zatien atzean izan behar dela. Kategoria sintaktikoen ordez kategoria semantikoak erabiltzen dira [Wilks 75 eta 78]. Riesbeck-ek [Riesbeck 75] ezaugarri semantikoetan oinarritutako analisatzailea

definitzen du. Ideia honen ondorengo garapenek, ELI eta CA [Birnbaum and Seldfrige 79] [Riesbeck and Schank 76] kasu, Marcus-en teoria lantzen dute ikuspuntu semantikotik. Horrez gain, "bottom-up" analisi semantikoa osatu egiten da "top-down" erako tratamendu semantikoa egin ahal izateko, "script" edo maila haundiko ezagumendu-egiturak erabiltzen dira lan honetan. Tresna hauek lengoaiaren perpaus bakoitza konparatzen dute testu-eredu tipifikatu batekin edo landutako domeinuari buruzko elkarrizketa batekin, helburuak anbiguitasuna ebaztea eta ezagunak omen diren informazio eliptikoak osatzea direlarik. Ezagumendu semantikoaren erabilera ezinbestekoa gertatzen da elipsi, erreferentzien determinazio eta kuantifikazioa bezalako problema linguistikoen ebazpenean [Sebright 81], [Mellish 85].

3.3. ATNetan OINARRITUTAKO ANALISATZAILEEN AZKEN GARAPENAK.

ATNak William Woods-ek [Woods 80] sortu zituen BTN-etan oinarrituta (BTN: oinarritzko iragaite sareak), zeintzuk egoera finituzko automata-multzoak bait dira. Automata hauek onartu egiten dute beste automata bat deitzea egoeren arteko iragaite-baldintza gisa, deiak errekurtsiboak ere izan daitezkeelarik. Woods-ek proposatutako aberasketa sarearen arku bakoitzean baldintza berezia eta ekintzak eranstea izan zen. Sare bakoitzari, ekintzen bidez maneiatzen diren erregistrozko multzo bat dagokio. Honek analisiaren azken emaitza osatuko duten egiturak eraikitzea ahalbidetzen du. Baldintza bereziek osagaien arteko komunztadura aztertzen dute, erregistroen edukinak erabiliz.

ATNetan oinarritutako ondorengo analisatzaileetan aldaketak sartu ziren, eraginkortasun handiagoa lortzearren. Besteak beste, aldaketa hauek dira aipagarrienak: arkuetan pisuak definitzea, (prozesamenduan lehentasunak ezartzeko) eta arku eskluienteak biltzea (horietako arku bat onartzeak beste gainontzekoak baztertzea suposatzen bait du).

ATNei buruzko ondorengo azpiataletan azkenaldian izandako garapen kontzeptualak azaltzen dira.

3.3.1. EZAGUMENDU SEMANTIKOAREN ERABILPENA.

ATN sintaktiko hutsak perpaus sintaktikoki zuzenak onartzen ditu, nahiz eta ez eduki inolako esanahirik. Sintaktikoki zuzen izan arren semantikoki okerra den Chomsky-ren esaldi hau jarri ohi da adibide gisa: "Koloregabeko asmo berdeek amorruez lo egiten zuten". Garbi dago, beraz, nolabaiteko tratamendu semantikoaren burutzearen beharra, honelako analisiak baztertu eta, gainera, esanahia determinatzeko.

Semantikarako zuzendutako lehen lana Woods-ek berak gauzatu zuen LUNAR projektuan [Woods 72], lengoia errealaren prozesamendurako sistemak eraiki zitezkeela garbi asko frogatuz. BURTON-ek ere lan bikaina egin zuen ICAI-ren SOPHIE sisteman [Burton 76], kapitulu honen azken atalean deskribatzen dena.

Kategoria semantiko-lexikalen definizio esker, eta informazio semantikoaren manejurako test eta ekintza bereziei esker sistema hauek murriztapen semantikoak maneia ditzakete. Test semantikoek ea aldarazleak eta aldarazten diren objektuak bateragarriak diren egiaztatzen dute. Analisatzaileek edukitzen dute zenbait hobekuntza ere, osotasuna eta aztertutako osagaien zentzu zuzena egiaztatzeko. Analisatu gabeko alternatibak berrordenatzeko gauza baldin bada analisatzailea, orduan arkuetan erabilitako ekintza semantikoek zuzen ditzakete berrordenazioak, alternatiben egokitasun semantikoaren arauera.

ATNen bidez inplementatutako analisatzaile semantiko hutsek ez dute kategoria sintaktikorik erabiltzen. Aplikazio domeinua oso murriztua denean soluzio oso eraginkorrak lor daitezke, nahiz eta domeinu batetik bestera aldatu nahi direnean osoki birdefinitu behar diren. Gramatika semantikoak sintaktikoak baino askoz haundiagoak izan ohi dira eta landu beharreko domeinua oso mugatua ez denean erabiltezinak gertatzen dira. Gramatika-mota honen adibide gisa Burton-en lana aipatu behar da [Burton 76], izan ere, lan hau izan bait zen kapitulu honen azken atalean deskribatutako SOPHIE sistemaren oinarria.

3.3.2. "BARRUTIK KANPORANZKO" ANALISI-ESTRATEGIA.

Testu idatzien prozesaketan aurkitzen ez diren zailtasun bereziak agertzen dira ahotsaren prozesaketan. Esaldi bakoitzaren hasiera eta bukaera energia txikiagoz emititzen dira, nahasketak sortzeko arriskua handia delarik. Segidako bi hitzen ahozkerak beste hitz desberdin batena bezalakoa izan daiteke (*que eso / queso*). Ingelesa bezalako beste hizkuntzetan bi hitz diferente berdin ahozka daitezke (*why choose / white shoes, bear / bare*). Zailtasunok, perpausak ezkerretik eskuinera prozesatu beharrean, perpausaren tarterik ulerkorrenetik abiatu eta analisia perpausoko hasiera eta bukaeraraino hedatzen duten analisatzaileen sorrera iradoki dute. Perpausaren zati prozesatuaren aurretik eta atzetik zeintzu elementu egon daitezkeen aurrikusteko erabil daiteke gramatika, eta honetaz baliatu ere zalantzazko elementuen analisia gidatzeko.

Mota honetako bi sistema eraiki zituzten BBN enpresan Bates-ek [Bates 75 eta 76] eta Woods-ek [Woods 76], ATN motako analisatzaileetan oinarrituz.

Bates-en sistemak perpausan lor daitezkeen analisi partzialak tauletan gordetzen ditu. Geroago, gramatika erabiliz, analisatzaileak aurrean dezake zeintzu kategoria lexikal diren eskuinetik edo ezkerretik analisi partzialak hedatzeko erabil daitezkeenak. Bi analisi partzialen artean hitz baten luzeraren antzeko hutsa ba dago, orduan hitzaren kategoria lexikala zehaztasun haundiz determina daiteke. Analisatzaileen kontrol-egitura erraz alda daiteke backtraking, bilaketa sekuentzial eta paraleloa konbinatuz esperimendatzea ahalbidetzeko.

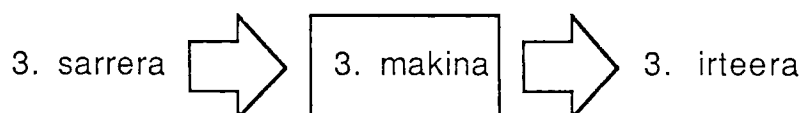
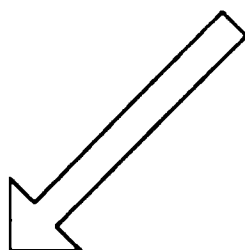
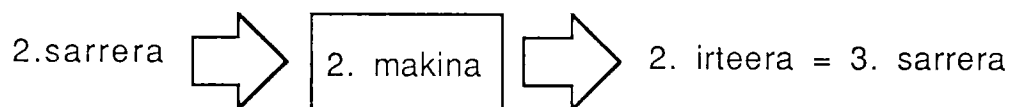
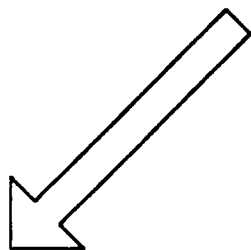
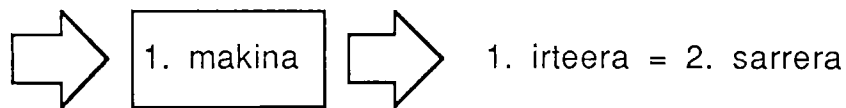
Woods-en sistemak, eraginkortasunari dagokionez, hobekuntza nabarmena suposatzen du Bates-en sistemarekiko. Gramatikaren aurreprozesaketa batean oinarritzen da, beronen bidez sareetako egoeren arteko erlazio-multzoa lortzen delarik. Edozein tarteko egoeretatik abiatuta, gramatika bi zentzuetan deskribatzen duten "array"-ak eraikitzen dira, ATNetako ezker-eskuin zentzu inplizituaren deskribapen hutsean geratu gabe. Gramatikaren egileak burutu behar duen aldaketa bakarra erregistroen edukinak erabiltzen dituzten arkuen ekintzak markatzea da, beren balioa zein

egoera edo egoeratan asigna zekiekeen espezifikatuz. Eraginkortasun-irabazpena eta aurrean segurtasuna kontutan hartuz gero, onerako aldaketa dela esan behar da.

3.3.3. ATNak TURRUSTAN.

ATNen sortzaileak berak [Woods 80] asmatu zituen ATNak turrustan. Analisi-maila desberdinei atxekitutako eta elkarren artean modu erraz eta ondo definituaz komunika daitezkeen ATNak eraikitzeke aukera eskaintzen du hobekuntza honek. Gramatika bakoitzari analisi-fase bat dagokio, adibidez: morfologia, sintaxia, semantika, pragmatika. 3.1. irudian 3 makina desberdinez osatutako turrusta ikus daiteke, horietako bakoitza ATN batean oinarritzen delarik. ATN kopuru finko bat dago bufferren bidez konektatuta (irudian 3 ATN). Sarrera ezkerretik eskuinera doa, makina analisatzaileen bektore honetan zehar. Makina baten irteera-bufferra bere eskuineko auzoaren sarrera-bufferra da. ATNen formalismoari TRANSMIT ekintza ere gehitu zaio. TRANSMIT-ek prozesuan dagoen makinaren irteera-buferrean idazten du bere argumentua eta kontrola eskuineko makinari pasatzen dio.

Bestalde, azken egoerara iritsi ezik prozesuan dagoen makinaren sarrera-bufferra hutsa den bakoitzean, kontrola ezkerreko makinari pasatzen zaio, honek balio berria sor dezan. Prozesu globalak bukaera arrakastatsua izango du baldin eta soilik makina bakoitza azken egoera batean badago eta balio guztiak onartuak izan ba dira iritsi diren sarrera-bufferretan.



3.1.irudia. Turrustan kokatutako ATNetan oinarritutako makinak.

Maila morfologiko, sintaktiko eta semantikoetan banatzen den analisatzaile bat turrustan eratuta dagoela suposa dezagun. Lehenengo hitza osatuta dagoenean maila morfologikora igortzen da. Honek analisia burutuko du, beheranzko estrategiaren bat erabiliz edo. Analisi morfologikoaren hipotesia konfirmatu egiten bada, emaitza maila sintaktikora bidaltzen da. Honek ere analisirako beheranzko estrategia erabil lezake. Maila sintaktiko eta morfologikoaren arteko koordinazioa inplizitua da eta kontrola erraza. Emaitza morfologikoa onargarria baldin bada maila sintaktikoarentzat, orduan honen sarrera-bufferra hutsik egongo da eta balio berri bat beharko du; beraz, kontrola maila morfologikora itzultzen da. Ebentualki, sintaxiak maila semantikora igortzeko

adinako infomazioa pila zezakeen, eta honek ere beheranzko estrategiaren bidez analisa lezake. Honela, analisirako posibilitateak sarrera-bufferrean baieztatu edo ezeztatu ditzakeen balioen bat baldin badago soilik hartuko dira kontutan, ez bestela. Ezeztatzen direnenan atzera jotzea edo backtracking-a beharrezkoa izango da.

Turrusten bidezko erakuntzari mailen arteko goranzko analisi-estrategia dagokio, baina ez du inolaz ere mailen barne-estrategia baldintzatzen. Edozein barne-estrategia hauta daiteke, beheranzkoa ere bai.

[Christaller and Metzger 83] lanean aurkezten den ezarrerak, hiru aldaketa egiten dizkio Woods-ek burutu zuen turrustan eratutako ATNei buruzko lanari, elkarrizketen analisisian malgutasun haundiagoa lortu asmoz.

Mailen arteko komunikazioa mezu-pasearen bidez gauzatzen da, eta ez da beharrezko maila guztiak ATNetan oinarrituak egon daitezkeen. Mezuen igorketa eta irakurketa parte hartzen duten prozesuak ATNen egoeren instantzi prozesuak diren edo ez jakin gabe buru daiteke. Christaller eta Metzger-ek ezarritako sisteman hiru analisi-maila desberdin bereizten dira. Maila sintaktikoa ATN bat bezala idatzita dago, maila semantikoa produkzio-sistema bat bezala eta diskurtso-maila ere ATNen ereduari jarraikiz eratuta dago. Azken honek hitzezko egintzen ereduaren (verbal act pattern) egitura definitzen du bi mintzaideentzat. Beraz, arkuen kategoriek elkarri bidaltzen dizkioten "hitzezko egintza" desberdinak jasoko dituzte. Adibidez, "informazio-eskari" egintza baten ondoren "informazio-emate" egintza bat etorriko da.

Beste aldaketa baten bidez backtracking itsua (kronologiko hutsa) ekiditzeko aukera lortzen du. Prozesu batek huts egiten duenean, zeukan informazio guztia mezu baten bidez pasatzen zaio kontrola hartzen duen prozesuari. Hortaz, prozesu hau errearen arrazoia aurkitzen eta, arrazoiaren arauera, aztertu gabeko arkuen antolaketa aldatzen saia daiteke. Sistemak arkuen sekuentziaren aldaketan hiru kasu desberdin aurrikusten ditu:

- Aztertu gabeko prozesuen instantzi kopuru edo denbora zehatz baterako soilik.

- Egoera zehatz batetik abiatuz prozesu-instantzia bakoitzerako, gainerako elkarrizketan ere erabilgarria.
- Egoera jakin baterako.

Christaller eta Metzinger-en hirugarren aldaketak maila batetik beste goiko maila batera pasatzea ahalbidetzen du tarteko mailarik zeharkatu gabe. Aldaketa hau berebizikoa gerta daiteke "mhm", "bai", "oh", "bada", ... bezalako hitzen azterketan, esannahi pragmatiko oso inportantea eduki arren analisi sintaktiko eta semantikoa (premiagabeko zaiena) dezente oztopatzen bait dituzte.

3.3.4. KONJUNTZIOEN TRATAMENDUA.

Nahiz eta ATNetan oinarritutako analisatzaileen ezaugarrietako bat orokortasun eta erregulartasun linguistikoak biltzeko erraztasuna den, formalismo honek problema latzak ditu egitura koordinatuak ebazterakoan. Orokorki, iragaite-sareetako ia egoera guztietatik abiatuta (WRD "y") arkuak beharko lirakeke maila bereko ia aurreko egoera guztietarako. Aldaketa honek (argitasunaren kaltetan) arku ugari eta era askotako test korapilotsuak sortuko lituzke. Gainera, kontutan hartu behar da egitura koordinatuen eraketan posible dela errepikatuta agertuko lirakekeen elementu linguistikoaren ezabaketa. Adibidez: "Europa (-ko herriak) eta Afrikako herriak", "Mehatzarien esku (*beltzak*) eta (*mehatzarien*) aurpegi beltzak". Geratzen diren fragmentuak osagai gramatikalak ez badira, konjuntzio laburtuen prozesamendua zaliagoa da. Adibidez: " ... oso- eta letra-sekuntziak".

Konjuntzioez hitz egiten denean, normalki adibide eta estudiogaitzat "eta" konjuntzioa ("and" edo "y" beste hizkuntzetarako) hartzen da, baina beste konjuntzioen tratamendua ere parekoa da.

ATNen inguruan konjuntzioen tratamenduari aurre egiteko hiru izan dira egin diren esfortzu nagusiak.

a) ATNen formalismoaren espezifikazioaren hedapena.

Blackwell-ek [Blackwell 81] Iragaitte-Sareetako bukaerako egoera guztietatik hasierako egoeretarako (WRD "and") arku simple bat eranstea proposatzen du, honela bigarren argumentu koordinatu baten analisia prestatuz. Horrez gain, bigarren elementua prozesatu aurretik konputazioa eteteko erraztasunak eskeintzen ditu, baita koordinazioan baztertutako elementuak bikoizteko ere. ATNa aldatzerakoan Zentzutasun (edo direkzionalitate) eta Ingurutasuna bezala definitutako murriztapenak errespetatu egin behar dira.

Murriztapen hauek lagungarriak gertatzen dira ezabaketa-kasuetan, elementu koordinatu batetik beste elementu koordinatuetara zeintzu erregistro eta zein zentzutan (eskerretik eskuina edo alderantziz) kopiatu behar diren erabakitzen laguntzen bait du. Osagaien ezabaketa ez da posible egitura koordinatuak berdinak ez badira.

Zentzutasunezko murriztapenaren arauera ezabaketak lehenengo elementu koordinatuko bukaeran eta bigarrenengo elementuko hasieran egin behar direla finkatzen da.

Ingurutasunezko murriztapenak zeintzu ezabaketa diren posible zehazten du, egitura koordinatuen korapilune nagusiarekiko korapilune osagaien auzatasunaren arauera. Adibidez: "Liburu berriak eta egunkariak" ezin daiteke ondorengo honen laburpena izan: "Liburu berriak eta egunkari berriak".

Sistema honen erakarpenik garrantzizkoena murriztapenen definizioan datza. Hala ere, sistema ez da eraginkorra erregistroen bikoizketen kopurua haundia bait da. Horrez gain, murriztapenen espezifikazioak gehiegitxo zaildu eta iluntzen du ATNen idazkera. Azkenik, estrategia hau ez da gauza konjuntzio laburtuak lantzeko.

b) ATNen interpretatzailearen espezifikazioaren hedapena

Woods-ek berak [Woods 73] eman zuen koordinazio-problemarako soluzio partzial bat. Honen arauera, "eta" konjuntzio bat aurkitzen denean interpretatzailea modu berezian

hasten da lanean. Presenteko prozesua bere horretan uzten du eta gainontzeko testuaren analisiari ekiten dio, modu ez-deterministaz hautatutako egoera batzuetatik abiatuz. Segitzen da testua analisatzen lehengo prozesuari jarraipena emateko aukera aurkitu arte. Hurbilketa hau konjuntzio laburtuen (naharoak LUNAR sisteman) tratamendurako diseinatu zen bereziki, baina oso inefikaza da konjuntzioen tratamendu orokorrerako, ez bait du modu deterministaz hautatzen koordinazio-aukera guztien artean.

Osagai semantikoak dituzten analisatzaileentzako hobekuntzaren bat iradoki zen [Boguraev 79]. Konjuntzioen anbiguetate-arazoen ebazpena, neurri haundi batean, ondorengo errutina semantikoaren esku utziz. Errutina hauek Izen Sintagmen (Konj Adj1 Adj2) edo Aditz Sintagmen (Konj IZ-S1 IZ-S2) zuzentasuna azter lezakete.

c) Arkuen eraiketa eta ebaluazio dinamikoak.

Kategoria berekoak ez diren osagaiak koordinaezinak direlako printzipioan oinarrituz, Boguraev-ek [Boguraev 83] gramatikaren definizioa apenas aldatu gabe, sistema bat eraiki zuen. Sistema honek konjuntzioak modu eraginkorrago batez aztertzea ahalbidetzen du, analisirako aukera probableenak baino ez bait ditu abiarazten. "Kategoria bera" terminoa kategoria lexikoei dagokie eta sintagma sintaktikoei ere bai.

Interpretatzaileak analisi-prozesuaren nondik-norakoa gordetzen du eta konjuntzioren bat agertzen denean esnatzen den "demon" bat sortzen du. Demon honek prozesu normala eten eta ordurarte ezagututako osagaiek dauzkaten erregistroak eta hurbileko historia ikertu ondoren lortzen duen informazioaren arauera ATN arku berri bat sortzen du dinamikoki. Arku hau, partzial edo osoki ezagutu den azken elementuaren kategoria bereko besteren bat ezagutzen saiatzen da, CAT edo PUSH motako arkua izan daitekeelarik. Arkuak huts egiten badu, ezagututako osagai globalago bati dagokion beste arku bat eraiki beharko da, eta honela segitzen da.

Demonak Blackwell-ek enuntziatu zituen bi murriztapenak biltzen ditu: Zentzutasuna eta Ingurutasuna. Beraz, zeintzu erregistro hasieratu behar dituen (jada ezagututako egituretatik

kopiatuz) eta zeintzu egitura bikoztu behar dituen, dinamikoki sortutako arkuaren iragaiten baten bukaeran, erabakitzen du. Mekanismo honek elementu linguistikoaren ezabaketa erabil dezake, bai konjuntzioaren lehenengo osagaikoa, bai bigarrenagokoa, baita bietan dagoeneko ere. Adibidez:

- "Etxeko aulki zuriak eta (*etxeko aulki*) beltzak"
- "Gizonezko (*jantziak*) eta emakumezko jantziak"
- "Taldea honetako Durango (*-ko ikasleak*)
eta (*taldea honetako*) Tolosako ikasleak"

Eraiketa dinamikoa hau iragaiten sintaktiko zuzen bat aurkitu arte luza daiteke, baina tratamendu semantikoa, kasuen batean, aurrekoa balio gabe utzi eta bilaketa-prozesuari berrekin diezaioke. Esate baterako: "Pato a la naranja y vino".

Lehenengo analisi zuzenean lortzen den "naranja y vino" balio gabe utziko da semantikoki eta bilaketa berri bati ekingo zaio.

Mekanismo honek premiazko du beste demon gehigarri bat definitzea konjuntzio laburtuen tratamendurako. Konjuntzio hauek, hain zuzen, era espezifiko batean lantzen zituen Woods-en sistemak. Boguraev-enak, aldiz, ez du interpretatzailearen portaera normala konplikatzeko, bakarrik ATNa aldatzen du dinamikoki konjuntzioaren bat agertzen denean.

Azken sistema hau eraginkorragoa da interpretatzaileak darabilen prozesamenduaren historiari esker. Bere helburua osagai posibleen artean bilaketa determinista gidatzea da, honela ugaltze konbinatorioaren arriskua gutxitzen delarik.

Blackwell-en sistemarekin alderatuz, Boguraev-ena errazago definitu eta egikaritzen da, konjuntzioaren sintaxiaren definizioa gramatikan bertan azaldu behar ez delako. Zentzutasun eta Ingurutasun printzipioak eta gramatika formal bat aski dira bere prozesamendurako.

3.4. "IRAKASKUNTZA" ETA "PROBLEMEN ENUNTZIATUEN ULERMENA" DIRELAKO EREMUETAN ERABILITAKO LENGOAIA NATURALAREN ULERMEN-SISTEMAK.

3.4.1. Ezaugarriak. Elkarren arteko arazoak.

Konputagailuz Lagunduriko Irakaskuntza Adimentsu (KLIA edo ICAI ingelesez) izenari "adimentsu" adjetiboa eransten diogu arlo honetako sistemek problemak edo gaiak ulertzeko duten ahalmenagatik eta erabiltzaileekin Lengoaia Naturelez komunikatzen direlako.

Ikerkuntz arlo honetan arrakasta lortu duten sistemek eta, salbuespenak salbuespen, ondorengo berezitasunak dituzte:

- Aplikazio eremuak teknikoak edo zientifikoak dira, beren formalizazioa errezagoa eta zehatzagoa bilakatzen delako.

- Erabilitako testuak bi mota desberdinetan sailka daitezke. Alde batetik objektu askoren deskribapen estatikoak eta beraien arteko erlazioak, eta bestetik gertaerezko sekuentziak. Sekuentzia hauetako ulerkuntza gertaera desberdinen arteko denborazko eta kausazko erlazioak aurkitzean datza.

Era bateko edo besteko testuak aukeratzeak agertuko diren problema-mota desberdinak sortaraziko ditu, eta baita berauek ebazteko erabiliko diren tresnak ere. Lehen erako testuetaz arduratzen bagara, ez da behar denboraren adierazpide sofistikatorik, ezta kausalitate- edo prezedentzia-azterketarik ere. Arazo nagusia analisi eta objektuen deskribapenen ulerkuntzan kokatuko da. Ikuspuntu linguistikotik begiratuta analisatzaile onak behar dira izen-sintagmak tratatzeko, hots, izenak, izenordeak, zenbatzaileak eta aldarazleak. Bestalde, testu baten barruan objektu erreal batekiko erreferentziak asko badira, ez dira beti egiten deskribapen osoa erabiliz, beraz, garrantzizkoa izango da erreferentzi ez-osoen tratamendua. Behin baino gehiagotan erreferentzia gainaditua geldituko da, beharrezko bihurtuz elipsiaren tratamendurako tresnen erabilpena.

Aurkezten dugun lana Novakek [Novak 76] definitzen duenez Lengoia Naturalaren "*natural language problem solver*" edo "Lengoia Naturelezko problemen ulerkuntza eta ebazpena" arloan kokatzen da. Eremu honetako lehen urratsak ondorengo ikertzaileek eman zituzten : Charniak-ek [Charniak 68 eta 72] ,Bobrow-ek [Bobrow 68], Heidorn-ek [Heidorn 72], Hayes eta Simon-ek [Hayes and Simon 74], Novak-ek [Novak 76] eta Ginsparg-ek [Ginsparg 78]. Gehienek nagusiki sintaktikoa zen egitura bat eraikitzen zuten aurretik, egitura hau lortu ondoren tratamendu semantikoari ekiten zioten barne-adierazpidea lortzeko.

Ondorengo ataletan interesatzen zaizkigun eremu honetan kokatzen diren hiru sistema deskribatzen dira: Brown eta Burton-ek [Brown eta Burton 75] sortutako KLIako SOPHIE sistema eta Mellish eta Gomezen problemen ulermenerako sistemak.

3.4.2. SOPHIE.

SOPHIE sistema Brown eta Burton-ek [Brown eta Burton 75] garatu zuten. Sistema honek diagnostikatzen eta konpontzen laguntzen dio ikasleari transistorez eratutako zirkuito elektronikoko batean osagai bat gaizki dagoenean.

Ikasleak ondorengo ekintzak burutu ahal ditu :

- Voltaiak, intentsitateak eta erresistentziak neurtu.
- Galdetu neurketa hauen ondorioetatik.
- Galdetu zirkuitoan aldaketa hipotetikoak egin ezkeroraino zein ondorio emango ligukeen.
- Laguntza eskatu diagnostiko-estrategia bat antolatzeko, orain arte jasotako datuetan oinarrituz.

SOPHIEk ikaslearen ekintzak exekutatzuz gain, inizatiba har dezake edozein unetan :

- ikaslea kuraiatzeko hipotesi bat asma dezan, orain arte hartu dituen neurriak azaltzearen. Geroago sistemak erakutsiko dizkio ikasleari hipotesi horrek sortzen dituen kontraesanak (horrelakorik badago).
- ikasleak eskatu duen neurri baten meritua ebaluatzeko (neurriak ekarriko dituen informazio berriak kontutan hartuz).

Nahiz eta ez eduki erregeletan oinarritutako inongo inferentzi-mekanismorik, eman dituen emaitza onakatik jarrera adimentsuak itxuratzen denez, SOPHIE lehenengotariko sistema aditutzat hartu da. Inferentzi-mekanismoa estrategia heuristikoetan oinarritzen da zirkuitoen simulaziozko ereduak maneiatzeko eta neurketa desberdinak, ondorioak eta hipotesiak prozesatzeko prozedura bereziekin.

Ezagumenduaren errepresentazioa sare semantikoak erabiliz osatzen da 50 kontzeptu abstraktu bilduz.

Lengoaia Naturalaren prozesatzaileak, ikasle eta sistemaren arteko elkarrizketan, erosotasun eta eraginkortasun handienak lortzea du helburutzat. Bilatutako nolakotasun hauek eta elkarrizketetan erabiltzen den lengoaiari lotutako ezaugarriek eragin handia dute diseinuan.

Inplementazioa gramatika semantikoen kontzeptuan oinarritu da, "neurria", "zirkuitoaren elementua", "transistoreak", "hipotesiak" eta abar kategoria kontzeptualak maneiatuz. Gramatika erregela bat dago kontzeptu bakoitza adierazteko, bere kontzeptu eratzailleak erabiliz. Erregela bakoitzean esplizitoki azaltzen da zer egin behar den kontzeptuaren azpikontzeptu eratzaille bat agertzen ez bada edo izenorde batez adierazita badator.

Behin gramatikaren erregelak formalizatuak izan direnean eta eraginkortasun handiagoa lortzeko, gramatikaren erregela bakoitza LISP funtzio bat legez kodetzen da (eskuz !).

Kodeketa prozesu honek beheranzko kontrol-egitura bat asignatzen dio gramatikari eta erregela bakoitzaren aukera desberdinen aplikazio-ordena zehazten du. Ondorio bezala LISP funtzio-bilduma bat lortzen da. Sortutako analisatzailea aplikazio honetan baino ezin da erabili.

Ikaslearen esaldi batetarako, analisatzaileak itzultzen duen emaitza LISPez idatzitako programa bat izango da, hain zuzen ere, erantzuna lortzeko prozedura berezi bati egindako dei bat. Analisisatzailearen emaitza lengoaia-sorkuntzarako errutinek erabiliko dute erantzunaren egitura eratzeke.

Gramatika semantikoak erabiltzen direnean, analisi sintaktiko-morfologikoa ez da hertsia-hertsia, hitz batzu salta daitezke esaldiaren gainontzeko ulergarria bada. Edozelan ere, SOPHIEk erizpide bi definitzen du zenbat hitz pasa dezakeen jakiteko. Alde batetik, gramatikaren erregela bakoitzak zenbat hitz salta dezakeen adierazten du. Bestetik, erregela bakoitzak informazio esplizittoa dakar gainaditu izan daitezkeen bere eratzaileak zehaztatzeko.

Ikaslearekin elkarrizketa "eroso" bat lortzeko, elipsia eta erreferentziak ebatzi behar dira. Informazio semantikoaz gain, elkarrizketaren testuingurua ere kontutan hartzen da, ikasketa-saioan sistemak ikaslearekin izan dituen elkarreragiketen historiaren bidez. Historia honek erreferente posibleen zerrenda ematen digu, ondoren zerrenda hori alderantzizko ordenan aztertzen da kategoria semantiko egokiena duen objektu aukeratzeko. Objektu eta kontzeptuen arteko anbiguitasuna maneiatzeko gai diren zenbait prozedura bereziek aztertzen dute zerrenda. Adibidez, Q2 transistore edo zirkuitoko puntu gisa izan daiteke hartua. Testuingurua tratatzen duen mekanismoak zirkuitoko puntu bat bilatzen badu, Q2 erreferentea onartuko da zirkuitoko puntu gisa erabilia izan bada eta ez transistore gisa izan den kasuetan.

Objektu batekiko erreferentzian bere osagarriren baten elipsia gertatzen bada, kontzeptu bakoitzari buruzko informazio esplizittoa erabiliz ebatzen da. Informazio honek adierazten digu zeintzu diren eratzaile elipsagarriak eta kasu bakoitzean zein den

egikaritu behar den tratamendua. Eratzailea elipsagarria ez bada eta hala ere agertzen ez bada, analisi posible hori baztertu egiten da.

Ondoren agertzen den bezala, ia esaldi osoko elipsiak ere tratatu egiten dira :

What is the voltage at Node 5? *Zein da 5. nodoaren voltaia?*
At node 1? *1. nodoan?*

Elipsi hau ondoz-ondoko esaldi bi nahiko antzekoak direnean gertatzen da. Bietako bigarrenean, bien arteko desberdintasuna baino ez da adierazten. Erregela bat dago gramatika esaldi eliptikoetarako. Lehenengoz aztertzen ari garen kontzeptuaren kategoria identifikatzen du. Orduan testuingurua lantzen duen mekanismoak ez du objektu erreferente bat bilatzen, aurretik erabilitako funtzio bat baizik. Funtzio hau aplikagarria izan behar da aztertzen ari garen kategoriarentzat. Horrelako funtzio bat aurkitzen bada, goian ageritako esaldia hartu egiten da argumentu zaharraren ordez, elipsiaz eraikitako esaldian ageritako kontzeptu berria ipiniz.

Sistemak tratatzen dituen beste konplexutasun batzu, ikasleak galdetutako ondorengo galderak izan daitezke :

What is the output? *Zein da irteera?*
Irteerako voltaia (output voltage) buruzko erreferentzi ezosoa.

What is it in a working instrument?
Zein da tresna egoki batean?
Izenordainaren erabilera, aurreko esaldian aipatutako voltaia buruzko erreferentzia.

What is the current thru R22 when it is shorted?
Zein da korronea R22tik zehar "bera" kortozirkuitatua dagoenean?
Izenordainaren erabilera, esaldi berean aipatutako R22 erresistentziari buruzko erreferentzia.

SOPHIEk ez ditu erlatibozko esaldiak tratatzen, ez zenbatzaileak, ez konjuntzioak. Muga hauetaz gain, aipatu behar dira eraginkortasuna eta gramatika semantikoen bilatzen dituzten

sistemenak: analisatzailearen definizioa nahiko korapilatsua da, ezagumendu informatiko sakona behar delarik bera maneiatzeko.

Sistema hedatzeko zaila da beste aplikazio-eremuetarako, analisatzaile linguistikoa, simulazio-modulua eta prozedura bereziak berreraiki behar baitira.

3.4.3. Mellishen analisatzailea.

Mellishek sistema hau [Mellish 85] Edinburgoko Unibertsitatean garatu zuen 1978-1981 tartean, sistema honek ingelesez enuntziatutako Mekanikari buruzko problemak ulertzen ditu Ulermen-mailan eraginkortasun handiena bilatzen saiatzen da, horretarako "anlisi semantiko goiztiarra" eta izen-sintagmekiko erreferentzien "gehikuntzazko ebaluazioa" erabiliz.

"Anlisi semantiko goiztiarrak", analisi sintaktiko-semantiko klasikotik urrunduz, ez dago zain testu osoaren analisi sintaktikoa burutu arte, ez eta esaldi bakar batena ere, analisi semantikoa martxan jartzeko. Anlisi semantiko goiztiarra testuko zati baten esannahia sortzen saiatzen da irakurtzen ari den unean eta ez du uzten geroko beste prozesu desberdin gisa. Eraginkortasun-mailan lortzen dugun aurrerapena ondorengo bi faktoreetan oinarritzen da :

- 1) Bilaketa sintaktikoaren espazioa laister txikitzen da zentzu gabeko egiturak baztertuz.
- 2) Egitura sintaktiko konplexuak sortzeko beharrik ez da ikusten, ez eta geroago bere osagaiak behatzeko funtzio bereziak erabiltzeko ere.

Estrategia honek arazoak sor ditzake esaldiaren osagai bat ulertzerakoan, batzutan ondoren agertzen den testu-zati batekin lotuta baitago, eta ezinezkoa baitaiteke bere esannahi osoa zehaztea geroko zatia ezagutu arte.

Testua perpausez osatua dagoela suposatzen da eta perpausak izen eta aditz-sintagmez. Testu osoaren esannahia bere sintagma

guztien esannahien konposaketan datza. Perpaua batek munduarekiko proposamen bat adierazten du eta izen-sintagma batek objektu bat edo objektu-multzo bat deskribatzen du. Izen-sintagmen ulerkuntza nahiko garrantzizkoa da testuko objektuekiko erreferentziak ondo burutu behar baitira.

Sistema honek urrats bi ematen ditu izen-sintagmak aztertzerakoan. Lehenengoa izen-sintagma mailan burutzen da hitzen ezaugarriak eta berauek nola konbinatzen diren jasoz. Urrats honetako tratamendu semantikoa markatzaile semantikoaren arteko komunztadurak egiaztatu baino ez da. Bigarren urratsa erreferenteen mailan egiten da. Informazio "kuasisintaktikotik" abiatuz, izen-sintagmak erreferentziatzen duen objektu edo objektu-multzoa ezagutzen saiatzen da. Aztertutako perpausek izen nagusiak edo izen-sintagma determinatu singularrak dauzkatenean kasuetan ez da diferentziarik egongo bi mailen artean. Izen-sintagma indeterminatuak edo pluralak baldin badauzkate diferentziak nabariak izango dira. Mellish-ek, nahiz eta kasu batzutan prozesua bitan banatzen den, erreferenteen zehaztapenean ahalmen eta argitasun handiagoa lortzen dela esaten du.

Izen-sintagmen analisi semantiko goiztiarrarekin azaltzen den arazoa zera da: erreferenteekiko informazio osoa ez dela beti posible eskuratzea testuinguru osoa prozesu aurretik, konkretuki ondorengo kasuetan gartatzen da :

- a) Izen determinatuen eta izenordainen ulerkuntzan, tinkotasun orokorrezko erizpideak garrantzizkoak baitira erreferente posibleak mugatzeko. Adibidez :

Estatu Batuetako lehendakaria...(1984an)

- b) Izen-sintagma indeterminatuen ulerkuntzan, kardinalitatearen informazioa ezin baita zehaztu ondorengo izen-sintagma batzu irakurri arte. Adibidez :

M masako zenbait bloke txiki barra zurrun baten erdian eta ertzetan daude finkatuta.

(=> 3 bloke)

Mellish-ek analisi semantiko goiztiarraren ideiei euste-arren, kasu horietan agertzen den informazio partziala errepresentatzeko tresnak eraikitzen ditu. Erreferenteak izen-sintagmeen bidez soilki zehazten direla esan beharrean, tinkotasun-baldintza orokorren bidez egiten dela proposatzen du, izen-sintagmek hauetariko batzu sartzen dituztela baina ez guztiak. Eremuaren eredian ikur berezien erabilpena proposatzen du azpiespezifikatutako objektu-mota desberdinak errepresentatzeko. Espezifikazioak autoaberastu egiten dira partzialki "erreferentzi-ebaluaketa" eragiketaren bidez. Eragiketa hau "murriztapen-beteketaren prozesua" bezala definitzen da, eta baita partzialki ere kardinalitate eta esparruarekiko "menpekotasun informazioa" komunikatuz. Menpekotasunezko zerrenden nozioa zenbatzaileen esparru eta multzoen informazioa errepresentatzeko sartzen da. Menpekotasunezko zerrenden gainean aplikatzen diren "matching" moduko eragiketen bidez multzoen kardinalitatearen eta zenbatzaileen esparruaren informazioa hedatu egiten du Mellish-ek.

Lana ez da amaitzen ezagumendua partzialaren adierazpide berri bat proposatzean, behar-beharrezko den bezala, proposatutako errepresentazioa maneiatzeko algoritmo egokiak ba daudela zihurtatu ere egiten da eta.

3.4.4. Gomezen programazioari buruzko problemen ebazpen-sistema.

Fernando Gomezek sistema hau [Gomez 81] Ohio Estatuko Unibertsitatean garatu zuen. Sistema honek lengoia naturalen adierazita datozen bankoetxeetako kontuen gestiorako problemak, nominak eta azterketetako notak ulertu eta ebatzi egiten ditu. Ebazpena PL/1-eko programa baten sintesian datza.

Programa-sortzaileak analisatzaileak ezagutu dituen kontzeptuen izenez osatutako zerrenda bat jasotzen du sarrera gisa. Oro har, kontabilitatea gestionatzeko edozein problema ondorengo urratsetan banatzen da : 1) Kontuak irakurri ,2) Salerosketak irakurri eta prozesatu eta 3) Emaitzak inprimatu.

Urrats bakoitzarentzat berau nola burutu behar den dakien "prozedura berezilari" bat definitzen da, mota desberdinetako problemak ebazteko prozedura ugari daudelarik. Adibidez,

READ-ACCOUNTS berezilariak kontuek definitzen dituzten balioak hartzen ditu; ondoren balio horien tinkotasuna egiaztatzen du (hots, ea identifikazio motarik erabili den ala ez aztertzen du kontuaren zenbaki, izena, ... eta ea bere mota zenbakia, karakterea,..... den); eta azkenean kontuen "datu-amaiera" balioa hartzen du bere tinkotasuna eta presentzia egiaztatzeko. Hiru "berezilarietatik" gain, planifikatzaile (PLANNER) bat dago exekuzioa koordinatzeko eta berezilari bakoitzak dagokion lana egin dezan.

Baina, lan honetan gehien interesatzen zaigun osagaia enuntziatuen ulermena da. Analisi-prozesuan goi-mailako ezagumendu iturriek lehentasuna hartzen diete behe-mailako ezagumendu linguistikoei. Beharrezkotzat jotzen da behe-mailako analisatzaile bat konjuntzioen esanahia, erabilpen anitzetako aditzena eta izen-sintagma zehazteko. Baina analisatzaile honen funtzioa ez da emaitza bat sortzea ondoren errutina semantikoak interpreta dezan, baizik eta, eremuarekiko informazioak bideratuko duen analisiari hasiera ematea, kontzeptu aipagarri bat ezagutu arte segituz. Une honetan kontzeptuari atxekitako prozedura espezialista batek hartzen du analisiaren kontrola, itzalean utziz behe-mailako ezagumendu linguistikoa. Prozedura espezialista hauek ondorengo analisia gidatzen dute, programazioari buruzko problema definitzen duten kontzeptu aipagarrien multzoa emaitza bezala aurkeztuz.

Ezagumenduaren errepresentazioa prozedurala da. Kontzeptu bakoitza prozedura berezilari bat legez errepresentatzen da, nola trata daitekeen adierazten duelarik. Beste alde batetik, enuntziatuen ulerkuntza frame "pasibo" gisa errepresentatzen diren enuntziatu-ereduen bidez zuzentzen da. "Pasibo" adjektiboa erabiltzen dugu, framearen edukina prozesu osoan aldatzen ez delako. Horietariko bakoitzak problemarentzako aipagarrien diren kontzeptu-zerrenda dauka, hala nola, sarrera eta irteera ohizkoenak, eta abar. Testutik ateratako kontzeptu aipagarriak egitura etiketatuz bezala errepresentatzen dira beraien propietateak deskribatuz.

Behe-mailako ezagumendu linguistikoak ez du gramatika sintaktiko edo semantiko baten definizio zehatzik ulertzen. Tratamendu semantiko bakarra kategorien erabilpenean datza, izen-sintagmak non amaitzen diren heuristikoki erabakitzeko

erabiltzen direnak, aditz bat, esaldi bat edo konjuntzioak detektatzeko. Testuaren segmentazio arrunt batetik abiatuz, testuko kontzeptu esanguratsuekin lotuta dauden hitzak bilatzen dira segmentu bakoitzean. Izen-sintagma baten segmentazioa ondoko eran burutzen da: artikulua, adjetiboak, izenordeak eta izenak DESCR deskriptorearen markarekin daude karakterizatuak hiztegian. Artikulu, izenorde eta erakusleak STD, deskripzio berriaren hasera-marka daramate. Izenorde pertsonal eta inpersonalek gainera, ERU marka daramate erabilpen erreferentziala adierazteko. Hitz bat DESCR markarekin agertzen denean, izen-sintagma baten hasieran gaudela suposatzen da. Hitzak, gainera, ERU marka badauka amaitutzat ematen da. Bestelako kasuetan prozesua aurrera doa :

- 1) Hitz batek DESCR marka galdu arte.
- 2) Hitz batek STD edo ERU marka hartu arte.
- 3) Aurretik erabili den izen nagusia bat aurkitu arte (gutxienez bere izena erabili da aurretik).

3.5. KOKAPEN BIBLIOGRAFIKOA.

Zehaztasun orokorrak :

- Aurkezten dugun CAPRATE sistema ATNen formalismoetan oinarritzen da. Hedapen bat garatu da informazio semantikoa erabiltzeko, berorren helburua elipsia maneiatu eta erreferentziak zehaztea izanik. Problema hauek maiz agertzen dira objektuak deskribatzen dituzten testuak erabiltzen dituzten aplikazio-eremuetan; adibidez, problema-enuntziatuen ulermenean, matxurak konpontzeko eskuliburuak maneiatzerakoan eta abar. Informazio semantikoa ez da analisiaren gidari bakarra, Gomezen analisatzailean edo Burtonen gramatika semantikoan (SOPHIEn) gertatzen den bezala. Ezagumendu sintaktiko eta semantikoaren kooordinazioa, kategoria sintaktiko eta semantiko murriztuetako osagarriak biltzen dituzten eskaeren bidez egiten da. Eskaera hauek erabiliz, zenbait

analisi sintaktiko posible baztertu izango da bere analisi osoa bukatu baino lehenago. Diseinu-aukera honen arrazoa zera da: beste aplikazio-eremuetarako hedapenak errazak izan zitezten nahi genuela. Nagusiki semantiko diren analisatzaileak beste domeinuetara hedatzeko nahiko zailak direla kontutan izan behar dugu.

- Boguraev-ek proposatutako arkuen eraikuntza eta ebaluapen dinamikoa nahitaezko abiapuntu da konjuntzioak tratatzeko, sistemaren ondorengo garapenetarako.
- "Barrutik-kanporantz" analisi-estrategiak "eskuinetik-ezkerrera" estrategiaren definizioa iradokitzen du euskarazko perpausen analisisian. Hizkuntza honetaz idatzitako enuntziatuak etorkizun hurbilean aztertuko dira; baina lehen azterketa batean ezkerreko errekurtsibitatea agertzen da izen-lagunetan eta erlatibozko esaldietan, ezinezkoa bihurtuz "ezkerretik-eskuinera" strategiaren erabilpena analisisian.

Ondoren aintzinean, 3.4 apartatuan, deskribatutako hiru sistemekin konparatuko dugu gure sistema.

a) Konparaketa SOPHIEekin.

Gure sistemak ez du behar denborarekiko eraginkortasun handirik SOPHIE sistemak bezala, honek ikaslearekin denbora errealean lan egiten du eta. Arkuen arteko bersailkapena eta eksklusibotasuna, edo ATNen behe-mailako kodeaketa ez dira ikusten premia handiaz gure sisteman. Ondorengo berrikuntzetan aztertuak izan daitezke eraginkortasun handiagoa lortzeko.

Elipsiak eta erreferentziak ebazteko CAPRATEk definitu duen metodoa SOPHIE-renean oinarritzen da, hots, jasandako elkarrizketaren historian zehar kategoria semantiko mugatuzko objektuen bilaketan.

SOPHIEk ez ditu erlatibozko esaldiak tratatzen. Nahiko zaila da beste eremuetara garraiatzeko, batez ere bere diseinu prozedural eta semantikoarengatik.

b) Konparaketa Mellish-en analisatzailearekin.

CAPRATEk Mellish-ek garatutako gehikuntzazko analisi semantiko goiztiarraren ideiak jasotzen ditu, baina gurean tinkotasunezko baldintzak maila orokorrean soilki aplikatzen dira eta era murriztaile txikiagoan izen-sintagma eta aditz-sintagma mailan. Mellish-ek sartutako formalismoa objektu azpiespezifikatuak errepresentatu eta maneiatzeko egokiena dirudi zenbatzaileak eta izen-sintagma indeterminatu pluralak tratatzeko, baita ere erreferentzi ez-osoak ebazteko ere.

Mellish-en sisteman analisi linguistikoa formalismo logikoa erabiliz burutzen da, CAPRATEn bestalde ATNak erabiliz. Eskakizunen definizio eta kontrol-sistema ere CAPRATE-ren ezaugarria dira.

c) Konparaketa Gomez-en sistemarekin.

Gomez-ek sistema bat eraiki zuen ez bakarrik programazioko enuntziatuak ulertzeko, PL/1ean idatzitako programa ere sortzen baitzuen enuntziatua ebazteko. Baina, alde batetik, aplikazio-eremua (ohizko gestio-problema) murriztagoa da, eta bestetik, eremuarekiko menpekotasun handia duen eta oso prozedurala den definizioak nahiko zaila egiten du beste eremu desberdinetan aplikatzea.

Egiten duen analisi sintaktikoa ia-ia hutsa da: hitzen kategoria sintaktikoak maneiatzen dituzten heuristikoak erabiltzen dira testua segmentatzeko.

Nahiko interesgarria da beheranzko estrategiaren implementazioa. Estrategia honek enuntziatuen ulerkuntza gidatuko du domeinuaren ezagumendua hartuz abiapuntutzat. Ezagumendu honek enuntziatuaren egitura modelizatzen du, baita ezagutzen den elementu bakoitzarekin egin behar den tratamendua ere. Enuntziatuetan agertutako objektuen errepresentazioak ateratzeko CAPRATE-n implementatutako goranzko estrategia soila osatua izan daiteke beheranzko estrategia batekin funtzio, ekuazio edo espresio konplikatuen definizioen ulerkuntza gidatzeko. Ulerkuntza gidatzeko erabiltzen den aplikazio-eremuari buruzko ezagumendu

hau beheranzko estrategia bezala implementa liteke modulu analisatzaile berri baten bidez analisatzaile sintaktiko-
-semantikoarekin turrustan koordinatuz.

4. DISEINU OROKORRA ETA BESTE APLIKAZIOTARAKO TRESNA ERABILGARRIAK.

4.1. SISTEMAREN EGITURA ETA FUNTZIONAMENDUA.

4.2. BESTE APLIKAZIOTARAKO TRESNA ERABILGARRIAK.

4.2.1. ATNen interpretatzailea.

4.2.1.1. Ezaugarri orokorrak.

4.2.1.2. Arkuak.

4.2.1.3. Ekintzak.

4.2.1.4. Testak.

4.2.1.5. Forma ebaluagarriak.

4.2.1.6. Analisarako estrategia.

4.2.2. Hiztegiaren egitura.

4. DISEINU OROKORRA ETA BESTE APLIKAZIOTARAKO ERABILGARRI DIREN TRESNAK.

4.1. SISTEMAREN EGITURA ETA FUNTZIONAMENDUA.

CAPRATE sistemaren helburutzat Lengoaia Naturelez enuntziatutako problemen itzulpen automatikoa dugu, CAPRA projektuaren espezifikazio-lengoiaren bidez formalki adierazteko asmoz. Zehatzago esateko, esandako itzultzailearen sarrera lengoaia naturala da, eta irteera, enuntziatuari dagozkiokeen espezifikazio guztiez osaturik dago (baita espezifikazio bakoitza lortzeko erabili den analisi sintaktikoa ere). Kontutan hartu behar da enuntziatu batetik abiatuz zenbait interpretazio lor daitezkeela lengoiaren anbiguitasunaren ondorioz. Kontutan hartu ere ondo definitutako interpretazio bat era desberdin batzuek espezifika daitezkeela. CRAPATE-k enuntziatu problema batetarako lor daitezkeen interpretazio guztiak sortuko ditu, baina interpretazioetarako espezifikazio bana sortuz.

4.1. irudian sistemarako aukeratu den diseinua azaltzen da. Sistema itzultzaileak hiru osagai dauka, tratamenduaren ondoz-ondoko hiru faseekin lotuta daudelarik.

Lehenengo osagaia analizatzaile lexiko-morfologikoa da. Sarreran enuntziatu oso baten testua hartzen du (hau da, sistema itzultzaile orokorraren sarrera), eta emaitza bezala, enuntziatuan ageritako unitate lexikoek osatzen duten lista bat lortzen du. Orokorrean, unitate lexikoen osagaiak hiztegiko sarrerarako erreferentzia bat eta hiztegian ez dagoen (baina textuko hitzak dakartzan) informazio morfologikoak dira.

Bigarren osagaia analizatzaile sintaktiko-semantikoa da. Berau ATNen interpretatzaile, hiztegi eta aplikazio konkreturako definitutako ATNmultzoren bidez inplementaturik dago. Bai hiztegia, bai ATN sintaktiko-semantikoen multzoa birdefini daitezke sistema beste baldintetan erabili ahal izateko

interpretatzailea aldatu beharrik gabe (beste hizkuntzekin, beste esaldi-motekin edo beste aplikazio-eremuekin). Analisi fase honen sarrera lehenengo fasean lortutako unitate lexikalen lista da. Lortzen den emaitzak sintaktiko-semantikoki posible diren analisi guztiak dauzka. Beraiek adierazteko, alde batetik dagozkien zuhaitz sintaktikoak eta bestetik enuntziatuan deskribatzen diren eremuko objektuak errepresentatzen dituen asertzio-multzoa erabiltzen dira.

Hirugarren eta azken osagaia espezifikazioen sortzailea da. Fase honetan sarrera gisa aurreko fasean sortutako asertzio-multzo bat hartzen da eta CAPRA sisteman definitutako espezifikazio-lengoiaren bidez dagokion espezifikazioa lortzen da. Sorkuntz fase hau funtzionatzen has dadin, ez da itxaron behar aurreko faseak analisi sintaktiko-semantiko posible guztiak sortu arte, analisi posiblea bakoitza lortu ahala, espezifikazio-sortzailea martxan jartzen da bere espezifikazio formala lortzeko.

Kapitulu honetako bigarren atalean ATN beharrezkoen multzoa eta hiztegia nola eraiki daitezkeen azaltzen dira. Bostgarrenean, ordea, analisi-fase bakoitzean zer ezagumendu-mota behar den eta nola erabiltzen den aplikazio konkretu honetarako xeheki azaltzen da.

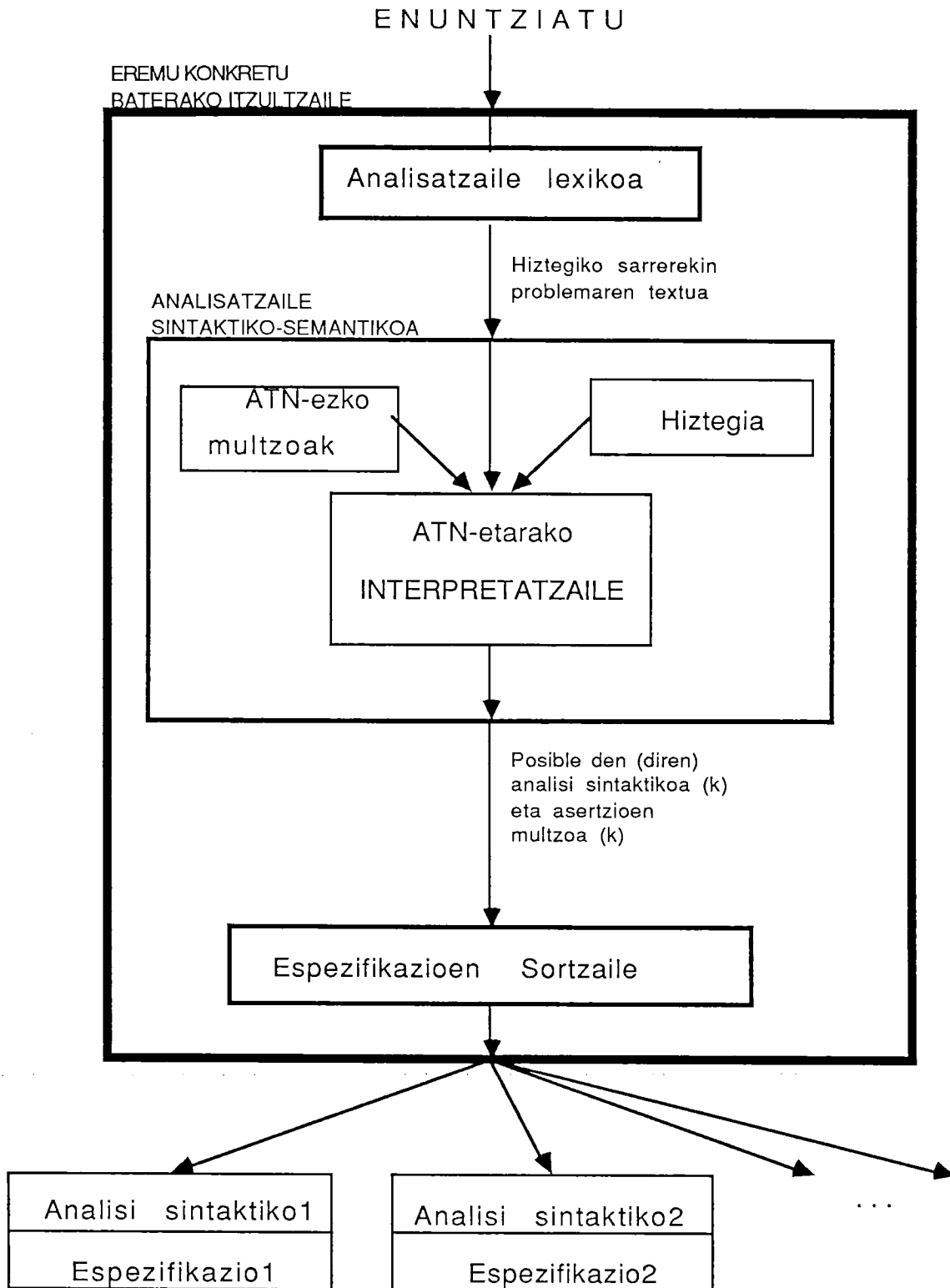


fig 4.1. Sistemaren diseinu globala

4.2. BESTE APLIKAZIOTARAKO TRESNA ERABILGARRIAK.

4.2.1. ATNen interpretatzailea.

4.2.1.1. Ezaugarri orokorrak.

Inplementatutako ATNen interpretatzailea Bates-ek [BATES 78] emandako definizio standard-a onar dezan diseinatua da. Horretaz gain, ezagumendu semantikoa maneiatzeko zenbait predikatu eta ekintzez zabaldu da. Ekintza berriek objektuen errepresentazioak eta beren propietate eta erlazioenak eraikitzeak aukera eskaintzen dute. Ekintzok, textuinguruko informazioa eta kategoria semantikoak erabiliz, erreferentziak zehaztatzeko ere balio dute.

Gehitzen diren zabalkuntza honek ondorengo ataletan deskribatzen dira. Bertan interpretatzaileak onartzen dituen arkuak, forma ebaluagarriak, testak eta ekintzak formalki espezifikatzen dira. Azkenik, interpretatzailearen analisirako estrategia deskribatzen da.

ATNa gutxienez korapilune bat duen korapilune-lista bezala definitzen da. Korapilunea bere izenaz eta bere baitatik ateratzen diren arkuez osaturiko lista da.

```
<atn> ::= [<korapilune> <korapilune>*]  
<korapilune> ::= [<korapilune_izena> <arkuak>]  
<arkuak> ::= <arku> | <arku> <arkuak>
```

4.2.1.2. Arkuak.

Bates-ek bere ereduan definitzen dituen arku-mota guztiak erabil daitezke sistema honetan.

```
<arku> ::= <cat_arku> | <word_arku> | <push_arku> | <pop_arku>  
| <jump_arku> | <vir_arku>
```

Arku-moten deskribapena ondokoa da:

<cat_arku> ::= (CAT <kategoria> <test> <ekintzak>)

CAT arku baten iragaketa posiblea da sarrerako lehenengo hitza <kategoria>-n sailkatuta badago eta <test> *egiazkoa* bezala ebaluatzen bada. Sarrerako testutik hitz bat jaten du.

<word_arku> ::= (WRD <hitz> <test> <ekintzak>)

WRD arku baten iragaketa posiblea da sarrerako lehenengo hitza eta <hitz> berdina badira eta <test> *egiazkoa* bezala ebaluatzen bada. Sarrerako testutik hitz bat jaten du.

<push_arku> ::= (PUSH <korapilune_izena> <test> <ekintzak>)
| (PUSH <korapilune_izena> <test> <aurrekintzak> <ekintzak>)

<test> *egiazkoa* bezala ebaluatzen bada, PUSH arkua <korapilune_izena> korapilunean hasten den ATNrako dei errekurtsiboa abiarazten du. Orduko ATNari lotutako erregistroak pilaratzen ditu. Beste ATNari egindako deia arrakastaz bukatzen bada, lehengo erregistroak berreskuratzen ditu eta PUSH arkua iragaten da.

<pop_arku> ::= (POP <forma> <test>)

ATN baten bukaera-korapilune bat definitzen du. Hartutako azken PUSH arkuari kontrola ematen dio, <forma> informazio bezala itzuliz prozesatutako emaitza bezala.

<jump_arku> ::= (JUMP <korapilune_izena> <test>)

<test> *egiazkoa* bezala ebaluatzen bada, besterik gabe <korapilune_izena> korapilune edo egoerarako iragaketa burutzen du.

<vir_arku > ::= (VIR <osagai> <test> <ekintzak>)

VIR arkua eta HOLD ekintza elkarrekin erabiltzen dira ATNen arteko edozein dei-mailatik erregistro globalen balioei atzitu ahal izateko. Kontutan hartu ATN bakoitzeko erregistroak bertakoak direla, beste ATNetatik ikustezinak direlarik.

4.2.1.3. Ekintzak.

Bates-en eredian definitzen diren oinarrizko ekintza guztiak erabili ahal dira sistema honetan. VERIFY, SETRQ eta SENDRQ ezik, aukerazko ekintza gehigarri guztiak ere inplementatu dira. VERIFY arku baten iragaketa abortatzeko aukera eskaintzen du, argumentu bezala laguntzen dion forma ebaluatzeko ezina denean. Bere Inplementazioa ez da zaila, baina ez dugu beharrezkotzat ikusi. SETRQ eta SENDRQ ekintzak ez ditugu inplementatu, ondoko modu alternatiboaz idatz daitezkeelako:

```
[SETR <erreg> [QUOTE <balio>]] eta  
[SENDR <erreg> [QUOTE <balio>]].
```

Sei ekintza berri erantsi dira. Beraietako bik (COND eta TO) ATNerako kontrol egiturak dira, eta beste lauek ezagumendu semantikoa erabiltzeko aukera eskaintzen dute.

COND ekintza (<test> <ekintza>) moduko bikote-sekuentzia batekin definitzen da. Testak sekuentzialki ebaluatzen dira, batetan *egiazkoa* emaitza lortu arte. Test honi dagokion ekintza exekutatu ondoren, bukatutzat ematen da COND ekintza osoaren exekuzioa. Test guztiak *faltsuak* badira ez da ekintzarik exekutatuko.

TO ekintza arkuaren azken ekintza bezala ageri behar da, JUMP eta POP motako arkuetan ezik (hauetan ez baita agertzen). Arkua arrakastaz gainditzen deneko kasuetan, zein korapilunera jauzi behar den zehazten du ekintza honek. Bates-en ATNen eredian ez da definitzen ekintza-mota gisa, JUMP eta POP ez diren beste arku-motetako azken osagai bezala baizik. Definizio formalak desberdinak izan arren, Bates-en ereduaren eta CAPRATE-ren erabilpenak baliokideak dira TO ekintzarako.

Ekintza semantikoek osagai sintaktikoen iragazketa semantikoa gidatzen dute. Unitate sintaktiko-semantikoen nukleoak aurkezten dituen eskaerak piztu, bete eta aztertuz lan egiten dute ekintza hauek. Tratamendu semantikoa osatzeko, lau ekintza hauetatik aparte, hurrengo apartatuan deskribatzen den USOSEMCOMPLETO testaren laguntza beharrezkoa da.

<ekintzak> ::= <ekintza> | < ekintza> <ekintzak>

<ekintza> ::= <setr_ekintza> | < to_ekintza> | <liftr_ekintza>
| <hold_ekintza> | < addr_ekintza> | < addl_ekintza>
| <cond_ekintza> | <waitsem_ekintza>
| <possem_ekintza> | <filtrar_ekintza>
| <tomar_modif_gen_ekintza>

<setr_ekintza> ::= (SETR <erregistro-izena> <forma>)
<forma>-ren balioa asignatzen dio <erregistro-izena>
erregistroari.

<to_ekintza> ::= (TO <korapilune_izena>)
JUMP eta POP motako arkuetan ezik (hauetan ez baita
agertzen). Arkua arrakastaz gainditzen deneko
kasuetan, zein korapilunera jauzi behar den zehazten
du.

<liftr_ekintza> ::= (LIFTR <erregistro-izena> <forma>)
Oraingo ATNari deitu dion ATNko <erregistro-izena>
erregistroari <forma>-ren balioa asignatzen dio.

<hold_ekintza> ::= (HOLD <osagai> <forma>)
<forma>-ren balioa <osagai> erregistro globalari
asignatzen dio. (ikus VIR arku-mota)

<addr_ekintza> ::= (ADDR <erregistro-izena> <forma>)

<addl_ekintza> ::= (ADDL <erregistro-izena> <forma>)

<erregistro-izena> erregistroaren edukia aldatzen du.
lehenengo edukia lista bat izan behar du. Balio berria
lortzeko, eskuinetik (addr) edo ezkerretik (addl)
<forma>-ren balioa eransten da.

<cond_ekintza> ::= (COND <alternatibak>)

<alternatibak> ::= (<test> <ekintzak>)

| (<test> <ekintzak>) <alternatibak>

permite la ejecución de acciones bajo condiciones

<waitsem_ekintza>::= (WAITSEM <osagarri> <kat_sintaktikoa>)
Osagarri sintaktiko-semanticoko bat berari dagokion unitatearen nukleoa baino lehenago agertzen denean erabiltzen da ekintza hau. Osagarria eta bere kategoria sintaktikoa gorde egiten dira nukleo ageri arte. Nukleo agertzen denean, iragazi egiten da gordetako osagarria. Lan hau burutzeko, "salaespera" erregistro berezia erabiltzen da.

<possem_ekintza>::= (POSSEM <unitate_lexikal>)
Ekintza honek unitate lexikal bati (normalean, hiztegiko sarrera bati) unitate sintaktiko-semanticoko bezala dagozkion definizio semanticoko guztiak aktibatzen ditu. Horrez gain, "salaespera" erregistro berezian itxaroten ari ziren osagarri sintaktikoak iragazten ditu definizio semanticoen arauera. Ekintza hau agertzekotan, arkuko azkenaurreko ekintza bezala ageri behar da. Murrizketa honen arrazoia zera da: POSSEM ekintzak analisi-egoera berri bat sortzen du hitz nukleoaren definizio semanticoko bakoitzerako, eta behar-beharrezkoa da arkuko beste ekintza guztiak analisi-egoera guztietan exekutatzea. Azken ekintza TO modukoa izango da.

<filtrar_ekintza>::= (FILTRAR <osagarri> <kat_sintaktikoa>)
Ekintza honek unitate sintaktiko-semanticokoaren nukleoa piztutako eskaeraren batekin (betebeharrekoa edo gehigarria) osagarria ea bateragarria den ala ez aztertzen du. Bateragarria bada analisi-egoera berri bat sortuko du, bestela ez du egoera berririk sortuko eta beraz, puntu horretaraino egindako analisisa bazterturik geratzen da. Eskaera bat eta osagarri bat bateragarriak dira baldin eta bion kategoria sintaktikoa eta semanticoko berdinak edo bateragarriak badira

<tomar_modif_gen_ekintza>::= (TOMAR_MODIF_GEN)
Ekintza honek ez du argumenturik. Osagarri gehigarri orokorrak xahutzen dituzten ATNei deitzen dieten

PUSH moduko arkuetan erabiltzen da ekintza hau. Kasu hauetan osagarriaren nukleoak zuzentzen du iragazketa. Osagarri gehigarri orokorra izango denaren nukleoak kategoria semantiko bat eskatuko du berak aldarazten duen nukleorako.

Propietate gehigarri orokorra aplikagarria baldin bada, (hots, aldarazitako nukleoa menpekoak egindako eskaerakin bateragarria bada eta menpekoen beste eskaera guztiak ere bete badira), interpretatzaileak bere ekintza semantikoak goiko ATNra igoko ditu (LIFTR modif_orok) ekintzaren bidez. Goiko ATNean dagoen (TOMAR_MODIF_GEN) ekintza "modif_orok" erregistroaren edikuna propietate gehigarrien listari gehitzeaz arduratzen da, unitate sintaktiko-semantikoaren analisia bukatutakoan igondako ekintzak ere exekuta daitezten.

<aurrekintzak> ::= <aurrekintza> | <aurrekintza> <aurrekintak>

<aurrekintza> ::= (SENDER <erregistro-izena> <forma>)

Deitutako sarearen <erregistro_izena> erregistroari <forma> balioa asignatzen dio. PUSH moduko arkuetan erabiltzen da.

4.2.1.4. Testak.

Bates-ek proposatutako oinarritzko test guztiak erabilgarriak dira sistema honetan. CATCHCHECK, ENDOFSENTENCE, x-AGREE eta x-START aukerazko testak ez dira inplementatu beharrezkotzat jo ez direlako. Beste aukerazko testez gain (NULLR, CHECKF, NIL, T, AND, OR eta NOT), beste hiru gehiago erantsi dira: GETF, GETFLISTA eta USOSEMCOMPLETO.

GETF eta GETFLISTA berez ez dira testak, forma ebaluagarriak baizik, baina test bezala erabil daitezke. Lortutako emaitza NIL ez bada "true", eta NIL bada "false" egi-balioak kontsideratzen dira.

USESEMCOMPLETO izaera semantikozko testa da, zeina unitate sintaktiko-semantikoei dagozkien ATNetako POP arkuetan ageriko bait da. Unitate sintaktiko-semantikoaren nukleoak piztutako betebeharreko osagarri-eskaera guztiak bete direla egiaztatzeko balio du. Hala balitz definizio semantikoari dagozkion ekintza semantikoak, baita agerturiko eskaera gehigarriei dagozkienak ere, exekutatuko lirateke. Baina betebeharreko eskaera guztiak ez badira bete, aurretik agerturiko objektuen artean kategoria semantiko berekoa eta bete ez den osagarrian espezifikatzen diren propietateak dituen objekturen bat bilatzen da. Bilaketa-erizpidea erraz alda daiteke. Inplementazio honetan erizpide sinplea baina nahiko eraginkorra hartu da. Erizpide hau, bilatzen den kategoria semantikokoa den azken objektu prozesatua hartzean datza. Bilaketa hau arrakastatsua bada eskaera betetzat jotzen da, dagozkion ekintza semantikoak exekutatu direlarik, bestela USESEMCOMPLETO testaren ebaluaketaren emaitza NIL izango da, eta beraz, segitutako analisia baztertu egingo da.

Prozesatzen ari den unitate sintaktiko-semantikoa nukleoak eskatzen dituenetakoa ez bada (hau da, eskaera gehigarri orokorra prozesatzen denean), USESEMCOMPLETO-k ez ditu dagozkion ekintza semantikoak exekutatu, baizik eta LIFTR ekintzaren bidez goiko mailara igotzen dituela. Informazio hau TOMAR_MODIF_GEN ekintzaren bidez berreskuratuko da goiko mailan.

Izen-sintagma baten analisisian, kategoria semantiko eta propietateak zehaztuta dituen objektu baten sorkuntz-ekintza semantikoa exekutatu baino lehen aldaraz daiteke, izen-sintagmaren determinazio eta zenbakiaren arauera. Aldaketa posibleak ondokoak dira:

- a) indeterminatu singularra: ez dago aldaketarik.
- b) indeterminatu plurala: objektua elementu bakun bati dagokion kategoria bada "sekuentzi deskribatzaile" kategoriako objektu bat sortuko da. Kategoria hau "objektu jeneriko" baten deskripzioa bezala interpretatu beharko litzateke, baina ez da beti sekuentzia izango.

- c) indeterminatu singularra: objektua aldaketarik gabe sortuko da, baina lehenago deskripzio hori betetzen duen objekturik ez dela existitzen egiaztatzen da. Objektu hori existituko balitz ez litzateke berri bat sortuko zeren eta aurreko objektuaren erreferentzia bat baino ez bait da.
- d) determinatu plurala: "indeterminatu plurala" kasuan bezalaxe "determinatu plurala"-ren kasuan "sekuentzi eskribatzaile determinatu-plurala" kategoriako objektu bat sortzen da, praktikan "sekuentzi deskribatzaile" delakoaren baliokidea delarik.

Ondoren interpretatzaileak erabiltzen dituen testak deskribitzen dira:

<test> ::= T | <tgetf12> | <tnullr> | <tgetr>
 | <tand > | <tor> | <tnot> | <tequal>
 | <tusosemcompleto>
 T-ek egiazkoa balioa adierazten du

<tget12> ::= (GETF <ezaugarri> <erregistro-izena>)
 |(GETF <ezaugarri>)
 ezaugarria definiturik badago, NIL ez den balio bat itzuliko du, zeina egiazkoa-tzat interpretatzen bait da, bestela faltsua bezala interpretatuko den NIL balioa itzuliko du.

<tgetflista> ::= (GETFLISTA <forma_lista> <identifikadore>)
 listaren barruan lehenengo elementu <identifikadorea> daukan azpilista bat existitzen bada egiazkoa-tzat interpretatzen da bestela faltsua-tzat.

<tnullr> ::= (NULLR <erregistro-izena>)
egiazkoa itzultzen du erregistroak NIL balioa badauka, bestela faltsua.

<tgetr> ::= (GETR <erregistro-izena>)
egiazkoa itzultzen du erregistroak balio asignatuta badauka, bestela faltsua.

<tand>::= (AND <test> <test>)

<tor>::= (OR <test> <test>)

<tnot>::= (NOT <test>)

baldintza konposatuak erabiltzeko eragileak.

<tequal>::= (EQUAL <forma> <forma>)

egiazkoa itzultzen du formak berdinak badira. Bestela faltsua.

<tusosemcompleto>::= (USOSEMCOMPLETO)

prozesuan dagoen definizio semantikorako osotasunezko testa.

4.2.1.5. Forma ebaluagarriak.

Bates-ek proposatutako forma ebaluagarri guztiak, ABORT ezik, inplementatu izan dira. Forma honek egungo arkuaz baztertzeko aukera eskaintzen du, baina ez dugu inplementatu, ez delako arku baten prozesamendua porrot egiteko forma naturala eta ez dugulako "in extremis" delako baliabide hau gure gramatikan erabili behar izan.

Sei forma berri sartu dira. Lauk balio egituratu batetik azpilistak edo elementuak ateratzeko balio dute (GETFLISTA, MD, TL ETA [] forma hutsa), bostgarrenak (LORTU_DESKRIPZIOA) aurretik analizatu eta errepresentatutako enuntziatuko objektuen deskribapenak lortzeko, eta seigarrenak (GEN_ESPEC) espezifikazioa sortzeko.

(HD "lista") eta (TL "lista") formek, hurrenez hurren, lista baten lehenengo elementua eta hondarra itzultzen dute.

() forma hutsak NIL lista hutsa itzultzen du.

(GETFLISTA <lista> <identifikadorea>. Forma honek listaren barruan lehenengo elementua bezala identifikadorea daukan azpilista bat bilatzen du. Bilaketa arrakastatsua bada azpelistaren gainerakoa itzuliko du, bestela NIL.

(LORTU_DESKRIPZIOA <obj_erref> <obj_lista> <descr>). Forma honek <obj_lista> objektu-listako <obj_erref> erreferentziako objektuari dagokion <descr> izeneko deskripzioa itzultzen du. Objektu lista normalki "obj_bild" da, zeina sorturiko baieztapen guztiak gordetzen dituen aldagai orokorra bait da. Deskribapenak izan daitezke bai orokorrak ("mota" edo "izena") bai propietateenak, esate baterako: "secuenciaincidente", zeinak sekuentzia batekiko "contenido" propietatea duen objektu batetarako sekuentziaren erreferentzia itzultzen bait du. Objektu baten erreferentzia orokorrean, objekturako errepresentazioaren barruko izena da, baina bilaketa baten emaitza bezala ere adieraz daiteke.

(GEN_ESPEC) formak enuntziatu osoaren analisia bukatutakoan lortu den asertzio-listari dagokion espezifikazio formala lortzen du.

Jarraian iterepretatzaileak ebalua ditzakeen forma guztiak azaltzen dira:

```
<forma> ::= * | <fgetr> | <fquote> | <fget12> | <fgetflista>  
          | <fbuildq> | <fappend> | <fcond> | <fnil> | <fhd> | <ftl>  
          | <flortu_deskripzioa> | <fget_espec>
```

* ikurrak momentuan analisitzen ari den sarreraren perpausa edo hitza adierazten du. CAT arkuetan hitzaren erroa da. PUSH arkuetan, aldiz, deituriko ATNak itzultzen duen hitz-talde osoa da.

```
<fgetr> ::= (GETR <erregistro-izena>  
            <izena_erreg> erregistroan dagoen balioa lortzen du.
```

```
<fquote> ::= (QUOTE <balio>  
             <balio> ebaluatu gabe itzultzen du.
```


<fget12> ::= (GETF <ezaugarri> <erregistro-izena>)
|(GET <ezaugarri>)
<izena_erreg> erregistroan dagoen hitzaren
<ezaugarri>-ari dagozkion hiztegiko balioa itzultzen du,
erregistroa ez bada aipatzen momentuan analizatzen ari
den hitzarena hartzen da.

<fgetflista> ::= (GETFLISTA <lista> <identifikadore>)
<lista> barruan lehenengo elementu bezala
<identifikadore>-a daukan azpelistaren gainerakoa
itzultzen du. Horrelakorik existitzen ez bada NIL
itzultzen du.

<fappend> ::= (APPEND <nombre_lista> <forma>)
lista argumentuen kateadura itzultzen du.
Formaren ebaluazioaren emaitza lista bat izan behar da.

<fcond> ::= (COND (<klausulak>))
<klausulak> ::= (<test> <forma>)
|(<test> <forma>) <klausulak>

test-en arauera formen artean hautatzeko aukera ematen
du.

<fnil> ::= () | nil

<fbuildq> ::= (BUILDDQ <pattern> <espresioak>)

<pattern> ::= (@ <xs>) | (<xs>)
<xs> ::= <x> | <x> <xs>
<x> ::= <pattern> | <atomo> | + | * | #
<espresioak> ::= <espresio> | <espresio> <espresioak>
<espresio> ::= <erregistro> | <espresio_ebaluagarria>
|

pattern-aren egitura daukan balio bat itzultzen du,
zeinean +, * eta # ikurren ordean dagozkien balioak edo
<izenak>-eko erregistroen balioak agertzen bait dira.
+: erregistro baten balioa adierazten du.

*: analizatzen ari den hitz edo perpausa adierazten du.
#: erregistro baten balioa, kanpoko parentesiak gabe, adierazten du.

BUILDQ-en adibidea:

(BUILDQ ((F + + + (GVER +)) TIPO SUJ AUX VER)
 1 2 3 4 1' 2' 3' 4'

Demagun erregistroen balioak ondokoak direla:
TIPO afirm ; SUJ Juan ; AUX ha ; VER visto
itzuliko den emaitza ondokoa izango da:
(F afirm Juan ha (GVER visto))

<fhd> ::= (HD <lista>)
 listaren lehenengo elementua itzultzen du

<ftl> ::= (TL <lista>)
 listaren gainerakoa itzultzen du

<flortu_deskripzioa> ::= (LORTU_DESKRIPZIOA <obj_erref>
 <objektu-lista>
 <deskr>)
 <objektu-lista> objektu-listako <erref_obj> referentzia
 duen objektuari dagokion <deskr>-ren deskribapena
 itzultzen du.

<fgen_espec> ::= (GEN_ESPEC)
 enunziatu osoaren analisia bukatutakoan lortu den
 asertzio-listari dagokion espezifikazio formala itzultzen
 du.

4.2.1.6. Analisisirako estrategia.

Interpretatzaileak segitzen duen estrategiak analisiaren posibilitate guztiak aztertzen ditu. Zuzen ez diren posibilitateak, informazio sintaktiko eta semantikoa batera erabiliz, ahalik eta lehen baztertzen saiatzen dira. Edonola ere, posibilitateak

aztertze ordena ATNen definizioarena da, aukeratutako arku bakoitza sakoneran zehar arakatzen da hurrengo arku korapilunekideak kontsideraratu baino lehenago.

Fase lexikal, sintaktiko eta semantikoetan analisirako posibilitate berriak sor daitezke.

Analisi lexikoaren fasean enuntziatutako testuko hitz batek bat baino deskonposaketa morfologiko gehiago onar dezakeelako. Deskonposaketa desberdinak hiztegiaren sarrera desberdinetara joatea ere gerta daiteke.

Analisi sintaktikoaren fasean ATNko korapilune batetik analisi desberdinei jarrai dakiekeelako, korapilunetik abiatzen diren arku desberdinak hartuz.

Analisi semantikoaren fasean hiztegiaren sarrera berberari definizio semantiko bat baino gehiago egoki dakiokelako, edo erreferentzia bat zentzait objektuekin ebatz daitekeelako.

Analisi-posibilitate desberdinak agertzen direnean, (edozein fasetan bada ere) lehenengoari jarraitzen zaio eta analisiaren egoera gordetzen da, geroago gainerako posibilitateak prozesatzeko.

Enuntziatu baten testua emanda, aukeratu den estrategiaren ondorioz emaitza bezala interpretazio posible guztiak lortzen dira. Analisia ez da bukatzen interpretazio bat aurkitzen denean.

ATNen interpretatzailea analisi sintaktiko hutsa egiteko ere erabil daiteke informazio semantikoa kontutan hartu gabe. Hala izateko nahikoa da ekintza eta test semantikoak ez erabiltzea.

4.2.2. Hiztegiaren egitura.

Atal honetan ATNren interpretatzaileak erabili behar duen hiztegiaren definiziorako eredua deskribatzen da. Hiztegia lista egitura erabiliz definitzen da, dauden hiztegiko sarrera berriak

gehitzeko edo sarrerentzako ezaugarri berriak erantzeko erraztasunak ematen baititu.

Hiztegia bere sarrera guztiez osatutako lista bezala definitzen da. Sarrera bakoitzerako definitzen diren osagaiak ondokoak dira: informazio semantikoaren azpilista, informazio sintaktikoaren azpilista eta lema (edo hobe esanda "sarrera" bera).

```
<hiztegia> ::= (<sarrerak>)  
<sarrerak> ::= <sarrera> | <sarrera> <sarrerak>  
<sarrera> ::= (<lema> <inf_sint> <inf_sem>)
```

Informazio sintaktikoaren azpilista osatzen duten osagaiak sarreraren kategoria sintaktikoa (izen, adjektibo, preposizio, izenorde etabar), eta bi osagaitako beste lista batzuk dira. Azken hauek analisi lexikorik gabe ezagut daitekeen informazio lexikoa aurkezten dute, hau da, numeroa, generoa, determinazioa, etabar.

```
Adibideak: [los [cat art] [num pl] [gen masc] [determ det]]  
           [par [cat adj]]
```

```
<inf_sint> ::= <inf_kat_sin> <beste_inf_sint>
```

```
<inf_kat_sin> ::= (CAT <cat_sint>)  
<beste_inf_sint> ::= <bikote_lista>  
<bikote_lista> ::= Ø | <bikote> <bikote_lista>  
<bikote> ::= (<ident> <balio>)
```

Informazio semantikoak hiru zati izan dezake, ez badago ere beharrezkorik. Unitate sintaktiko-semantiko baten nukleoei ez dagozkien sarrerek zati horietako bat ere ez dute.

```
<inf_sem> ::= <betebeharreko_def_sem>  
           <def_sem_gehigarriak>  
           <def_sem_gehigarri_orokorrak>
```

Lehenengo zatia lista da. Bere lehenengo osagaia SEM hitza da, eta gainontzeko osagai bakoitzak hiztegiko sarrerari dagozkien betebeharreko osagarrizko definizio semantiko bat zehazten du.

Informazio semantikoaren zati hau unitate sintaktiko-semantikoaren nukleoak diren sarreretan ageriko da, besteetan ez.

$$\langle \text{betebeharreko_def_sem} \rangle ::= (\text{SEM } \langle \text{def_sem_ak} \rangle) \mid \emptyset$$

Bigarren zatia PROP_SEM lehenengo osagai duen lista da, eta gainontzeko osagaiek osagarri gehigarrien definizio semantikoak zehazten dituzte. Hauek betebeharreko osagarriak bezalaxe adierazten dira. Nukleo izan daitezkeen eta osagarri mota hau onartzen dituzten hitzetan definituko da, besteetan ez.

$$\begin{aligned} \langle \text{def_sem_gehigarriak} \rangle &::= (\text{PROP_SEM } \langle \text{def_sem_gehigarriak} \rangle) \mid \emptyset \\ \langle \text{def_sem_gehigarriak} \rangle &::= \langle \text{def_sem_ak} \rangle \end{aligned}$$

Hirugarren zatia sekuentzia bat (eta ez lista) izango da. Beste unitate sintaktiko-semantiko bati gehitu dakizkiokeen osagarriak definitzeko erabiltzen da. Lehen aipatutako osagarri gehigarriekin konparatuz, hirugarren osagarri hauen kasuan, menpean dagoen osagarriaren nukleoa da iragazketa-prozesua gidatzen duena, eta ez osatuarena. Honexegatik deitzen dira *osagarri gehigarri orokorrak*.

$$\begin{aligned} \langle \text{def_sem_gehigarri_orokorrak} \rangle &::= \emptyset \\ &\mid \langle \text{def_sem_gehigarri_orokorra} \rangle \\ &\quad \langle \text{def_sem_gehigarri_orokorrak} \rangle \end{aligned}$$

Definizio semantiko bakoitzan (betebeharreko eta gehigarri osagarriak dituena) ondoko osagaiak agertzen dira: eskaeren lista, hauei buruzko baldintza globalen lista eta egikaritzeko ekintza semantikoak.

$$\begin{aligned} \langle \text{def_sem_ak} \rangle &::= \langle \text{def_sem} \rangle \mid \langle \text{def_sem} \rangle \langle \text{def_sem_ak} \rangle \\ \langle \text{def_sem_} \rangle &::= ((\langle \text{eskaera-lista} \rangle) \\ &\quad \langle \text{baldintza_globalak} \rangle \\ &\quad \langle \text{ekintza_sem_ak} \rangle) \end{aligned}$$

Eskaera-listako eskaera bakoitzak eskaera sintaktikoa, semantikoa eta osagarri posibleei buruzko beste baldintza batzu dauzka. Eskaera eta baldintza guztiak bete behar dira eskaera osoa betetzat onar dadin.

<eskaera-lista> ::= <eskaera> | <eskaera> <eskaera-lista>
<eskaera> ::= (<esk_sint> <esk_sem> <baldintzak>)

Eskaera sintaktikoak nukleo eta osagarriaren arteko erlazio sintaktiko posiblea mugatzen du. Ondoren, aukeratutako aplikazioaren eskaera sintaktiko posible batzu azaltzen dira.

<esk_sint> ::= (<elem_sint> <erlazio_sint>)
<elem_sint> ::= SUJ | OBJDIR | GPREP | GADJ | ...
<erlazio_sint> ::= Ø | <preposizio> | ...

Adibideak: [SUJ Ø], [GPREP DE], [GPREP CON].

Eskaera semantikoak osagarriaren kategoría semantiko posiblea mugatzen du. Ondoren, aukeratutako aplikazioaren eskaera semantiko posible batzu azaltzen dira:

<esk_sem> ::= <kat_sem>
<kat_sem> ::= (OBJ) |
 (ELEM) | (ELEM <mota oinarri>) |
 (SEK) | (SEK <mota oinarri>) |
 (SEKDESK) | (SEKDESK <mota oinarri>) |
<mota oinarri> ::= NUM | NAT | ENT | REAL |
 CAR | LETRA

Adibideak: (ELEM) , (SEK CAR) , (SEKDESK ENT).

<baldintzak> delakoak osagarriari buruzko beste baldintza extra batzu espezifikatzeko balio du. LET baldintza moduan zehaz daitezke, baldintza-mota honek definizio semantikoaren gainerakoan balio bat identifikadore baten bidez adierazteko balio duelarik.

<LET_bald> ::= (LET <ident>) |
 (LET <ident> <forma>)

LET baldintza, argumentu bakar batekin erabiltzen denean "egiazkoa" ebaluatutzat hartzen da, eta gainera, eskaeran deskribatu osagarria identifikadorearen bidez erreferentzia daiteke

definizio semantikoaren gainerakoan.

LET baldintza, bi argumentuekin erabiltzen denean (identifikadore eta forma), forma ebaluatzean NIL emaitza lortzen ez bada, baldintzak itzulitako emaitza "egiazkoa" izango da, bestela "faltsua". Lehenengo kasuan formaren balio identifikadorearen bidez erreferentzia daiteke definizio semantikoaren gainerakoan.

Definizio semantikoaren <baldintza-globalak> delakoak eskaeren osagarriak erlazionatzen dituzten baldintzak definitzeko balio du. Aurkeztutako aplikazio konkretuan ez dugu erabili behar izan posibilitate hau, baina gerta liteke beste batzutan.

Ekintza semantikoak forma ebaluagarri gisa definitzen dira, baina haietariko edozein forma ebaluazio-emaitzak ekintza semantikozko listaren sintasia bete behar du.

```
<ekintza_sem_ak> ::= <forma>  
                ;;; emaitza <ek_sem-lista>-ren sintasikoa da
```

Honelako lista bat osatzen duten ekintza semantikoak bi motatakoak izan daitezke: emandako ezaugarriekin objektu bat sortu ala aurretik sortutako objektua bati deskribapen berriak gehitu.

```
<ek_sem-lista> ::= (<ek_sem_emaitza>  
<ek_sem_emaitza> ::= <ek_sem> | <ek_sem> <ek_sem_emaitza>  
<ek_sem> ::= <sortu_ekintza> | <erantsi_ekintza>
```

Deskribapenak eta ezaugarriak izena, mota edo beste edozein propietateri dagozkie.

```
<sortu_ekintza> ::= (SORTU_OBJEKTU <balio_desk_s>  
<erantsi_ekintza> ::= (ERANTSI_DESKRIPZIOA <balio_desk_s>  
<balio_desk_s> ::= <balio_desk> | <balio_desk> <balio_desk_s>  
<balio_desk> ::= (MOTA <mota-balio>) |  
                 (IZENA <izen-balio>) |  
                 (PROP <prop-ident> <arg-lista>)
```

Definizio semantiko gehigarri orokorrak ez dira definitzen beste definizioak bezala. Definizio semantiko orokor bakoitzerako PROP_OROK lehenengo osagaia duen lista bat sortzen da. Bigarren osagaia eskaeren lista da, baina hauen arteko lehenengoa osagarri gehigarriaren jabeari dagokio. Hirugarrenak eskaera guztiei buruzko baldintza globalak ditu (aurreko definizioan bezala). Laugarren eta azken osagaia dena egikaritzeko ekintza semantikoak definitzen dituen forma da. Beste definizio semantikoen motetatik bereizten da, bi ebaluazio jasan behar baitu. Lehenengoan osagarriari dagokion informazioa gehitzen da, eta bigarreanean osatuari dagokiona.

*, +, #, @ eta "izarsem" erregistro bereziak beharrezkoak izan daitezke ekintza semantikoak errepresentatzen dituzten adierazpenetan, eta bietariko edozein ebaluazio-mailatan. Idazkera alternatiborik ez badago, bere interpretazioa lehenengo ebaluazioan burutuko da beti, batzutan bigarren mailari egokitzen bazaio ere. Arazo honek erregistro berezi berrien definizioa motibatu du. Azken hauek aurrekoen antzekoak dira, baina bigarren mailan soilik ebaluatuko dira. Definitutako korrespondentzia ondoko hau da:

1ª evaluación

izarsem

*

+

#

@

2ª evaluación

fathersem

father

plus

almohadilla

arroba

Lehenengo ebaluazioan erregistroak, osagarri gehigarri orokorrarenak izango dira. Bigarren mailako erregistro berezien balioak ondokook izango dira: fathersem ---> izarsem, father ---> *, plus ---> +, almohadilla ---> #, arroba --->@.


```
<def_sem_gehigarri_orokorrak> ::=  $\emptyset$   
      | <def_sem_gehigarri_orokorra>  
        <def_sem_gehigarri_orokorrak>
```

```
<def_sem_gehigarri_orokorra > ::=  
       $\emptyset$  |  
      (PROP_OROK  
        (<jaberako_esk> <beste_eskaerak>)  
        <baldintza_globalak>  
        <ek_sem_orokorrak>)
```

Jabeari dagokion eskaera beste edozein eskaera gisa definitzen da, baina listaren lehenengo osagaia PROPDUN da. Kasu honetan, eskaera sintaktikoak menpeko osagarri eta bere jaberaren arteko eralzioa mugatzen du. Eskaera semantikoak jabearen kategoria semantiko posiblea mugatzen du.

```
<jaberako_esk> ::=  
      (PROPDUN <esk_sint> <esk_sem> <baldintzak>)
```

5. BEHARREZKO EZAGUMENDUAK ETA BEREN ERABILPENA.

5.1. APLIKAZIOAREN EREMUARI DAGOKION EZAGUMENDU EZ-LINGUISTIKOA.

5.1.1. Errepresentazio-lengoaia.

5.1.2. Azken errepresentazioaren osotasunari buruzko ezagumendua.

5.1.3. Identifikadoreen tratamendua.

5.2. EZAGUMENDU LINGUISTIKOAREKIN LOTUTAKO EREMUARI BURUZKO EZAGUMENDUA.

5.2.1. Hiztegia.

5.2.2. Analisatzaile lexikoa.

5.2.3. ATN sintaktiko-semantikoak.

5.2.3.1. Enunziatuak interpretatzeko ikuspuntua.

5.2.3.2. Aukeratutako kategoria sintaktiko eta semantikoak.

5.2.3.3. Sortutako ATNen deskribapena.

5. BEHARREZKO EZAGUMENDUAK ETA BEREN ERABILPENA.

5.1. APLIKAZIOAREN EREMUARI DAGOKION EZAGUMENDU EZ-LINGUISTIKOA.

5.1.1. Errepresentazio-lengoaia.

Enuntziatuan azaldutako objektuen adierazpidea sinplea da, asertzio-multzo batetan datza. Asertzioak sekuentzialki eraikitzen dira, edozein momentuan aurretik sortutakoei buruzko informazioa eskuratu daitekeelarik. Jadanik sorturik dagoen asertzio bat ezin daiteke aldatu edo modifikatu. Hala ere, aurretik emandako balioarekin konpatiblea den balio konkretuago bat espezifikatuz gero, objektu eta predikatu berberaiek dituen beste asertzio bat sar daiteke, lehengo balioa birdefinitzearren.

Objektu guztiek izena eta mota semantikoa egokiturik izan behar dutenez, izena eta motaren maneiua, beste propietateena baino errezagoa izan dadin diseinatu da errepresentazio-lengoaia.

Errepresentazio-lengoiatzeko adierazpenak maneiatzeko bost funtzio definitu dira. Horietako bik asertzioak sortzeko erabil daitezke eta beste hirurak momentuan daukagun adierazpideari buruzko informazioa lortzeko.

Definizio semantikoetan agertzen diren ekintzek, asertzioak sortzeko daukagun funtzioei deitzen diete, informazioa lortzeko dauden hiru funtzioak, deiaren argumentuak adierazteko erabil daitezkeelarik. Informazioa lortzeko dauden funtzioak ATNen arkuetan definitzen diren testetan eta erreferentzi-bilaketa prozesuan ere erabil daitezke.

Hona hemen asertzioak sortzeko dauden funtzioen deskribapena:

"sortu-objektu": objektu bati dagokion deskribapen-lista harturik, objektu berri bat sortu erazten du eta berau errepresentatzen duen asertzio-multzoa eransten du. Hau lortzeko, lehendabizi barne-identifikadore bat sortzen du, ondoren berau erabiliz, deskribapen bakoitzeko asertzio bat eransten du. Deskribapenak ondorengo motatakoak izan daitezke:

mota: [mota <balio>]
izena: [izena <balio>]
orokorra: [prop <prop-ident> <argumentu-lista>]

Sortutako asertzioak ondorengo ereduen arauera izango dira:

[mota <barne-ident> <balioa>]
[izena <barne-ident> <balioa>]
[<prop-ident> <barne-ident> <argumentu-lista>]

"erantsi-propietatea": "sortu-objektu" funtzioaren baliokidea da, baina honek ez du objektu berririk sortzen. Asertzioak objektu batentzat sortuko dira eta funtzioen lehenengo argumentua objektu honi dagokion barne-identifikadorea izango da. Bigarren argumentua asertzio-lista bat izango da. Bere aplikazioaren ondorioz, espezifikatu den deskribapen bakoitzeko asertzio bat erantsiko da.

Adirazpenari buruzko informazioa lor dezaketen funtzioak honako hauek dira:

"lortu-deskripzioa": objektu baten erreferentzia (normalean objektuari dagokion barne-identifikadorea edo bestela, bilaketa baten emaitza ere espezifika daiteke) eta propietate-identifikadore bat hartuz objektu horrentzako propietateari dagokion balioa ematen du.

"konprobatu-deskripzioa": objektu baten erreferentzia eta pattern bat. Pattern honek egiaztatu behar den propietate baten deskribapena adierazten du. Egiaztapena objektuari dagozkion asertzioekin egiten den matching prozesu baten bidez burutuko da. Funtzio honek balio boolear bat itzultzen du, ATNen arkuetako testen definizioan erabil daitekeelarik.

"bilatu-objetuak": propietate-deskribapen baten pattern bat hartuz, propietate hori betetzen duten errepresentatutako objektuak dauzkan lista bat itzultzen du. Objektu bakoitzaren barne-identifikadorea aurkezten da listan.

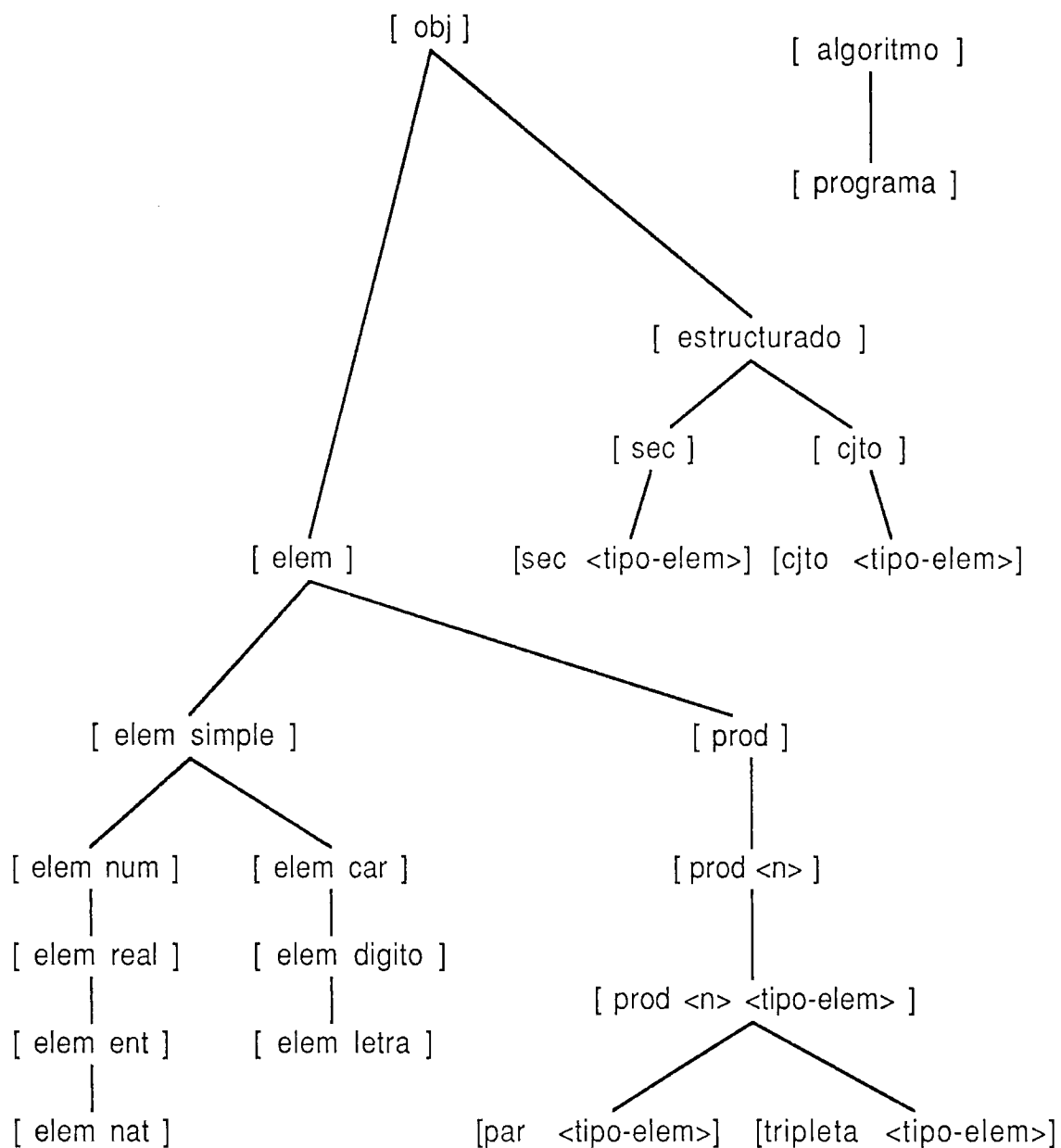
Asertzio bakoitza lista bat balitz bezala adierazten da. Listaren lehenengo osagaia propietatearen identifikadoreari dagokio. Bigarren elementua barne-identifikadorea edo objektuaren izena da (izen hau eta problemaren enuntziatuan objektuari egokitu zaion kanpo-identifikadorea desberdinak dira). Hirugarren elementua asertzioaren beste argumentuen lista da; argumentu gehiago ez balego lista hutsa ageriko litzateke.

Definitutako adirazpide hau beste aplikaziotarako ere erabil daiteke. Objektu motak eta propietate konkretuak, azaltzen dugun programaziozko problemen enuntziatuen aplikazioari loturik daude.

Erabili diren propietateak bigarren eranskinean deskribatzen dira.

Objektu-motak 5.1 irudian erakusten dira.

5.1 irudian objektu-mota baten definizioan <tipo-elem> agertzen denean, horren ordez irudi honetan [elem] motaren azpian dagoen edozein mota ager daitekela esan nahi da. Adibidez, [sec <elem-mota>] delakoak ondorengo edozein mota adieraz dezake: [sec [elem]], [sec [elem ent]], [sec [elem car]], [sec [par [elem ent]]]. Idazkera errazagoa izan dadin, [elem num], [elem real],... [elem digito] eta [elem letra] motak, *num*, *real*, ..., *digito* eta *letra*-z ordezka daitezke.



5.1 Irudia. Definitutako objektu-motak

Bi azpimota, erabilpen semantikoetarako bateragarritzat hartzen dira, berdinak baldin badira edo bietako batek besteak adierazten duen motaren barruan dagoen mota bat adierazten baldin badu. Adibidez, [elem ent]ATEGARRIA DA [elem num], [elem simple], [elem] eta [obj] motekin, baina ez [sec] edo [elem car] motekin.

Espezifikazioa egiterakoan [sekdesk] eta [sekdeskdetpl] motak ere erabiltzen dira (orokorkiako: [sekdesk <tipo-elem>] eta [sekdeskdetpl <tipo-elem>]). Honek [estructurado] motarekin eta bere azpian dauden guztiekin bateragarriak dira.

5.1.2. Azken errepresentazioaren osotasunari buruzko ezagumendua.

Enuntziatuko testua analisatu ondoren, ageritako objektuak errepresentatzen dituen asertzio-multzoa aztertu egiten da, egiaztatua izan dadin. Tratamendu hau lau azpitratatamendutan banatzen da, asertzio-multzoaren gainean arau batzuk aplikatuz burutzen delarik:

1) "Ezabatu" motako objektuen ezabapena.

Objektu hauek "deskribatzaile" motarekin sortu ziren, aurrerantzean deskribatzaile gisa erabiliak izan zitezten. Ez badira erabili objektu baten deskribatzaile gisa, objektutzat hartu direla suposa daiteke. Deskribatzaile gisa erabili izan badira, bere aurreneko mota, "ezabatu" motaz ordezkutzen da. Horrela ez dira berriro objektutzat hartuko. Objektuen ezabapenerako azpitratatamendu honetan kendu egiten dira "ezabatu" motako objektuei dagozkien asertzio guztiak.

Azpitratatamendu hau beste aplikazio-eremu batzutan erabili ahal da, zeharo definitu gabe dauden izen-sintagmei behin-behineko adirazpide semantikoa egokitzeko lagungarria baita. Analisiaren ondorengo pausuetan definizioa zehatzago beteko da edo baztertu egingo da objektua. Azpitratatamendu hau gehikuntzazko analisisa egiteko oso tresna egokia da.

2) Predikatu batzuren definizio oso eta zuzenen egiaztapena.

Arau honen erabilpenaren adibide tipiko bat asertzio desberdinen arteko harremanen egiaztapena da. Bestalde, aurretik sortuta zeuden asertzioak elkartuz, asertzio konplexuagoak sortzeko ere erabili daiteke. Ikus 2. Kapituluaren zehaztatutako adibidea.

Azpitratamendu honekin bi helburu bete nahi dira. Alde batetik, sortzaroan kontutan hartu ez diren ezagumendu osoaren inkoherentziak detektatu. Adibidez, analisisian objektu bat erreal motakoa dela alde batetik eta bikoitia bestetik inferi daiteke. Kasu honetan, beste deskribatzaileak ez balego, osoa motakoa dela inferi daiteke, baina zenbaki errearen digitu hamartarrak definituko lukeen deskribatzailearen bat balego, inkoherentzia garbi legoke, eta beraz, analisisia baztertuta izango litzateke.

Azpitratamendu honen bigarren helburua, adierazita dagoen ezagumenduaren gainean inferentziak egitea litzateke, geroko sorkuntzaren fasean errazago erabil zedin. Bigarren kapituluaren azpitratamendu honen aplikazio baten adibidea zehaztatzen da. Obj13 deskribatzaileak, obj12 sekuentziako osagaiak bereizten zituen. Bestalde, obj13 deskribatzaile bera, bikoitia propietatea dagokiolarik, obj15 sekuentzia baten partaide bezala agertzen da. Orduan zera inferi daiteke: obj12 objektua obj15 sekuentziaren azpisekuentzia bat dela, obj13 deskribatzaileak bereizten dituen osagaiak dauzkana, hau da, obj15-ren osagai bikoitiez osaturiko azpisekuentzia.

Azpitratamendu hau, beste aplikazio desberdinetarako era berean defini daiteke, nahiz eta predikatu desberdinen osotasun-arauak, aplikazio bakoitzari egokituak izan.

3) Objektuen sailkapena ondorengo hiru mailetan: sarrera, irteera eta ezsarreraezirteera.

Propietate hauetako bat definiturik ez duten objektuei hiruetako bat egokitzen zaie. Propietate hauek, espezifikazio formalaren sorkuntzan erabiliko dira, honela "sarrera" propietatea duten objektuak espezifikazioaren input-eko partaideak izango dira, era berean, "irteera" propietatea dutenak, output-ekoak izango dira. "Ezsarreraezirteera" propietatea dutenak ez dira input edo output-eko objektuak bezala azalduko, baizik eta beste objektu batzuren definizioan elementu deskribatzaile bezala erabiliko dira.

Definitutako arauak ondorengo eritziak batzen ditu:

- a) Izena eta mota bai, baina propietaterik definituta ez daukan objektu bat datutzat hartzen da (sarrerakoa).

- b) Objektu-deskribatzaileak beste objektu batzuren definizioan elementu deskribatzaile bezala erabiltzen dira soilki, beraz, ez dira datutzat hartzen, ezta emaitzatzat ere (ezsarreraezirteera).
- c) Enuntziatuan balio konstante eta esplizitua definituta duten objektuak espezifikazioan sarrerako objektu bezala definitzen ez direnez, ez dira datutzat hartzen ezta emaitzatzat ere (ezsarreraezirteera).
Adibidez: "... 3ren anizkoitza ...".
Kasu honetan "3" zenbakia ez da azalduko espezifikazioan input-eko elementu bereiztu bezala:
ENTRADA: Sea 3 ent ...
- d) [programa] edo [algoritmoa] motako objektuak ez dira datutzat hartzen ezta emaitzatzat ere (ezsarreraezirteera). Mota honetako objektuak " ... *egiten duen programa eraiki ezazu* " motako enuntziatuetan besterik ez dira azaltzen. Enuntziatu bat, eraiki behar den programa edo algoritmo baten espezifikazioa dela suposatzen denez, sarrera hau ez litzateke beharrezkoa. "Algoritmoa" edo "programa" motako objektuak ez dira azalduko espezifikazioan.

Azpitratamendu hau, aplikazio-eremu konkretuari oso lotuta dago eta ezin daiteke eremu desberdinetara zabaldu.

4) Izena definitu gabe duten objektuei izena egokitu.

Izenik ez duten objektuei edozein izen-identifikadore egoki dakieke, adibidez I, J, J48. Dena dela, objektuaren mota nolabait gogoratzen duten identifikadoreak erabiltzen badira espezifikazioan, hau irakurgarriagoa gerta liteke, adibidez sek1, sek2, oso1, nat2. Espezifikazio formalean deskribatzen diren objektuek bere izen-identifikadorea behar dute, baina enuntziatuan beti ez dira zehaztatzen objektu guztien izenak, hau dela eta, azpitratamendu hau beharrezkoa bihurtzen da. Honen implementazioa, definituta dauden objektuak banan banan aztertu eta izenik ez dutenei izen bat egokitzean datza. Izena identifikadore-sortzaile batek eraikitzen du, zeinek erro bezala objektuaren mota hartzen baitu.

Objektuak identifikadore edo izen propioen bidez erreferentziatzen dituzten aplikazio-eremuetan ere erabil daiteke azpitratamendu hau.

5.1.3. Identifikadoreen tratamendua.

Azkenik, aplikazio-eremuari buruzko ezagumendu barruan, enuntziatuan azaltzen diren identifikadoreekin egiten den tratamendua nabari daiteke. Lehendabizi, identifikadore bat kardinal baten posizioan agertzen bada, osoa mota duen elementu sinple bat sortzen da identifikadoretzat bere izen berbera duelarik. Identifikadorea objektu bat errepresentatzen duen izen baten ondoren agertzen bada, identifikadore bera objektuaren izentzat hartzen da. Azkenik izan-sintagma bat bezala agertzen bada eta izen hori duen objekturik ez balego definiturik, [obj] mota orokorra duen objektu bat sortzen da eta izentzat aurkitu den identifikadorea egokitzen zaio.

Aurretik azaldu den azpitratamenduarekin gertatzen zen bezala, azpitratamendu hau objektuak identifikadore edo izen propioen bidez erreferentziatzen dituzten aplikazio-eremuetan ere erabil daiteke.

5.2. EZAGUMENDU LINGUISTIKOAREKIN LOTUTAKO EREMUARI BURUZKO EZAGUMENDUA.

5.2.1. Hiztegia.

Proba bezala hartu diren enuntziatuak lantzeko, 70 sarrerataz osaturiko hiztegi bat erabili da. Sarrerak honela banatzen dira: 19 izen, 10 aditz, 9 artikulua, 2 konjuntzio, 3 adjektibo mugatzaile (determinanteak), 17 preposizio eta 10 adjektibo. Ondoren hiztegiko lau sarrera desberdin aurkezten dira:

[de [cat prep]]
;;; "de" sintaktikoki preposizioa da
;;; Ez dakar informazio semantikorik.

[**productorio** [cat nom] [gen masc]
 ;;; *Sintaktikoki, jenero maskulinoa duen izena da.*
 [sem [[[[gprep desde] [elem ent] [let x1]]
 [[gprep hasta] [elem ent] [let x2]]
 [[gprep de] [exp num] [let x3]]]
 ;;; *Izen-sintagma baten nukleo bezala ager daiteke,*
 ;;; *baina hiru osagarri behar ditu:*
 ;;; *Oso bat definitzen duen "desde" preposiziodun*
 ;;; *talde bat. Balego x1 bezala erreferentziatuko*
 ;;; *litzateke.*
 ;;; *Oso bat definitzen duen "hasta" preposiziodun*
 ;;; *talde bat. Balego x2 bezala erreferentziatuko*
 ;;; *litzateke.*
 ;;; *Espresio bat definitzen duen "de" preposiziodun*
 ;;; *talde bat. Balego x3 bezala erreferentziatuko*
 ;;; *litzateke.*
 []
 ;;; *Osagarriak koordinatzeko baldintza orokorrik ez*
 ;;; *dago.*
 [[sortu_objektu [mota [elem num]]
 [prop propproductorio
 [@ # # #]
 [[getflista [getr x1] sem]
 [getflista [getr x2] sem]
 [getflista [getr x3] sem]]]]]
 ;;; *[elem num] motako objektu bat sortzen du.*
 ;;; *Objektu hau x1, x2 eta x3-k definitzen duten*
 ;;; *biderkaria da.*
]]]

[**secuencia** [cat nom] [gen fem]
 ;;; *"secuencia" sintaktikoki jenero femeninoko izena da*
 [sem [[] [quote [[sortu_objektu [mota [sek]]]]]]]
 ;;; *Betebeharreko osagarririk behar ez duen*
 ;;; *izen-sintagma baten nukleo bezala ager daiteke,*
 ;;; *[sek] motako objektu bat adieraziz.*
 [propsem
 ;;; *Ondoko osagarri gehigarriarekin etor daiteke:*

```

[[[gprep de] [deskribatzailea elem]
    ;; "de" preposiziodun talde bat ementuz
    ;; -deskribatzaile bat adieraziz.
[let des1]
    ;; Bedi "des1" deskribatzailea
[let xmota [hd [tl [lortu_deskripzioa
                [hd[getflista[getr des1] sem]]
                obj_bild
                [quote mota ]]]]]
    ;; Bedi "xmota" deskribatzailearen mota
    ;; semantikoa
[let xcard
    [cond [[getflista [getr des1] card]
          [hd [getflista[hd[getflista
                        [getr des1]
                        card]]
              sem]]]]]]
    ;; Bedi "xcard" deskribatzailean azaldu
    ;; behar den kardinala
[ ]
    ;; Ez dago baldintza orokorrik
[buildq [ [erantsi_deskripzioa [quote +] [mota {sek +}]]
        [erantsi_deskripzioa [quote +] [prop longitud [+] ] ]
        [erantsi_deskripzioa des1 [mota [borrado]]]]
    izarsem xmota
    izarsem xcard]]
    ;; Aldez aurreko baldintza guztiak bete badira,
    ;; "xmota" mota duten osagaiez osaturiko eta
    ;; "xcard" kardinalak adierazten duen luzerako
    ;; [secuencia] motako objektu bat sortuko du.
    ;; Aurrerantzean, deskribatzailea ezin daiteke hartu
    ;; sekuentzia bezala.

```

```

[[[gprep de] [deskribatzailea elem]
  ;; "secuencia" sarreraren bigarren osagarri
  ;; gehigarri hau lehengoaren baliokidea da, baina ez
  ;; du behar kardinalaren presentzia deskribatzaile
  ;; barruan.
  [let des1
    [let xmota [hd [tl [lortu_deskripzioa
                      [hd [getflista [getr des1] sem]]
                      obj_bild
                      [quote mota ]]]]]
  [ ]
  [buildq [ [erantsi_deskripzioa [quote +] [mota [sek +]]]
            [erantsi_deskripzioa [quote +] [prop longitud [+]] ]
            [erantsi_deskripzioa des1 [mota [borrado]]]]
            izarsem xmota]]]]]

```

```

[par [cat adj]
  ;; "par" sintaktikoki adjektiboa da
  ;; Jeneroa ez da definitzen, "masculino" edo ""
  ;; "femenino" izan daiteke eta.
  [prop_orok
    [ [[propdun [adj] [elem num]]]
      ;; osagarri gehigarri orkor baten nukleo
      ;; bezala ager daiteke, zenbakizko elementu
      ;; baten adjektiboa izanik.
      ;; Ez du behar bestelako osagarririk
    [ ]
      ;; Ez dago baldintza orokorrik
    [buildq [buildq [[erantsi_deskripzioa [quote +]
                                                    [prop prop_par [ ]]]]
              izarsem ]
              plus] ]]]
  ;; "prop_par" predikatu eta zenbakizko
  ;; elementuaren identifikadorearekin
  ;; asertzio berri bat sortuko du.

```

```

[[[cat nom] [gen masc]
  ;; "Par" izen masculinoa ere izan daiteke.

```

```

[sem [[[[gprep de ] [elem] [let x1]]
      ;;Izen-sintagma baten nukleo bezala
      ;;agertzen bada, elementu bat adierazten
      ;;duen "de" preposiziodun taldea behar du.
[let xmota [hd [tl [lortu_deskripzioa
                  [hd [getflista [getr des1] sem]]
                  obj_bild
                  [quote mota ]]]]]
      ;; Bedi "xmota" elementuaren mota
      ;; semantikoa
[ ]
      ;; Ez dago baldintza orokorrik
[buildq [[sortu_objektu [mota [par +]]
                [t_q_elem_par +]]
        xmota +]]]]
      ;;Aldez aurretiko baldintza guztiak bete
      ;;badira, "par" motako objektu bat sortuko
      ;;du.

```

5.2.2. Analisatzaile lexikoa.

Egindako analisatzaile lexikoa sinplea da eta pretentsio handirik gabekoa, indarririk handiena analisi sintaktiko-
-semantikoaren faseari eskaini bait zaizkio.

Analisatzaile lexikoak enuntziatu osoa hartzen du sarrera gisa eta hiztegian ez dagoen informazioaz osatutako unitate lexikale sekuentzia itzultzen du. Unitate lexikaletan 5 mota bereizten dira: Puntuazio-zeinuak, Identifikadoreak, Espresioak, Adizkiak, Aditz eta hiztegian sarrera duten Hitz normalak.

Analisatzaile honek itzultitako emaitzak ondoko espezifikazio formala du:

```

<salida> ::= [<lista-posibilidades-palabra>*]
<lista-posibilidades-palabra> ::= [<posibilidades-palabra> *]
<posibilidades-palabra> ::=
    <posib_expresion> |
    <posib_identificador> |

```

```

        <posib_palabra> |
        <posib_signo_puntuacion>
<posib_expresion> ::= [ ADIERAZPENA <expresion> <tipo> ]
    <expresion> ::= <cadena_caracteres>
    <tipo> ::= <cadena_caracteres>
<posib_identificador> ::= [ IDENT <cadena_caracteres> ]
<posib_palabra> ::= [ HITZA <entrada-diccionario>
    <lista_sufijos> ]
    <atzizki-lista> ::= Ø | [ ATZIZKIAK <sufijo1>* ]
<posib_signo_puntuacion> ::= [ PMARKA <signo-punt> ]
    <signo_punt> ::= "." | "," | ...

```

5.2.3. ATN sintaktiko-semantikoak.

5.2.3.1. Enuntziatuak interpretatzeko ikuspegia.

Eremu honetarako sortutako ATNen deskribapen sakona azaldu baino lehen, behar beharrezkoa da tratatuko diren enuntziatuei buruz dugun ikuspegia agertzea, azken errepresentazioari begira osagai sintaktiko bakoitza nola kontsideratu behar dugun ikusiz.

Enuntziatu bat osagai sintaktiko desberdinez eratutako perpaus-sekuentziatzat hartzen da. Osagai sintaktikoen jario horren analisiaren helburua, aipatzen diren objektuak atera eta bakoitzaren ezaugarriak karakterizatzea da. Objektu eta ezaugarrien eransketa-prozesu hori sortzen den informazio semantikoak gidatu eta iragazten du, enuntziatuaren analisi osoaren amaiera aplikatzen diren osotasun-erregelak burututako gainbegiratzeak ere parte hartzen duelarik. Objektuen eransketa-prozesu honetan zenbait osagai sintaktikok duen eginkizuna berezia da, hala nola izen-sintagma, adjektibo, adjektibo-talde, erlatibozko perpaus eta perpausa berbera.

Izen-sintagmen garrantzia eremuko objektuen erreferentzi bide izatetik dator. Objektu bat testuan lehendabizikoz aipatzen denean era zabalean deskribatzen da, geroago askoz erreferentzia laburrago erabiltzen delarik. CAPRATE sisteman, izen-sintagmetatik ateratzen dira objektu berrien definizioak

adierazten dituzten asertzioak, edota aldez aurretik errepresentatutako objektuentzako erreferentziak.

Adjektiboak, adjektibo-taldeak eta erlatibozko perpausak ere garrantzi handikoak dira objektuen propietateak ematen bait dituzte. Preposiziodun taldeak ere garrantzi handikoak ditugu baina zeharka tratatzen dira beren barneko izen sintagma tratatzean.

Azkenik, perpausa ere propietate berriak lortzeko bide dugu, bere baitako osagai sintaktikoetan agertzen diren objektuen arteko erlazioak direla eta (izen-sintagma desberdinak subjektu, objektu zuzen eta preposiziodun taldeen funtzioetan). Galderazko perpausak ez dira kontutan hartu, ez bait dira enuntziatuetan agertzen. Aginduzkoak aldiz aintzakotzat hartu dira, zenbait problema-enuntziatutan ageri bait dira zein objektu kalkulatu behar diren adierazteko.

5.2.3.2. Aukeratutako kategoria sintaktiko eta semantikoak.

Erabili diren 14 kategoria sintaktikoak honela sailka litezke:

Kategoria klasikoak: izena, aditza, adberbioa, preposizioa, artikuloa eta izenordea. Ez dago zalantzarik berorien definizioan oso erabiliak bait dira edozein analisatzaile sintaktikotan.

Adjektibozko kategoriak: adj, adj9, adj9def eta det9. Lauak bil litezke "adjektibo" izenaren pean baina 4 azpikategoria definitzea hobetsi da izen-sintagmaren barruan har ditzaketen posizio desberdinak berezi ahal izateko. "adj" kategoriak izenaren ondoan doazen adjektiboak errepresentatzen ditu, adibidez: "un número par". "adj9" delakoak izenaren aurretik doazenak biltzen ditu, adibidez: "el primer elemento". "adj9det" kategoriak normalean izenaren aurretik joaten diren arren, izenik gabe ere erabil litezkeenak hartzen ditu, adibidez: "el anterior a ellos". "det9" kategorian adjektibo erakusle eta posesiboak daude, izenaren aurretik edota "adj9" kategoriako baten aurretik erabiltzen direlarik, adibidez: "su primer elemento", "aquellos numeros naturales que ...".

Aurredeterminatzaile kategoria (predet). Zenbatzaileak biltzen dira kategoria honetan. Daudenean izen-sintagmaren lehen posizioan egon ohi dira, adibidez: "cada numero", "todos los divisores del número z".

Hiztegiko sarrera ez diren hitzen kategoriak: puntuazio-zeinuak (signop), identifikadoreak (ident) eta espresioak (exp). Analisatzaileak kategoria horien barruan aintzakotzat hartutako unitate lexikalei dagozkie, adibidez: ".", "S", " $(l+1)*(l+2)/l$ ".

Erabilitako kategoria semantikoen multzoak 5.1 irudian agertzen diren eremuko objektu-mota guztiak hartzen ditu. Beste bi kategoria erantsi dira izen propio ez diren edo singularrean ez dauden izen-sintagmak ulertu ahal izateko. Pluralean dauden izen-sintagmak errepresentatzen duten kategoria semantikoa, ezin da ziurtasunez determinatu analisiaren bukaeran. Azaltzen den testu-inguruaren arabera, sekuentzia-deskribatzaile soilaren edota objektu egituratu baten (sekuentzia edota multzoa) erreferentzia izan liteke. [sekdesk <tipo-elem>] eta [sekdeskdet <tipo-elem>] kategoria semantikoei esker anbigutasun hori errepresenta liteke harik eta ebatzi ahal izan arte.

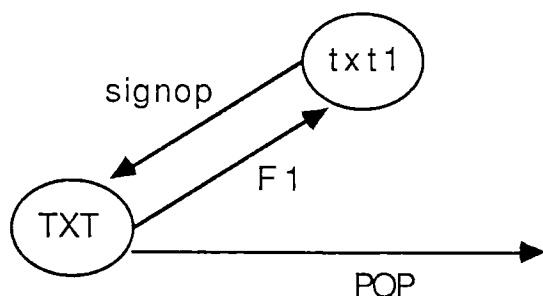
5.2.3.3. Sortu diren ATNen deskribapena.

10 ATN desberdin definitu dira aplikazio honetan. Bakoitzaren bereizgarri analisatzen duten osagai sintaktikoa eta hasierako korapilunearen identifikadorea ditugu, eta honako hauek dira:

<u>Analizatzen duen osagaia</u>	<u>Hasierako korapilunea</u>
Problemaren enuntziatuaren testoa	TXT
Perpaua	FI
Izen-sintagma	GN
Preposiziodun taldea	GPREP
Erlatibozko perpaus menpekoak	REL
Adjektiboa	ADJ
Adjektibo-talde	GDDJ
Espresio matematikoa	EXPR
Kardinala	CARD
Identifikadorea	IDENT

Atal honetan ATN guzti hauen deskribapena emango da, horretarako bere definizio-testua, diagrama sinplifikatua, analiza dezakeen egitura sintaktikoaren deskribapena, tratamendu semantikoaren justifikazioa eta azkenik zenbait adibide azalduz.

5.2.3.3.1. Testu osoaren analisirako ATNa (TXT)



```

[TXT
  [push F1 t
    [addr texto *]
    [to txt1]]
  [pop [buildq [[analisi sintactico: ] +
    [especificacion correspondiente:] #]
    texto [gen_espec]]
  t ]
]

[txt1 [cat signop t

```

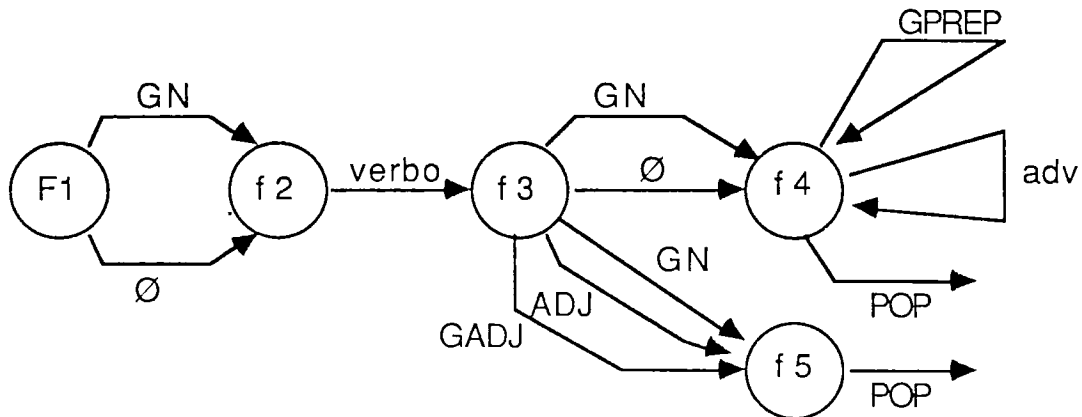
```
        [addr texto [buildq [signo_punt *]]]
    [to TXT]]
]
```

"TXT" ATNa gai da puntuazio-zeinuen bitartez bereiztutako esaldi-sekuentzia analisatzeko. Esaldi bakoitzak "F1" ATNak analisatzeko modukoa izan behar du. Oraingoz definitutako puntuazio-zeinu bakarrak, koma eta puntua dira. Perpaus bakoitzaren analisi sintaktikoaren emaitza "texto" lista-erregistroari eranstean zaio eskuinetik eta honek esaldi guztien analisia biltzen du. Esaldi bakoitzean sortutako objektuen adierazpideko asertzioak "obj-bild" izeneko aldagai orokorrari erantsi zaizkio osagai sinpleagoen analisia burutu ahala. ATN horixe denez lehendabizikoz deitzen dena enuntziatu oso bat analisatzerakoan, POP arkuak ez da gaitutako hitz guztiak (unitate lexikalak, hobe esanda) pasa arte. POP arkuak gaitutzen bada, itzulitako espresioak analisi sintaktikoaren emaitza osoa eta dagokion espezifikazioa edukiko ditu.

ATN honi ez zaio egokitu inongo tratamendu semantikorik. Beraz, ez dira definitu ez nukleo semantikorik ez iragazketa-ekintzarik. Hala ere, analisiaren emaitza definitzen duen forma ebaluagarriaren barruan [gen_espec] delako azpiforma erabiltzen da, zeinek objektuen errepresentazioa diren asertzioen listari dagokion espezifikazioa sorrerazten bait du. [gen_espec] ebaluatzean lortzen den emaitza nahi dugun espezifikazioa da.

ATN honek analisa litzakeen testuen adibide gisa, proposatutako enuntziatuen arteko edozein har daiteke.

5.2.3.3.2. Perpaus baten analisirako ATNa (F1)



```

[F1
  [push GN t
    [waitsem * [quote [suj 0]]]
    [setr suj *]
    [to f2]]
  [jump f2 t
    [waitsem nil [quote [suj 0]]]
  ]
]
[f2
  [cat verbo t
    [setr verbo [buildq [[verbo *]]]]
    [possem *]
    [to f3]]
]
[f3
  [push GN t
    [filtrar * [quote [objdir 0]]]
    [setr objdir [buildq [[objdir *]]]]
    [to f4]
  ]
  [jump f4 t
    [filtrar nil [quote [objdir 0]]]
  ]
  [push GN [equal [quote [predicativo]]
    [getf tipoverbo verbo]]
    [filtrar * [quote [atrib 0]]]
    [setr atrib [buildq [[atrib *]]]]
    [to f5]
  ]
]

```

```

]
[push ADJ [equal [quote [predicativo]]
               [getf tipoverbo verbo]]
  [filtrar * [quote [atrib 0]]]
  [setr atrib [buildq [[atrib *]]]]
  [to f5]
]
[push GADJ [equal [quote [predicativo]]
                  [getf tipoverbo verbo]]
  [filtrar * [quote [atrib 0]]]
  [setr atrib [buildq [[atrib *]]]]
  [to f5]
]
]
[f4
  [push GPREP t
    [addl gprep *]
    [filtrar * [buildq [@ [gprep] #]
                    [getflista [getr gprep] prep]]]
    [to f4]
  ]
  [cat adv t
    [addl adv [buildq [[adv *]]]]
    [filtrar * [quote [adv]]]
    [to f4]
  ]
  [pop
    [buildq [ [@ [orac] + + + + +]
              suj verbo objdir gprep adv]
    [usosemcompleto]
  ]
]
]
[f5
  [pop
    [buildq [ [@ [orac] + + + ]
              suj verbo atrib]
    [usosemcompleto]
  ]
]
]

```

"F1" ATNak esaldi deklaratiiboak eta aginduzkoak analisatzen ditu, oinarrizko bi egitura sintaktiko hauekin: *subjektua + aditza + obj.zuzena* eta *subjektua + aditza + atributoa*, non subjektua edo objektu zuzena ez azalduta ere ontzat ematen den. Gainera, etor litezke preposiziodun taldeak zein adberbioak esaldi baten amaieran, esaldi horren esannahia aldaraziko dutelarik. Subjektuak eta objektu zuzenak (baldin badaude) "GN" izeneko ATNak analisatzeko moduko izen-sintagma izan behar dute.

"Oración" (perpauza) unitate sintaktiko-semanticotzat hartzen da eta bere nukleoa aditza da. Hortaz, f2tik abiatu eta aditza pasatzen duen arkuak POSSEM izeneko ekintza bat dauka zeinek aditzaren erabilera semantiko guztiak askatzen bait ditu analisi-posibilitate guztiak sortzearen.

F1etik abiatzen den GN arkuak esaldiaren subjektu bezala prozesatutako izen-sintagma iragazi behar luke, baina aditza (unitatearen nukleoa) oraindik ageri ez denez izen-sintagma zai ezartzen da [WAITSEM * [QUOTE [suj 0]]] delako ekintzak iragaz dezan.

f3tik abiatzen den GN arkuak objektu zuzentzat iragazten du izen-sintagma. JUMP jauzi-arkuek subjektua edota objektu zuzena ez daudeneko kasuak iragazten dituzte.

f4tik abiatzen den GPREP arkuak izen-sintagma semantikoki iragazten du, hasierako preposizioa kontutan hartuz. "gprep" erregistrotik lortzen da preposizioa [getflista [getr gprep] prep] formaren bitartez. Arku honi esker, aditzaren definizio semantikoan ezarritako betebeharreko osagarriei erantzuten dieten preposiziodun taldeen analisia egin daiteke. Aditzaren osagai gehigarri orokor bezala analisatu behar diren preposiziodun taldeak prozesatzeko premia egongo balitz nahikoa litzateke honako arku hau eranstea.

```
[push GPREP t
      [addl gprep *]
      [hartu_modif_orokorrak]
      [to gnf]]
```

"f4" korapiluneko "adv" arkuak adberbio bat analisatzen du

aditzak eskatutako osagarri bezala, eta horrela iragazten du. Pentsa liteke arkuak bikoiztea GPREP delakoarekin egin den bezalaxe.

"f3"tik abiatu eta "f5"era doazen GN, ADJ eta GADJ arkuak izen-sintagma, adjektiboa edo adjektibo-taldea iragazteko dira, esaldi kopulatiboetan atributo zeregina betetzen duten kasuetan. Horrexegatik eskatzen du test-ak aditz predikatiboa.

"f5"tik abiatzen den POP arkuak bukaera ematen dio perpausaren analisiari. "f4"tik abiatzen denaren antzekoa dugu baina egitura sintaktiko desberdina sorrerazten du.

"f4" azken korapilunetik abiatzen den POP arkuak bukaera ematen dio perpausaren analisiari. Aukeratutako erabilera semantikoa osaturik baldin badago erabilera semantiko horri lotutako ekintza semantikoak exekutatu eta analisi sintaktikoari dagokion arbola sortzen du.

Analizatutako esaldien zenbait adibide:

Dada una secuencia de enteros

(V GN)

Ordenar la secuencia ascendentemente

(V GN ADV)

Dados dos puntos en el espacio bidimensional

(V GN GPREP)

M representa el número de comas

(GN V GN)

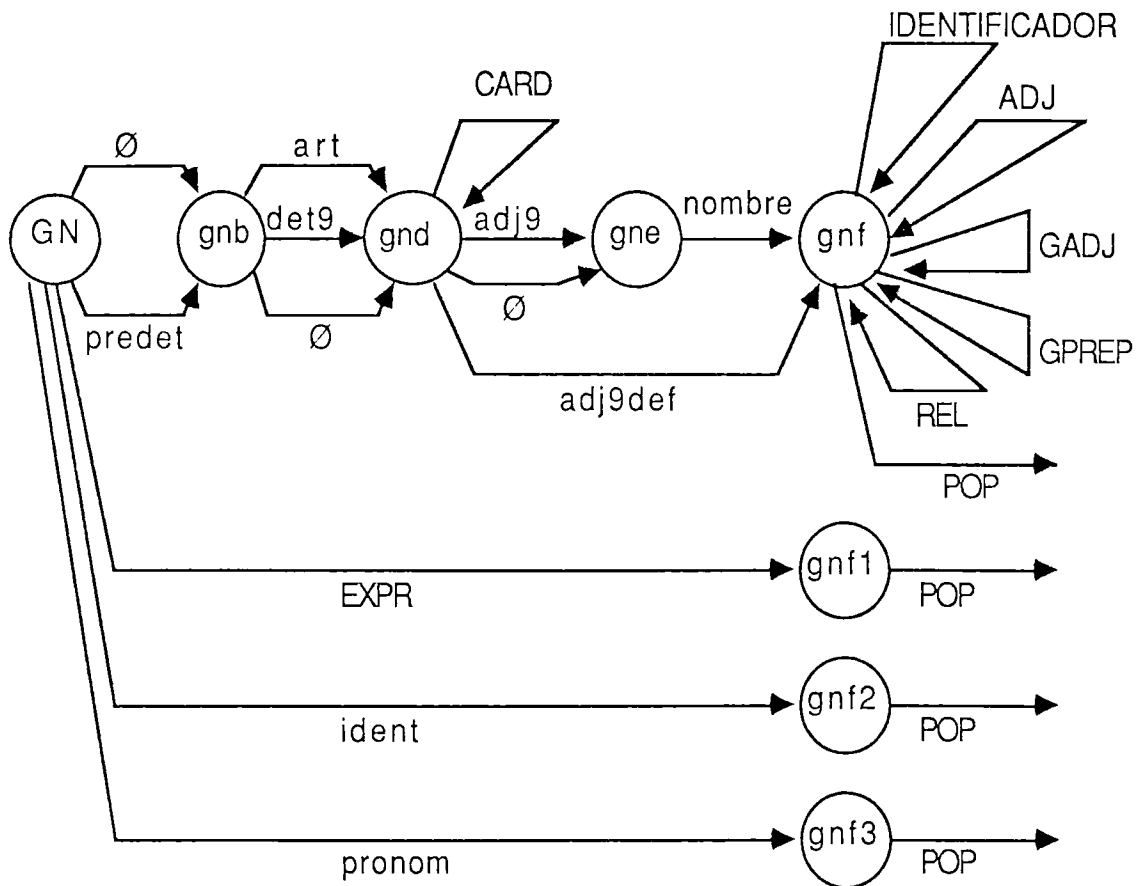
La longitud de la secuencia es par

(GN V atr)

M es un número entero

(GN V atr)

5.2.3.3.3. Izen-sintagmaren ATNa (GN)



```
[GN
  [jump gnb t]
  [cat predet t
    [setr predet [buildq [[predet *]]]]
    [to gnb]
  ]
  [push EXPR t
    [possem [buildq [exp *]]]
    [to gnf1]
  ]
  [cat ident t
    [possem [buildq [ident *]]]
    [to gnf2]]
  [cat pronom t
    [setr nom [buildq [pronom *]]]
    [to gnf3]]
]
```

```

]
]
[gnb
  [cat art t
    [setr art [buildq [[art *]]]]
    [setr determ [getf determ]]
    [to gnd]
  ]
  [cat det9 t
    [setr det9 [buildq [[det9 *]]]]
    [setr determ [getf determ]]
    [to gnd]
  ]
  [jump gnd t]
]
[gnb
  [push CARD t
    [addl adj [buildq *]]
    [to gnd]
  ]
  [cat adj9 t
    [addl adj [buildq [[adj9 *]]]]
    [to gne]
  ]
  [jump gne t]
  [cat adj9def t
    [addl adj [buildq [[adj9def *]]]]
    [setr nom [quote [elipsis]]]
    [to gnf]
  ]
]
]
[gne
  [cat nom t
    [setr nom [buildq[[nom *]]]]
    [setr numero [getf num]]
    [possem *]
    [to gnf]
  ]
]
]
[gnf

```

```

[push IDENTIFICADOR t
  [setr identificador [buildq [identificador *]]]
  [hartu_modif_orokorrak]
  [to gnf]
]
[push ADJ t
  [addl adj1 *]
  [hartu_modif_orokorrak]
  [to gnf]
]
[push gadj t
  [addl adj1 [buildq [[gaj *]]]]
  [hartu_modif_orokorrak]
  [to gnf]
]
[push GPREP t
  [addl gprep *]
  [hartu_modif_orokorrak]
  [to gnf]
]
[push GPREP t
  [addl gprep *]
  [filtrar * [buildq [@ [gprep] #] [getflista [getr gprep] prep]]]
  [to gnf]
]
[push REL t
  [addl adj1 *]
  [hartu_modif_orokorrak]
  [to gnf]
]
[pop
  [buildq [@ [gnom] [[sem +]] + + + + + ]
    izarsem predet art det9 adj nom adj1 gprep ]
  [usosemcompleto]
]
]
[gnf1
  [pop
    [buildq [ gnom [sem +] [@ [exp] *] ] izarsem ]
    [usosemcompleto]
  ]
]

```

```

    ]
  ]

  [gnf2
    [pop
      [buildq [gnom [sem +] [ident *] ] izarsem]
      [usosemcompleto]
    ]
  ]

  [gnf3
    [pop
      [buildq [gnom [pronom *] ] ]
      t]
    ]
  ]

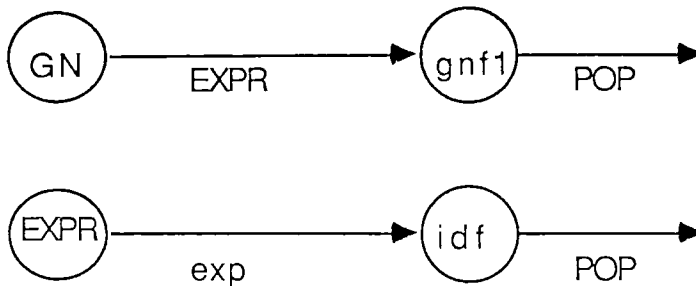
```

Hau da sortutako guztien artean ATN konplexuen eta inportanteena. Honen arrazoa tratatu diren textuetan datza, non lanik garrantzitsuena normalean izen-sintagmen bitartez adierazitako objektuak errepresentatzea bait da. "GN" ATNak analisa litzakeen izen-sintagmak sinpleak izan litezke, hala nola espresio bat identifikadore bat edo izenorde bat, edota oro konplexuak, beren baitan era askotako osagaiak biltzen dituztelarik, hala nola predeterminanteak, (aurremugatzaileak), artikuluak, mugatzaileak, kardinalak, adjektiboak, identifikadoreak, adjektibo-taldeak, preposiziodun taldeak ete erlatibozko esaldi menpekoak.

Lau POP arku desberdin definitu dira (gnf, gnf1, gnf2 eta gnf3 korapilunetan), beraz lau bide analisia behar den bezala amaitzeko. ATNaren deskribapenean lau kasu hauek bereiztuko ditugu. Bide horietako bakoitzak analisi sintaktikoaren arbola bat emango digu. Gainera, "izen-sintagma" unitate sintaktiko-semantikoa denez, bide horietako bakoitzarentzat nukleoa eta tratamendu semantikoa definitu beharra egongo da.

Lauetarik hiru bide labur eta sinpleak dira; laugarrena da luzeen eta konplexuena. Azkenerako utziko dugu horren deskribapena.

Lehenengo bideari jarraituz espresio matematiko soil batez osaturiko izen-sintagmak analisa litezke:



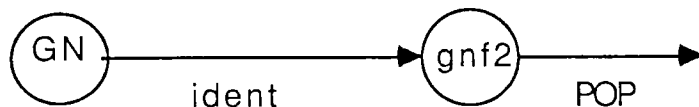
Analisa lortzeko "EXPR" ATNari egiten zaio dei eta honek "exp" kategoriako izan behar duen testu-elementu bat pasako du. Kategoria honetako bezala sarrera-listatik analisatzaile lexikoak "EXPRESSION"tzat analisatu dituen elementuak hartuko dira.

Expresioa bera da bide honen nukleo semantikoa. Hiztegian edozein espresioaren erabilera semantikoa eduki ez litekeenez, POSSEM funtzioaren lana da espresio guztientzako definizio semantikoa sortzea. Sortutako definizio semantikoak ez du inolako osagarri eskatzen eta [exp <tipo exp>] motako objektua sortzen du (<tipo exp> espresioaren mota delarik), espresioaren adierazpidea eta objektua identifikatuz. Espresioko identifikadoreetako bat ez badagokio lehenago errepresentatutako objektu bati objektu berri bat sortuko da izentzat identifikadore hori hartuz.

Adibidea: "...($l+1$)*($l+2$)/ l ..." espresioaren analisiaren emaitza ondoko hau izango da:

sintaktikoa: [gnom [sem obj1] [exp ($l+1$)*($l+2$)/ l]
 semantikoa: [mota obj1 [exp real]]
 [equal obj1 [($l+1$)*($l+2$)/ l]]

Bigarren bideak identifikadore bakarraz osatutako izen-sintagmak analisatzeko aukera eskaintzen du:



Testuko lehen elementuak derrigorrez "ident" kategoriakoa behar du izan. Kategoria honetako bezala sarrera-listatik lexikoki IDENTIFICADOR bezala analisatutako elementuak hartuko dira.

Bide honen nukleo semantikoa identifikadorea bera izango da. Espresioen kasuan bezala, POSSEM funtzioa izango da nukleo horren erabilera semantikoa sorreraziko duena, hiztegian ez bait dago definiturik. Baina erabilera semantiko hori askatu aurretik, POSSEM funtzioak alde aurretik asertzioen bidez errepresentatutako objektuen artean bilatuko du identifikadore hori izentzat duen objekturen bat. Aurkituz gero ez da erabilera semantiko berririk askatuko, izen hori duen objektuari erreferentzia itzuliz. Aurkitu ezean aldiz, POSSEM funtzioak askatutako erabilerak motarik abstraktueneko objektu berri bat sortzeko du: [obj]. Mota honen egokitzapena egiteko, eremuari buruzko eta erabat pragmatikari dagozkion erizpideak erabiltzen direla hartu behar da kontutan.

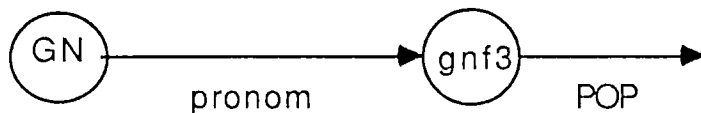
Adibidea: "... S ..."ren analisiak bi emaitz mota eman ditzake: Aldez aurretik errepresentaturiko objekturik baldin badago, esaterako obj3, S izentzat zuelarik ([nombre obj3 S]), honako hau lortuko da:

sintaktikoa: [gnom [sem obj3] [ident S]]
 semantikoa: ez da asertzio berririk erantsiko

Ez badago halako objekturik lehendik, objektu berri bat sortuko da, obj7 esaterako, eta hona emaitza:

sintaktikoa: [gnom [sem obj7] [ident S]]
 semantikoa: [tipo obj7 [elem ent]] eta
 [nombre obj7 S] erantsiko dira

Hirugarren bideari esker, izenorde soilaz osatutako izen-sintagmak analisa litezke:

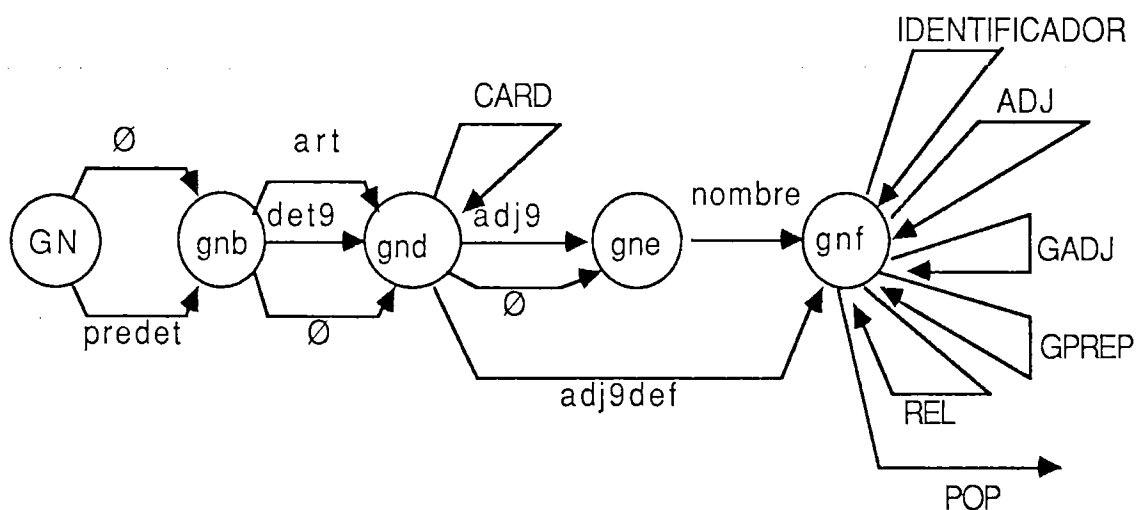


Testuko lehen elementuak "pronom" kategoriakoa izan beharko du derrigor. Egokitu zaion tratamendu semantikoa ez da oraingoz oso aproposa, "ad hoc" xamarra baizik. Erabilera semantiko baten erabilera osoa egiaztatzen denean, bete gabeko betebeharrak alde aurretik ageritako objektuekin osatzen saiatzen denez, izenordeei egokitutako tratamenduak egiten duen gauza bakarra sintaktikoki ezagutzea da. Asma liteke ebazpide hoberik, esaterako izenordeak sortutako erreferentzia, pertsona eta numeroaren komunizadura erizpideak erabiliz ebaztea. Dena dela, hartutako ebazpidea gure domeinuko kasu anitzetarako baliagarria da.

Adibidea: "...escribirlos..." horretako izenordearen analisiak honako emaitzak ematen ditu:

sintaktikoa: [gnom [pronom los]]
 semantikoa: ez du asertziorik eransten

Laugarren bideak elementu anitzek osaturiko izen-sintagma zabalagoak analisatzeko aukera eskaintzen du:



"todos" edo "cada" bezalako aurreterminatzaile bat aurkitzea bide honen hasieran aukerazkoa da.

Halaber, aukerazkoa da ondoren artikulua edo det9 delako bat etortzea (det9 kategorian adjektibo erakusleak eta posesiboak biltzen ditugu). Honetako osagaien bat egonez gero, berari dagokion determinatu/indeterminatu balioa jasoko da izen-sintagma osoari eranstea. Besterik ezean "indeterminatu" balioa hobetsiko da.

Ondoren (gnd egoeratik), kardinal bat egon badaiteke ere, hiru aukera ditugu: adj+izena, izena soilik edota adjektibo bat bakarrik, izena eliptiko egonik. Hirugarren aukera honetarako ez da tratamendu semantikorik diseinatu. Aukeratariko edozein jarraituz gnf amai-egoerara iristen da, non izen-sintagmari amaiera ematen dion POP arkuaz gain edozein ordenatan adjektiboak, adjektibo-taldeak, preposiziodunak eta erlatibozkoak pasa litzaketen arkuak bait ditugu. Era berean, errepresentatutako izen-sintagmari deitura bat ematen dion identifikadore bat pasatzeko arkuak ere badago. Komeni da ohartzea alabaina, era berdintsuan definiturik egonik ere, identifikadorea gehienez behin ager ditekela, hori bai posizioa aukerazkoa izanik. Adibidez:

La secuencia \underline{S} de enteros ...

La secuencia de enteros \underline{S} ...

Izen-sintagmaren analisi-bide honen nukleo semantikoa izena da. Beraz, gure egoeratik abiatuz izena hartzen duen arkuko azkenurreko ekintza [possem *] da. Arku honentantxe hartzen da izenaren numeroaren balioa ere izen-sintagmari hedatzearen.

Mota eta osagarri guztiak bikoiztu zitezkeen izenak eskatutako osagarri edo osagarri gehigarri orokor bezala analisatu ahal izateko, baina oraingoz preposiziodun taldearekin besterik ez da egin. Gainerako bikoizketa lan erraza litzateke. Izenak eskaturiko osagaien eskarrietan iragazketa-ekintza bat txertatzen eta eskari gehigarri orokorretan [hartu_modif_orok] delako ekintza bat, erantsi beharreko asertzio semantikoak bildu ahal izateko.

gnf egoeratik abiatzen diren osagarri-arku guztietan ADDL ekintza erabiltzen da osagarriaren edukin sintaktikoa gordetzeko

dagokion erregistroan. SETR-ren ordeaz ADDL erabiltzearen arrazoia osagarri hauek errepikatzeko posibilitatean datza, kasu horretan lista batean bildu beharko direlarik. IDENTIFICADOR delakoari dagokion arkuak aldiz SETR ekintza erabiltzen du, osagai hau behin bakarrik ager bait liteke.

POP arkua da unitate sintaktiko-semanticoren amaia eta [USOSEMCOMPLETO] testa dauka bere baitan.

Azken bide honen bitartez analisatutako izen-sintagmen adibideak:

los enteros negativos que pertenecen a la secuencia de enteros S
 art nombre adj rel

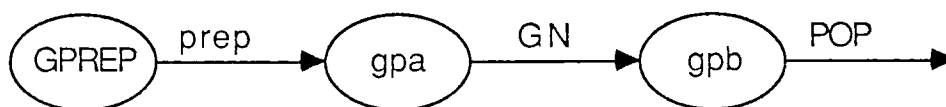
una secuencia de enteros S
 art nombre gprep ident

los N primeros números primos
 art card adj9 nombre adj

estos 5 números
 det9 card nombre

todos los elementos pares que tiene dicha secuencia
 predet art nombre adj rel

5.2.3.3.4. Preposiziodun taldearen analisirako ATNa



[GPREP
 [cat prep t
 [setr prep *]
 [to gpa]

```

    ]
  ]
  [gpa
    [push GN t
      [setr gn *]
      [to gpb]]
  ]
  [gpb
    [pop [buildq [gprep [prep +] +]
              prep gn]
      t]
  ]
]

```

Preposizio bat eta izen-sintagma batez osatutako preposiziodun taldeak analisa litzake ATN honek. Ez du ez testik ez ekintza semantikorik bere baitan. Beronen lan semantikoa "zubi"-tzat har liteke, alegia, "GPREP" ATNari dei egiten dioten perpaus edo izen-sintagmetatik askatutako eskariak eta hark behar duen izen-sintagmaren analisi sintaktiko-semantikoa lotzen dituen zulia. ATN 'GPREP' delakoaren barneko izen-sintagmaren analisisian objektu bat sortzeko edota bere erreferentzia lortzeko tratamendu semantikoa burutu da. ATN honi egindako deia sortu duen PUSH GPREP arkuari aipaturiko izen-sintagma sortu duen preposiziodun taldearen egokitasuna egiaztatuko da.

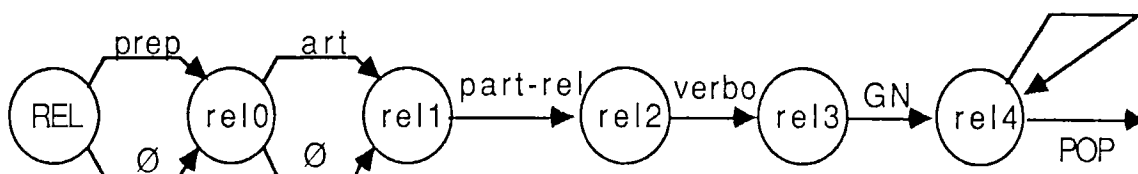
Ez du garrantzi semantikorik ez bait da unitate sintaktiko-semantikotzat hartu.

Analizatutako preposiziodun talde zenbait:

de 20 enteros
 prep GN

con los elementos pares de S
 prep GN

5.2.3.3.5. Erlatibozko perpaus menpekoen analisirako ATNa (REL)



```

[REL
  [jump rel0 t
    [setr preprel [ ] ]]]
  [cat prep t
    [setr preprel [buildq [prep *]]]
    [to rel0]
  ]
]
[rel0
  [cat art t
    [setr determ [getf determ]]
    [setr art [buildq [art *]]]
    [to rel1 ]
  ]
  [jump rel1 t]
]
[rel1
  [car part_rel t
    [setr p_rel *]
    [to rel2]
  ]
]
[rel2
  [cat verbo t
    [setr verbo [buildq [[verbo *]]]]
    [possem * [buildq [@ [rel] #]
      [getflista [getr preprel] prep]]]
    [to rel3]
  ]
]
[rel3
  [push GN t

```

```

        [filtrar * [quote [gn 0]]]
        [setr gn [buildq [[gn *]]]]
        [to rel4]
    ]
    [jump rel4 t]
]
[rel4
    [push GPREP t
        [filtrar * [buildq [@ [gprep] #]
                    [getflista [getr gprep] prep]]]
        [addl gprep *]
        [to rel4]]
    [pop [buildq [[@ [rel] + + + + + +]]
        preprel p_rel art verbo gn gprep]
    [usosemcompleto]]]

```

ATN honek erlatibozko perpausak analisatzen ditu, perpausak aurretik preposizio edo artikulua bat eraman dezaketelarik. "part-rel" delakoak erlatibozko partikulen kategoria sintaktikoa errepresentatzen du, hala nola "que", "cual" edo "cuales". Erlatibozko perpausa bera izen-sintagma eta preposiziodun talde batek edo batzuk osatzen dute.

Erlatibozko perpausak aurrekoari erantsiko dizkio zenbait propietate eta, beraz, unitate sintaktiko-semantikotzat hartu beharra dago. Nukleoa aditza da. Aditza honen hiztegi-sarrerak erlatibozko perpausarekiko erabilera gehigarri orokorren bat izan beharko du, eta honela hasiko da:

```
[prop-orok [[[propdun [rel ... ].....]...]]...]]
```

Nukleoko aditza pasatzen duen arkuak POSSEM ekintza dauka eta honek askatzen du erabilera semantiko gehigarri orokor hori (edo horiek, bat baino gehiago egonez gero).

Izen-sintagma eta preposiziodunak baldin badaude, FILTRAR izeneko ekintza semantikoaren arauera askatutako erabilera semantikoak iragaziko ditu.

Azkenik POP arkuaren ardura izango da erabilera semantikoaren behar den bezala bete dena egiaztatzea.

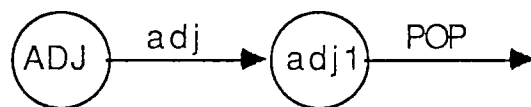
Hala bada, goiko ATNari erlatibozko perpausak erantsitako propietatea errepresentatzen duten ekintza semantikoak bidaliko zaizkio.

Adibideak:

que tiene dicha secuencia
que verbo GN

en la que existe un punto
prep art que verbo GN

5.2.3.3.6. ADJaren ATNa



```

[ADJ
  [cat adj t
    [setr adj [buildq [[adj *]]]]
    [possem * [adj]]
    [to adj1]
  ]
]
[adj1
  [pop [getr adj]
    [usosemcompleto]
  ]
]

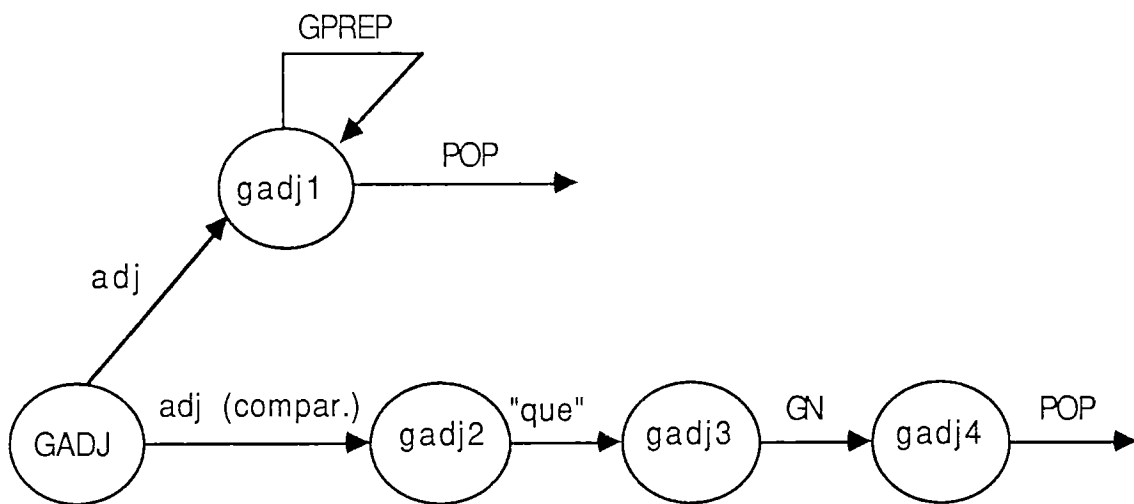
```

Lehen begiratu batean ATN oso bat definitzeak ez dirudi beharrezko, adjektiboa pasako lukeen CAT arku bat definitzea aski bait litzateke. ATN bezala definitzeko premia ordea, adjektibo bat unitate sintaktiko-semantikotzat hartu beharretik dator, unitate sintaktiko-semantiko batek bere ATN berezia eskatzen bait du. Normalean izenaren gaineko osagai gehigarri orokorraren zeregina betetzen du.

Nukleo semantikoa adjektiboa bera da eta POP arkuak [USOSEMCOMPLETO] testa dauka.

Adibideak : par
 positivo

5.2.3.3.7. Adjektibo-taldearen ATNa (GADJ)



```

[GADJ
  [cat adj t
    [setr adj [buildq [[adj *]]]]
    [possem * [gadj]]
    [to gadj1]
  ]
  [cat adj [getf comparativo]
    [setr adj [buildq [[adj *]]]]
    [possem * [adjcomp]]
    [to gadj2]
  ]
]
  
```

```

[gadj1
  [push gprep t
  
```

```

    [addl gprep *]
    [filtrar * [buildq [@ [gprep] #] [getflista [getr gprep]
                                                    prep]]]
    [to gadj1]
  ]
  [cat adv t
    [addl adv [buildq [[ adv *]]]]
    [filtrar * [quote adv]]
    [to adj1]
  ]
  [pop [buildq [@ [gadj] + + +]
        adj gprep adv]
    [usosemcompleto]
  ]
]
[gadj2
  [word "que' t
    [to gadj3]
  ]
]
[gadj3
  [push GN t
    [setr comp *]
    [filtrar *]
    [to gaj4]
  ]
]
[gaj4
  [pop [buildq [@ [gadj comp] + [que] +]
        adj comp]
    t]
]

```

Adjektibozko taldea adjektibo baten tankerakoa da baina adjektibo baten ondoren preposiziodun taldea edo adberbio osagarriak etor litezkeelarik.

Nukleo semantikoa adjektiboa da eta osagarri posibleak osagarri bezala iragaziak izango dira.

POP arkuak [USOSEMCOMPLETO] testa dauka.

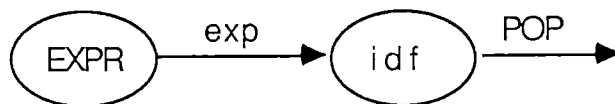
Adibideak:

el entero mayor de la secuencia
adj GPREP

múltiplo de 3
adj GPREP

mayor que M
adj GN

5.2.3.3.8. Espresio matematikoak analisatzeko ATNa (EXPR)

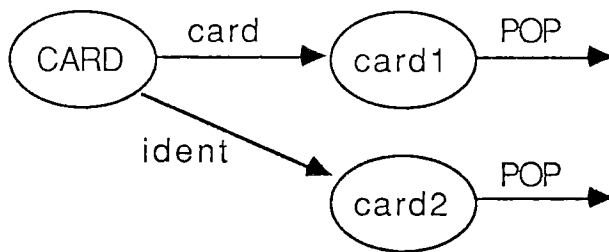


```
[EXPR  
  [cat exp t  
    [setr expresion [buildq [* [@ [mota] #]]  
      [getflista [getr izarosoa] mota]]]  
    [to idf]  
  ]  
]  
[idf  
  [pop [getr expresion]  
    t]  
]
```

ATN hau oso sinplea da, analisatzaile lexikalak espresiotzat hartutako unitatea pasa eta emaitza gisa itzuli besterik ez bait du egiten. ATN gisa definitu da gerora begira zabalkuntza baten posibilitatea kontutan hartuz. Espresio konplexuagoak eta formulak tratatuko lirateke zabalkuntza horretan, horretaz gain espresioan ageritako objektuen egiaztapena burutuko litzateke.

"izarosoa" delako erregistroa erregistro berezi bat da, "" delakoaren antzera bere balioa unitate lexikoen listatik hitz bat hartzen duenean automatikoki jasotzen duena. "" erregistroarekin alderatuz, horrek hiztegitik dagokion hitza besterik hartzen ez duen bitartean "izarosoa" delakoak unitate lexikal osoa hartzen du. Espresioen kasuan unitate lexikalak espresioaren mota ere badauka; mota hori ezagutzea beharrezkoa da espresioa objektu bezala definitu ahal izateko.

5.2.3.3.9. Kardinalentzako ATNa (CARD)



```

[CARD
  [cat card t
    [setr card *]
    [possem *]
    [to card1]
  ]
  [cat ident t
    [setr determ [quote det]]
    [setr numero [quote sing]]
    [setr card [buildq *]]
    [possem [buildq [identcard *]]]
    [to card2]
  ]
]
]
[card1
  [pop [buildq [[card [* [sem +]]]]] izarsem]
  [usosemcompleto]
]
]

```

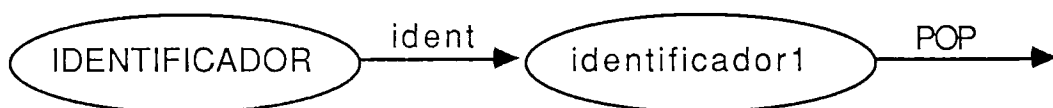
```

[card2
  [pop [buildq [[card [ident [* [sem +]]]]] izarsem]
    [usosemcompleto]
]

```

ATN honen zeregina izen-sintagma batean zenbakizko adjektibo kardinal bezala erabilitako zenbaki edo identifikadore bat prozesatzea da. Aplikazioaren eremuak zenbaki osoak izan litezkeen objektuak erabilzen direnez, ulergarria da kardinala bere izen-sintagmak errepresentatutako objektutik berezi nahi izatea. Kardinal hiztegian sarrera duen zenbakia baldin bada "card" kategoriarekin, berari dagokion erabilera semantikoak konstante bezala errepresentatzeko asertzio bat eranstea eragingo du. Kardinalaren lekuan identifikadore bat agertzen bada, POSSEM funtzioa izen bereko objektu errepresentatu bat bilatzen ahaleginduko da, eta aurkituz gero, asertzio berririk sortu gabe objektu horretarako erreferentzia itzuliko du. Halako objekturik aurkitu ezean, berri bat sortuko da identifikadorea izentzat hartuz eta elementu simple osoa motakoa definituz [elem ent]. Erabilera semantikoa askatu aurretik determinazio eta numerorako determinatua eta singularra balioak definitzen dira, erreferentzia zehatz bat da eta.

5.2.3.3.10. Identifikadoreen ATNa



```

[IDENTIFICADOR
  [cat ident t
    [setr ident *]
    [possem [buildq [identident *]] ]
    [to   identificador1]
  ]
]
[identificador1
  [pop [getr ident]
    [usosemcompleto]]]

```

ATN honek identifikadore bat prozesatzen du, identifikadore honek izen-sintagma bati deitura ematen dionean, adibidez: "...una secuencia S de 20 elementos ...". Argitu beharra dago erabilera hau ez dela identifikadore baten beste bi erabilera sintaktiko-semantikoen parekoa. Lehenik, identifikadore bat izan bait daiteke izen-sintagma baten osagai bakarra, 5.2.3.3.3 atalean azaltzen den bezala, adibidez: "...una subsecuencia de S...". Identifikadore baten beste erabilera posiblea zenbakizko kardinal baten zeregina betetzea da. Oraingoan ordea, izen-sintagman deskribaturiko objektuaren izena izango da. Era horretan definituko duen ekintza semantikoa ezingo da hiztegi-sarrerera baten barruan egon, identifikadoreentzako sarrerak ez bait daude bertan. Hartu den ebazpidea POSSEM ekintzari [identident] kasu sintaktikoarekin dei egitean datza. Ekintza honek jabetzat edozein objektu (hori beti beteko da) eskatzen duen propietate gehigarri orokorrezko erabilera semantikoa sortuko duelarik. Ebazpide honen betebeharra [hartu-modif-orok] ekintza IDENTIFICADOR delakoari dei egiten dion arkuaren barruan definiturik egotea da.

6. LORTUTAKO EMAITZAK.

6.1. TRATAMENDU SINTAKTIKOAREN EBALUZAZIOA.

6.2. TRATAMENDU SEMANTIKOAREN EBALUAZIOA.

6.3. TRATAMENDU PRAGMATIKOAREN EBALUAZIOA.

6.4. EBALUAZIO OROKORRA.

6.4.1. CAPRA projektuan azaldutako helburuak.

6.4.2. Enuntziatu-sailkapena.

6.4.3. Ebaluazioa.

6. LORTUTAKO EMAITZAK.

6.1. TRATAMENDU SINTAKTIKOAREN EBALUAZIOA.

Definitutako analisatzaile sintaktikoak perpaus erazagutzaile eta agintzailez osotaturako enuntziatuak prozesatzen ditu. Izen-sintagmek, aditzek, preposiziodun taldeek, adberbioek eta atributuek osatzen dituzte perpausak. Erabilitako aditzak iragankorrak, iragangaitzak edo kopulatiboak izan daitezke.

Problemaren eremua objektuen deskribapena denez gero, analisatzailearen definizioarako egindako ahaleginak izen-sintagmara zuzendu dira, objektuen deskribapenearako gehien erabiltzen den osagai sintaktikoa da eta. Beste domeinu batzutan gertaerazko narrazioek edota ekintza-sekuentziek osatzen dute testua, horrelako eremuetan perpausaren definizioak garrantzi haundiagoa hartzen du, baina ez gara hortaz ari.

Eremu mota honetan duen garrantzia kontutan hartuta, izen-sintagmarako implementatu den analisi sintaktikoa erabatekoa da, ondoko osagaien prozesamendua ahalbidetuz:

- a) Adjektiboak. Izenaren aurretik edo atzetik ager daitezke. Adibidez:

elemento par
el mayor elemento

- b) Adjektibo kardinalak. Zenbaki bezala edota zenbaki bat adierazten duen identifikadore bezala aurkeztuak. Adibidez:

20 números
M letras

Kardinalen analisisa posible egiten duen arkuak erraz zabal daitezke ordinalak ere onar ditzan. Nahikoa litzateke kardinalarekiko arku paralelua gehitzea, baina "ordinal" kategoria sintaktikoko hitz bat eskatu behar da kasu honetan.

c) Adjektibo-taldeak. Adjektibo konparatiboek, edo preposiziodun taldeez edo adberbioez osatutako adjektiboek osaten dituzte. Adibidez:

un entero múltiplo de 3
número divisible por 7
de longitud mayor que M

d) Erlatibozko perpausak. Adibidez:

las comas que tiene dicha secuencia
la secuencia en la que aparece el carácter K

e) Aurre-determinatzaileak. Sintagmaren hasieran kokaturik, zenbatzailearena egiten dute normalki. Adibidez:

todos los números pares

f) Determinatzaileak. Hauek taldearen determinazioaren balioa zehazten dute. Artikuluak, adjektibo erakusle edo posesiboak dira. Adibidez:

todos los números pares
esta secuencia
su media aritmética

g) Identifikadorea. Izen baten ondoan edo bakarrik (izen-sintagma baten osagai bakarra bezala) ager daiteke. Adibidez:

un número X de 5 cifras
la longitud de S

h) Esprezioa. Izen-sintagma osatzen duen espresio matematikoa. Adibidez:

de $(l+1)*(l+2)/l$

Sistemaren muga sintaktikoak ondoko osagaiei dagozkienak dira:

a) Konjuntzioak.

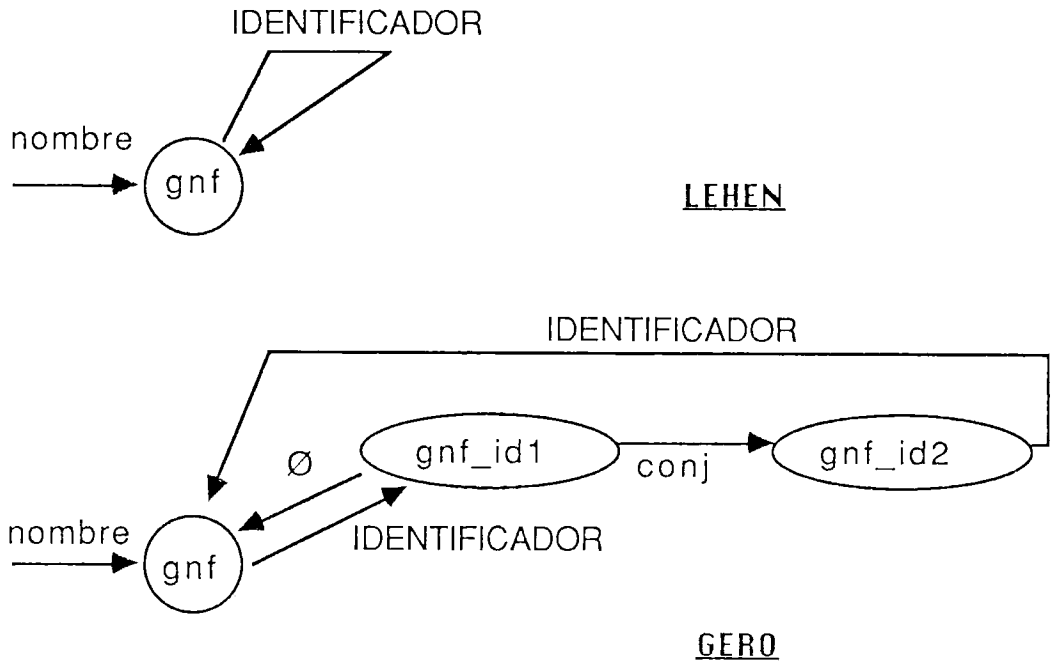
Konjuntzioen erabilpenak ulermen automatikoari oztopo latzak sortzen dizkio, analisirako sortzen dituen aukera-kopuru handiarengatik. Konjuntzioek sortutako anbiguitasuna ezin daiteke ebatzi ezagumendu sintaktiko soilaren bidez, sarritan beste ezagupen guztien erabilpena beharrezkoa egiten baita. Konjuntzioek sor litzaketan analisi-aukera guztiak modelatzeak definitu behar den ATN-kopururako gehikuntza izugarria ekarriko luke. Problema hau ebazteko soluziorik egokiena ondokoa dateke: konjuntzio bat agertzean dinamikoki zabaltzen diren ATNen erabilpena [Boguraev 83]. Beste soluzioa apalagoa baina epe laburrerako eraginkorragoa konjuntzioen erabilpen sinpleak soilik ukitzea da. Adibidez:

Dados dos números X_1 y X_2 , ...

Obtener...y calcular ...

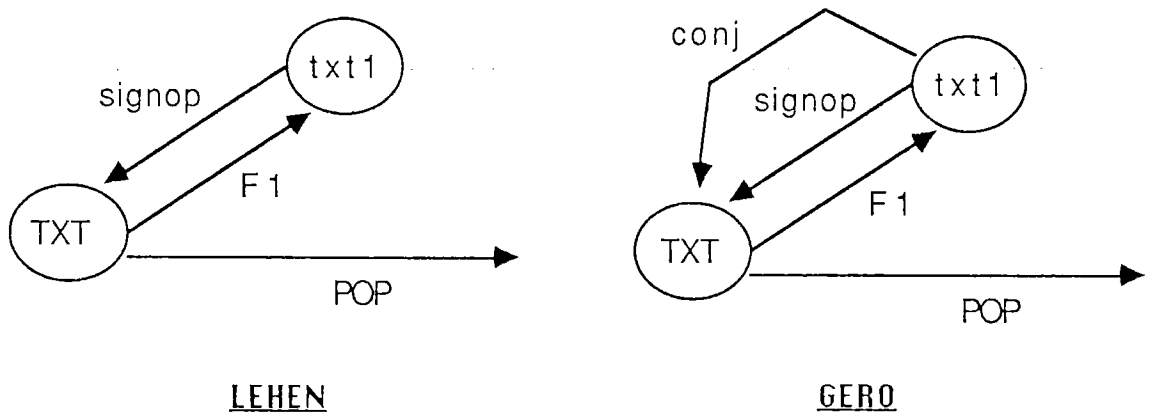
Un programa que lea 10 números enteros y escriba su suma un natural par y que pertenezca a S

Lehenengo adibidea sinplea da. Sistemari egindako hobekuntzetan espezifikoki onartzen da izen-sintagmaren ATNari bi korapilune gehituz (IDENTIFIKADOREA arku soila erabili beharrean). Honela "y" konjuntzioaren bidez bilduta datozen bi identifikadore hartu ahal daitezke. Ikus 6.1. irudia.



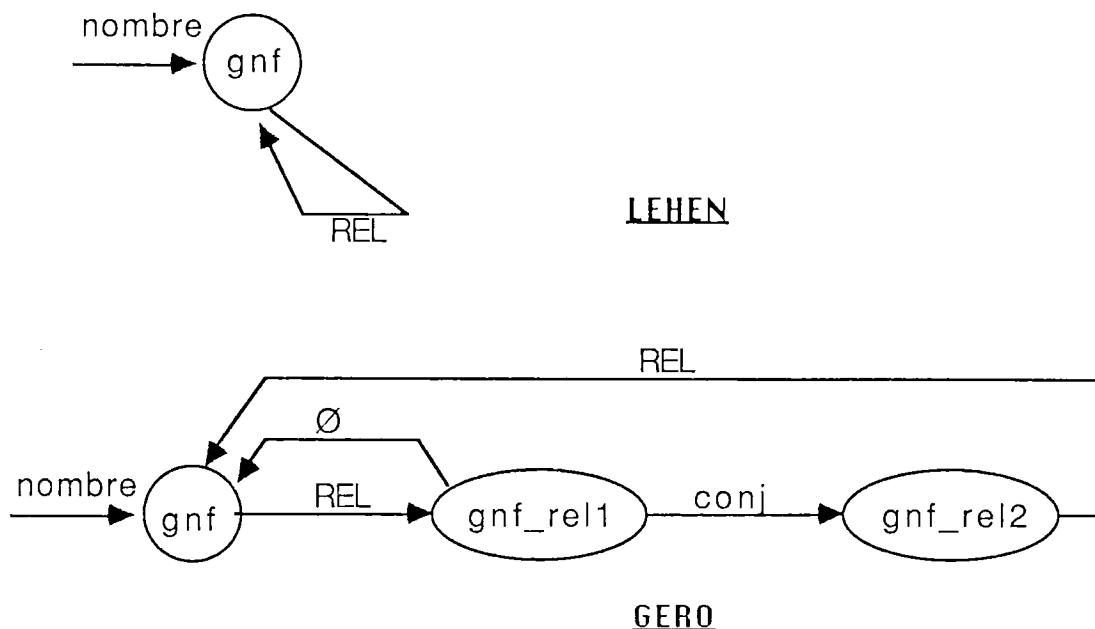
6.1. Irudia. Bi identifikadore konjuntzioren bidez loturik

Bigarren adibidea erraza da perpaus biak osorik agertzen badira, hau da, informazio eliptikorik ez badago. Bi perpausetan agertzen diren osagarri komunak ez errepikatzeagaitik baten bat kenduko balitz, analisia eta interpretazioa zailak lirateke. Horrelakorik gertatzen ez deneko kasua erraza da, perpausak banatzeko puntuazio-marka eskatzen duen arkoari beste arko paralelo bat gehitzen zaio konjuntzioa har dezan. Ikus 6.2. irudia.



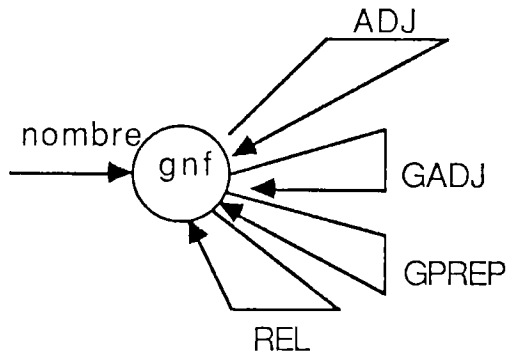
6.2. Irudia. Bi perpaus konjuntzioren bidez loturik.

Hirugarren adibidea erraza da, zeren eta konjuntzioak bi erlatibozko perpaus osoak biltzen bait ditu. Kasu honetan ere ez da kentzen osagarri errepikaturik. Izen-sintagmaren ATNari bi korapilune gehiago gehitzea nahikoa litzeteki. Ikus 6.6. irudia.

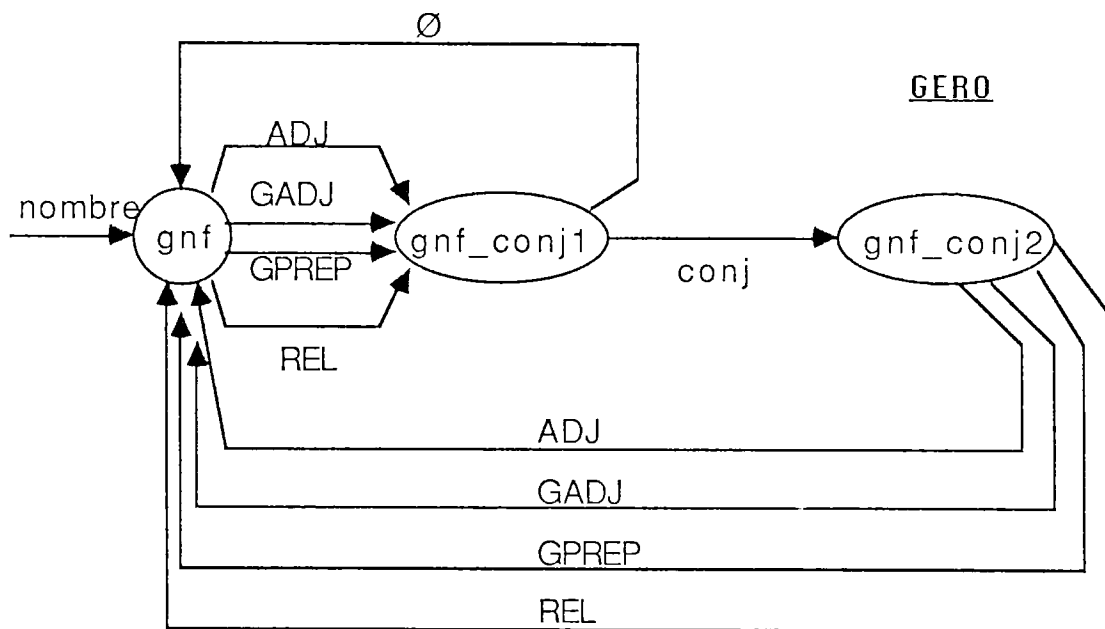


6.3. Irudia. Bi erlatibozko perpausak konjuntzioren bidez loturik.

Laugarren adibideak adjektibo eta erlatibozko perpaus bat konjuntzio baten bidez loturik aurkezten ditu. Posibilitate honek ebazpide orokorrago baten beharra dakar konjuntzioren bidez lotutako edozein bi osagarriren kasurako. Identifikadorea ez da tratatzen beste osagarrien modura, ez bait daitezke behin baino gehiagotan ageri. Ikus 6.4. irudia



LEHEN



6.6. irudia. Izen-sintagmaren osagarri desberdinak konjuntzioren bidez loturik.

b) Ukapena.

Osagarri desberdinetako perpaus batean ukapen bat agertzen denean era askotako interpretazioak egin daitezke [Tennant 81], ukatzailearen posizioarako aukerak ere ugariak direlarik. Anbiguotasun honek gogorki korapilatzen du ukapenaren prozesamendua. Predikatu batzuentzat bere ezezko predikatu bikia

definituz soluzio erraza inplementatu da. Predikatu hauen interpretazioa eremuari dagozkion erregelak aplikatu arte atzeratzen da. Baina soluzio hau ez da orokorra. Ukapenaren adibideak:

Adjektiboetan : número no primo
Perpausetan : que no acabe en 7

Lehenengo kasua ebazteko "no_primo" propietatea definitu da.

c) Ekuazio eta espresioen lorpena.

Analizatzaile lexikoak ekuazio eta espresioak ateratzeko gai dela eman dugu, baina batzutan lan hau oso zaila izan daiteke. Adibidez:

"Calcular W_n definido $W_n = V_n / V_{n-1}$ donde

$$V_0 = 1$$

$$V_1 = 2$$

.

.

.

$$V_n = V_n / V_{n-1}$$

el elemento n es mayor que 0"

Espresio-multzo hori ulertu ahal izateko, ezagumendu sintaktikoa, semantiko eta eremuari buruzkoa beharrezkoa dira. Idazkerari buruzko ezagumendua ere premiazkoa da, konkretuki "lerro-bukaera", "tabulazio", "zurigune", "eten-puntuak" eta abar karaktereen esannahiei buruzkoa. Orohar, testu-enuntziatuaren espazio-banaketa interpretatzen jakitea beharrezkoa da.

6.2. TRATAMENDU SEMANTIKOAREN EBALUAZIOA.

Tratamendu semantikoa sintaktikoarekin tartekatzen da eta betebeharrak bi betetzen ditu:

- Koherentzi semantiko gabeko analisi sintaktikoen posibilitateak ahalik eta lasterren baztertea.
- Erabilitako testuaren esanahia adierazten duen asertzioen multzoa sortzea.

Unitate sintaktiko-semantikoen nukleoek beren definizio posible guztiak askatzen dituzte, definizio bakoitzerako bere betebeharreko edo gehigarrizko osagarriak eskatuz. Osagarri sintaktiko guztiei definizio semantikorik ez dagokienez gero, ATN guztiek ez dute tratamendu semantikorik.

Betebeharreko eta gehigarrizko eskaerek mota sintaktiko bat (nukleo eta osagarrien arteko erlazioarena) eta mota semantiko bat (kategoria semantikoarena) eskatuz gain aztergaiari buruzko bestelako baldintzen egiaztapena ere eska dezakete.

Sintaktikoki posible den analisi-aukera bat ondoko edozein hiru arrazoigaitik baztertua izango da :

- Egungo definizio semantikoaren betebeharreko eskaera guztiak bete ez badira analisi sintaktikoa bukatutzat ematen denean.
- Osagarri bat egungo definizio semantikoaren barruan ezin bada sartu, ez betebeharreko bezala ezta gehigarri bezala ere.
- Osagarriak erlazionatzen dituzten baldintza globalak ez badira egiaztatzen, nahiz eta eskaera guztiak bete.

Objektuen ezaugarriak definitzen dituzten propietateak testutik ateratzeko hiru bide desberdin diseinatu dira:

- a) Definizio sintaktiko-semantiko baten barruan betebeharreko eskaera bezala definitutako propietatea.

- b) Definizio sintaktiko-semanticoren baten barruan eskaera gehigarria baina ez betebeharrekoa bezala definitutako propietatea. Eskaera gehigarriaren definizioa nukleo osatuari dagokion hiztegiaren sarreran agertzen da.
- c) Eskaera gehigarri orokorra bezala definitutako propietatea. Bere aplikazioa oso orokorra denez, hiztegiko jabe posible guztientzat definitzea luzea eta inefikaza suertatuko litzateke. Propietatearen nukleoaren definizioan eskaera gehigarri orokor bat sartzen da propietate horren jabe posible batek bete behar dituen baldintzak zehaztuz.

Asertzio-multzoa gehikuntzazko eratan sortzen da. Bere tinkotasuna eta osotasuna testu osoaren analisia egin eta gero soilik egiaztatzen da. Hala eta guztiz ere, elipsi eta erreferentzi kasuak ebazteko, analisiaren prozesua bukatu baino lehen ere erabil daiteke asertzio-multzoa.

Elipsi-kasuak ondoko eran ebazten dira: Unitate sintaktiko-semanticoren baten analisiaren bukaeran aukeratutako definizioaren betebeharreko eskaera bat bete gabe geratzen denean, aurretik ageritako objektuen artean eskaeraren kategoria semantikoko objektu bat bilatzen da. Bat baino gehiago existitzekotan, objektuen errepresentazioan sartu den azkena hautatzen da. Hautatze-funtzio hauek modu orokorrean diseinatu dira, etorkizunean asma litezkeen erizpide hobeen inplementazioa errazagoa izan dadin.

Erreferentzien ebazpena ondoko tratamenduetan bereiz daiteke:

- a) Identifikadoreak. Identifikadorea izen-sintagmaren osagai bakarra bezala eta identifikadorea izen-sintagmaren kardinal bezala modu desberdinean hartzen dira. Kasu bietan, izen bezala identifikadore bera duen aldeztu aurretik ageritako objektu bat itzultzen ahalegintzen da, baina objektu hori existitzen ez bada izentzat identifikadorea izango duen objektu berri bat sortuko da. Bigarren kasuan, identifikadoreak kardinal bezala agertzen danean, [elem ent] elementu osoen kategoria semantikoa hartuko liteke objektu

berriarentzat; lehenengo kasuan ordea, edozein objektu-mota adieraz dezaken [obj] kategoria orokorra hartuko liteke.

- b) Adjektibo posesiboak eta izenordeak. Elipsiaren antzera ebazten dira, hau da, bilatutako kategoria semantikoari dagokion sortutako azkenengo objektua aukeratuz.
- c) Izen-sintagma determinatu singularrak. Objektua errepresentatzen duten asertzioen sorkuntza burutu baino lehen, aztertu egiten da ea izen-sintagma aldeztatik aipatutako beste objektu baten erreferentzia izan zitekeen. Horrela balitz beraren erreferentzia itzuliko litzateke, bestalde objektu berri baten errepresentazioa sortuko litzateke.

Erreferentzia eta elipsien tratamenduaren adibide bezala, enuntziatu berberaren gainean egindako ondoko aldaerak erakus ditzakegu:

Dada una secuencia S de 20 números enteros, obtener otra secuencia con los elementos pares de S . Determinar su longitud.

Dada una secuencia S de 20 números enteros, obtener otra secuencia con los elementos pares de S . Determinar la longitud de dicha secuencia.

Dada una secuencia S de 20 números enteros, obtener otra secuencia con los elementos pares. Determinar la longitud.

Dada una secuencia S de 20 números enteros, obtener sus elementos pares. Determinar la longitud.

Dada una secuencia S de 20 números enteros, obtener la longitud de la subsecuencia formada por sus elementos pares.

Objektu-deskribatzaileak deskripzioaren objektu jabearen ausentziaz ere erabili ahal dira. Adibidez:

" ... una secuencia de 20 números ... "

" Dados 20 números ... "

Lehenengo adibidean "20 números"-ek sekuentziaren luzera eta mota oinarria deskribatzen du. Bigarren adibidean ez da agertzen inplizitoki kontutan hartzen den sekuentziaren erreferentzirik. Sortutako sistemak adibide bietan ezagutzen du sekuentzia.

Ondoko problemak tratatzeko ezintasunak definitzen ditu sistemaren muga semantikoak:

- a) Unitate sintaktiko-semantiko baten nukleoaren elipsia. Adibidez, "número" izena, zeina ondoko izen taldearen nukleoa bait da:

... número ... calcular otro que tenga tres cifras.

Problema honen ebazpenak nukleoa prozesatzen duen arku alternatibo baten definizioa exijituko luke. Arku berri honek aldeaz aurretik ageritako nukleoak aztertuko lituzke, beren arteko egokiena nukleo eliptikoa bezala aukeratzeko. Aukeraketa-prozesuan nukleo eliptikoa aldarazten duten osagarriak kontutan hartu beharko lirateke, baita osa litezkeen burutu gabeko eskaerak ere.

- b) Zenbatzaileak, izen aurreko adjektiboak eta kardinal txikiak (2 eta 3) sintaktikoki analisagarriak izan arren, ez da definitu berentzako tratamendu semantikorik. Ukapenak, sintaktikoki tratatzen ez zirenak, semantikoki ere ez dira tratatzen. Adibidez:

la diferencia entre cualesquiera dos elementos.

entre cada número y el promedio.

ningún número par.

los veinte primeros elementos que sean ...

la siguiente formula.

Dados dos enteros ...

que no acaben en 7.

Izen-sintagmaren analisia bukatu denean kardinal txikiak, 2 eta 3 bezalakoak, kasu berezia bezala trata litezke. Objektu berri baten sorkuntzarako ekintza semantikoak bi edo hiru aldiz exekuta litezke beste hainbeste objektu sortzeko.

Izen aurreko adjektiboak adjektibo gehigarri orokorrak bezala trata litezke, baina nukleo izena heldu arte itxarongo lukete beren esanahi semantikoa gehitzeko. Batzutan izen-sintagmaren beste osagarriek sortzen dituzten ekintza semantikoak alderazten dituzte.

Zenbatzaileak gorde egin litezke ekintza semantikoak egikaritu arte, eta une horretan, dagokien zenbatze-erizpidearen araueraz ekintza semantikoak alda litzakete, berauek egikaritu baino lehenago.

6.3. TRATAMENDU PRAGMATIKOAREN EBALUAZIOA.

Eremuari dagokion ezagumendua bi puntu nagusietan bildu da: alde batetik hiztegian (analisi linguistikoa arrakastaz egiten denean, bete behar diren ekintzetan mamitzen da ezagumendu hori), eta beste aldetik analisi linguistikoaren bukaeran aplikatutako erregela-multzoan (eraikitako asertzio-multzotik espezifikazioa lortzeko erabiltzen diren erregela-multzoa).

Hiztegiko ekintzek, eremuko objektuen errepresentazioa lortzeko aukera ematen dute izen, mota eta bestelako propietateen arauera. Hiztegiko sarreretan definitutako ekintza hauek, orokorki definitu daitezke (oso abstraktuak diren objektu-motentzat, eta propietate gehigarri orokorrentzat), baina definitu daitezke baita ere, oso kasu berezientzat (objektu-mota murrizentzat, eta betebeharreko propietateentzat). Objektu-motetan abstrakzio-mailarekin eta osagarrietan exijentzi-mailarekin jokatzuz, malgutasun handia lor daiteke errepresentaziorako ekintzen definizioan. Malgutasun honek ezagumendu pragmatikoaren erabilpenari laguntza handia ematen dio.

Analisi linguistikoa bukatutakoan aplikatzen diren erregela-multzoak, zenbait tratamendu egiteko aukera ematen du.

- a) Nukleo-izena faltatzen denean, agertzen diren objektuaren deskribatzaileak objektu bezala kontsideratzea. ("Dados 20 enteros ...", "Dada una secuencia de 20 números enteros ..." -ren ordez)
- b) Zenbait predikatuen definizioaren osotasun- eta tinkotasun-egiaztapena.
- c) Objektuen sailkapena ondorengo hiru kategorietako batean: sarrera, irteera eta ezsarrerazirteera.
- d) Izena egokitzea definitu gabe duten objektuei. Ahal den neurrian egokitzen den izena eta objektu mota erlazionatuta egongo dira.

- e) Objektu berri bat definitzen duten identifikadoreentzako tratamendu bereziak. Normalki objektuari egokitzen zaion mota, abstraktuena izaten da ([obj]), baina identifikadorea kardinalaren posizioan agertzen baldin bada, egokitu behar zaion mota, zenbaki osoa da ([element]).

Eremuari buruzko ezagumenduaren erabilpenaren mugak ondorengo zailtasunak tratatzeko ezintasunaren arauera definitzen dira:

- a) Funtzio berrien definizioa enuntziatuaren textuaren barruan. Adibidez:

"Escribir los coeficientes de la fila N del triángulo de Pascal. La función

$$K(i, j) = K(i, j - 1) + K(i - 1, j).$$

calcula los coeficientes de la fila i, siendo

$$K(i, 1) = i \text{ eta } K(1, j) = 1 "$$

Oztopo honen ebazpenak, kategoria semantiko berrien definizioa inplikatu luke, baita funtzioen definizioaren egokitasuna egiaztatzeko eta eraikitze ezagumenduaren erabilpena ere. Gainera arrazamendu prozesu bat beharko litzateke, kalkulatu behar den emaitzaren eta funtzio laguntzailaren artean dagoen erlazioa aurkitzeko. Testuan eskatzen diren emaitzak lortzeko asmoz, funtzioa nola erabili behar den asmatzeko aukera emango luke arrazamendu honen erabilpenak.

Aipatutako enuntziatuaren ebazpenean, lehendabizi K funtzioaren errepresentazioa lortu beharko luke, izena, parametroak, eta definizio errekursiboaren kasu tribiala eta orokorrak ateraz. Bigarrenez, kalkulatu behar diren elementuak Pascal triangularen N. lerroaren koefiziente guztiak direla atera beharko luke:

$$\begin{pmatrix} N \\ 0 \end{pmatrix} \quad \begin{pmatrix} N \\ 1 \end{pmatrix} \quad \dots \quad \begin{pmatrix} N \\ N \end{pmatrix}$$

eta azkenik, ezagutu behar da koefiziente guztiak kalkula daitezkeela K funtziorako ondorengo deiak eginez: $K(N,1)$, $K(N,2)$, ..., $K(N,N)$.

- b) Beste aplikazio-eremuei buruzko problemak. Eremu berriei buruzko ezagumendua behar da.

"Dado un complejo en forma de par ordenado (parte real, parte imaginaria) expresarlo en forma polar (módulo, argumento)."

"Dado un grafo orientado representando un conjunto P de puntos y el conjunto A de arcos. Encontrar los puntos aislados."

Enuntziatu-testu hauetan sisteman errepresentatu behar diren konzeptu berri asko erabiltzen dira. Zenbait eremuetan (adibidez zenbaki konplexuetan) oso erraz bereganatu daitezke, konzeptu berri asko ez daudelako edota kontzeptu berriak, inplementatutako objektu orokorren motekin berehala errerepresentatzen direlako. Baina beste zenbait eremuek (adibidez grafuak) konzeptu berri askoren definizioaz gain (grafu, korapilune, arku, puntu isolatu, ...), objektu-egitura berrien sartzeari eskatzen dute (grafu bat ezin daiteke sekuentziaz, multzoz, edo elementu bakunen bidez zuzenki adierazi). Kontutan hartu behar da egitura berrien definizioak aldaketa sakonagoak dakartzala, kategoria semantiko berriak, beste kategoriekiko bateragarritasun-erizpideak, eta osotasun- eta tinkotasun-erregelak definitu beharko lirerateke eta.

- c) Ekuazioen ulerkuntza eta espresioak. Adibidez:

"Calcular el término W_n definido $W_n = V_n / V_{n-1}$, donde

$$V_0 = 1$$

$$V_1 = 1$$

...

$$V_n = V_{n-1} + V_{n-2}$$

El elemento N es mayor que 0."

Oztopo hau muga sintaktikoez hitz egiterakoan aztertu bada ere, esan beharra dago oso erlazio estua daukala eremuari buruzko ezagumenduarekin. Formularen kokapen espazialari buruzko ezagumenduaz aparte (enuntziatuan agertzen den formula bakoitza ulertzeko balio duena), formula-multzoaren interpretazio bateratua lortzeko, beste ezagumendu-iturriekin kontaktzea beharrezkoa izango litzateke. Aurkeztutako adibidean, arrazonomendu analogikoa erabiltzea beharrezkoa izango litzateke formula guztien zati amankomuna lortzeko, eta beste aldetik, arrazonomendu induktiboa ere planteatzen den definizio errekursiboa ulertzeko.

6.4. EBALUAZIO OROKORRA.

6.4.1. CAPRA projektuan azaldutako helburuak.

CAPRA projektua adjudikatzeko aurkeztutako txostenean, helburuak era honetan definitzen ziren:

"Projektuaren asmoa, Konputagailuz Lagunduriko Irakaskuntza Adimentsurako balio duen sistema baten eraikuntza da, bere aplikazio-eremua programen eraikuntza delarik. Beraz, sistema ondorengo gauzak egiteko gai izango da:

- 1.- Enuntziatu batetik abiatuz, espezifikazio eta programa bat eraikitzea .*
- 2.- Eraiketa prozesua sistematikoki azaltzea.*
- 3.- Programa baten zuzentasuna aztertu, ea bere enuntziatuari dagokion.*
- 4.- Problema baten ebazpenean ikasleak egiten dituen erroren kausa posibleak arrazonatu.*
- 5.- Programa bat eraikitzen ikasleari lagundu.*
- 6.- Erabiltzailearekin Lengoia Naturelez elkarrizketan aritu.*
- 7.- Lengoia Naturelez problemen enuntziatuak sortu.*

Sistema, tutore batez, aditu batez, itzultzaile batez eta datu-base batez osaturik dago."

Deskribatzen dugun CAPRATE azpisistema problema-enuntziatuen ulerkuntzarako balioko duten tresnak eraikitze asmoarekin izan zen diseinatua, enuntziatuen itzulpenak eginez Lengoia Natureletik CAPRA-ren espezifikazio lengoiara. Lan hau burutu ondoren, lehendabiziko helburuaren ("enuntziatu batetik abiatuz, espezifikazio bat eta programa bat eraikitzea") lehendabiziko fasea egingo litzateke. Oso lagungarria izango litzateke baita ere, problema-bildumen ugaltze lanean.

CAPRATE-k enuntziatu baten interpretazio guztiak lortu ditzakeenez, enuntziatu berrien anbiguetatea aztertzeko laguntza bikaina izango da

Erabiltzailearekin Lengoia Naturalaz elkarrizketan arituko den CAPRA-ren modulu-itzultzaileak ezagumendu linguistikoaren osagaiak erabili ahal izango ditu. Itzultzailearen hiztegiak, CAPRATE barruan espezifikazioak lortzeko bildu diren termino guztiak izango ditu, eta gainera algoritmo eta programak eraikitzeko behar direnak. Elkarrizketetako azpilengoaia gramatikalki ez da izango enuntziatuetan erabiltzen dena bezain hertsia. Perpaus sekuentzi bat aztertzen duen ATNak (TXT), eta perpaus bakarra aztertzen duenak (F1), erabilgarritasun gutxienekoak izango dira, elkarrizketetako lengoiaren egitura informalagoa delako, baina elementu sinpleagoak definitzen duten beste ATN guztiak erabilgarriak izango dira. ATNaren interpretatzailea guztiz baliagarria da, nahiz eta agian ekintz semantikoak eta ezagumenduaren errepresentazio eredua desberdinak izan.

6.4.2. Ulerkuntz oztopoekiko enuntziatuen sailkapena.

100 enuntziatuz osaturiko programazio-problemen lagina definitu zen, proiektuaren modulu desberdinen diseinurako erreferentzi gisa balio duelarik. Aurkezten den enuntziatu-itzultzaile ebaluatzeko asmoaz, 100 enuntziatu horiek prozesatu direnean lortu diren arrakasta-maila aztertu da. Azterketa horretan aurkitu diren oztopoak kontutan harturik, enuntziatuen sailkapen bat egin da.

Sailkapenean bi bloke haundi bereiztu dira, batean (beren oztopoak oso zailak ez direnez) epe motzean ebazkarriak direnak sartuz, eta bestean epe luze batean ebatziko direnak. Bi talde handi hauen sailkapena, analisi-oztopo konkretuen arauera egiten da. Azpisaillkapen hauek ez dira disjuntu.

Epe motzean ebazkarriak diren oztopoak ondorengo hauek dira:

- a) Osagai osoak koordinatzen duten konjuntzioak, hots, koordinatutako osagarrien artean bat ere kendu egin ez

denean. Ikus adibideak eta proposatutako soluzioak 6.1.a atalean (8 enuntziatu).

- b) Ezezko eran enuntziatutako adjektiboak. Ikus adibideak eta soluzioak 6.1.b atalean (3 enuntziatu).
- c) Unitate sintaktiko-semantiko baten nukleoaren elipsia. Ikus adibidea eta posible den soluzioa, 6.2 atalean (4 enuntziatu).
- d) Izen aurreko adjektiboak. Adibidez: "los veinte primeros elementos que sean ... " (5 enuntziatu)
- e) Aplikazio eremu berriak, baina errepresentatzeko errazak direnak. Adibidez, hitzak letraz osaturiko sekuentzia bezala konsideratzerakoan sortzen diren problemak, era berean zenbakiak digitu-sekuentzia bezala, zenbakiak zortziko oinarrian, zenbaki konplexuak eta geometriako zenbait problema errazak (11 enuntziatu).

Epe luze batean ebazkarriak direnak, ondorengo hauek dira:

- a) Koordinatzerakoan elementu bat kendu duten konjuntzioak. Ikus 6.1.a (20 enuntziatu).
- b) Ezezko eran enuntziatutako perpausak. Ikus 6.1.b. (4 enuntziatu).
- c) Espresio eta ekuazioen ulerkuntza. Ikus 6.1.b eta 6.3.c (10 enuntziatu).
- d) Objektu anizkunen erabilpena. Adibidez:
"Escribir una función que dados 3 números enteros devuelva un valor verdad si éstos van en secuencia ascendente y falso en cualquier otro caso".
Ikus 6.3.b (12 enuntziatu).
- e) Zenbatzaile eta beste aldarazleak. Ikus 6.3.b. (6 enuntziatu).
- f) Aplikazio eremu konplexuak. Adibidez: Grafuak (grafuei

buruz 8 enuntziatu daude), segida aritmetikoak, erakusleak, geometria, azterketa-notei buruzko estatistikak, taula moduko irteerak, demografia. (20 enuntziatu).

g) Beste kausa batzuk. (10 enuntziatu).

6.4.3. Ebaluazioa.

Lehenengo sailkapena, eta talde desberdinen arteko ebaketa kontutan hartuz, ondorengo sailkapen disjuntu lortzen da:

- a) 20 enuntziatu eremu oso desberdinekin dute zerikusirik. Enuntziatu hauek, planteatutako asmo nagusiekin zerikusirik ez duen azterketa sakon bat eskatzen dute. Sistema ebaluatzeko, hasierako 100 enuntziatuetatik kanpo geratu daitezke aipatutako 20 hauek.
- b) 14 enuntziatu garatutako tresrekin ebazkarriak dira.
- c) 21 enuntziatu epe motzean ebazkarriak dira. Nahiko litzateke proposatutako soluzioak inplementatzea, pertsona batentzat hiru hilabetetako lan metodikoa suposatuz.
- d) 20 ebazkarriak dira epe luzeagoan. Bere ebazpena ATNen interpretatzailearen eta definitutako zenbait ATNen berdiseinu partziala suposa daiteke, baina inplementazioa egin daiteke gaur egungo diseinu orokorra mantenduz.
- e) 25en ebazpena zaila da, beraietan, eta era korapilatu batean, lehen azaldutako zenbait oztopo nahasten bait dira.

Laburtuz, sistema gai da lagin aleatorioaren %17,5 itzulpen osoa emateko. Tresnen garapen-lan baten bitartez, eta inolako diseinu aldaketarik egin gabe, itzulpenaren %42,75 lortu daiteke. Eta

azkenik zenbait aldaketa konplexu egin ondoren, zenbait kasutan sistemaren egitura ukituz, %67,75-raino hel liteke.

Gainera, kontutan hartzekoa da bildutako 100 enuntziatu ez direla antzekoak, bakoitzak enuntziatu posibleak diren multzo bat errepresentatzen dituelarik. Prozesatu daitezkeen 35 enuntziatuen atzetik, aldaera anitz aurki dezakegu.

ONDORIOAK

ONDORIOAK

CAPRATE Lengoaia Naturalaren Ulermenerako sistema da. Espaineraz idatzitako Programazioari buruzko oinarritzko problema-enuntziatuak aztertzen ditu, CAPRA sistemaren espezifikazio-lengoaia formalaren arauera dagozkien espezifikazioak lortzen dituelarik. CAPRA sistemaren barruan, CAPRATE azpisisetak problema-bilduma aberasteko modulu gisa jokatzeko du.

CAPRATE-ren diseinuak independentzia handiaz bereizten ditu aplikazio-eremuari lotutako ezagumendua eta hizkuntz ezagumendua. Beraz, erabilgarri litzateke sistema hau antzeko hizkuntz problemak azaltzen dituzten beste aplikaziotan, hain zuzen, objektuen deskribapenez eta beraien gaineko eragiketez edo harremanez oinarritzen diren aplikaziotan. Hala ere, aplikazio baten testuetan denborarako adierazpide zehatza behar duten gertakizun ugari agertzen bada, gure sistema ez da egokiena, ezta gertakizunen arteko erlazio klausulak aztertu behar ba dira ere. Antzeko aplikazio-eremuak beste jakintz-arlo bati buruzko problema-enuntziatuen ulermena eta matxura-konponketarako eskuliburuaren erabilpena izan daitezke. Hiztegiko informazio sintaktiko-semantikoaren egokipena eta tinkotasun-erregela globalen birdefinizioa beharrezkoak lirateke aplikazio-eremu berri batetarako zabalkuntza gauzatzeko, baina bai analisatzaile lexiko-sintaktiko-semantikoa, bai tinkotasun-erregelen interpretatzailea ezer ikutu gabe baliagarri lirateke. ATNek ez lukete garrantzizko aldaketarik behar azpilengoaia berria ezta ahal izateko.

Inplementatu den ATNen interpretatzailea Bates-ek [Bates 78] definitu zuen eredu klasikoaren hedakuntza da. Hedakuntzaren helburua tratamendu semantikoa gehitzea da. Tratamendu honetarako beharrezko den ezagumendua hiztegian definitzen da eta ez beste inon, bertako sarrera bakoitzerako posible edo beharrezko diren osagarriei buruzko eskaerak definituz. Eskaeren erabilpena ATNetan eransten diren ekintza eta test berrien bidez kontrolatzen da. Tratamendu semantikoak enuntziatuetan deskribatutako objektuen errepresentazio erazagutzailea eraikitzen du, eta horrez

gain, elipsia, erreferentziak eta anbiguotasun sintaktikoa ebazten ditu.

Mellish-ek [Mellish 85] garatu zuen "gehikuntzazko analisi goiztiar"-aren ideiak sartu dira CAPRATE barruan, baina tinkotasunezko baldintzak analisia bukatutakoan aplikatzen dira maila globalean (baita modu ez hain murrizgarriz izen-sintagma eta aditz-sintagma mailetan ere). Mellish-en sistemaren analisi linguistikoa formalismo logikoaren bidez gauzatu da. CAPRATE sisteman ATNen formalismoa erabili dugu analisi linguistikoa bideratzeko. Eskaera sintaktikoak definitu eta kontrolatzeko sistema ere CAPRATE sistemaren bereizgarria (eskaerak beharrezko, gehigarri eta gehigarri orokor gisa sailkatzen dira)

Sistema hedatzeko erakutsi diren lerro nabarmenenetan ondoko problemak azalduko lirateke: konjuntzioak, zenbatzaileak eta enuntziatuan bertan deskribatzen diren funtzioen ulermena.

Konjuntzioen tratamenduak sortzen duen analisi-aukera kopuru erraldoia nola edo hala mozteko aurkezten den soluzio tinkoena Boguraev-ena da [Boguraev 83]. Soluzio honen arauera konjuntzio bat agertzen denean ATNak dinamikoki zabaltzen dira. Hobekuntza honek ATNen interpretatzailearen aldakuntza suposatzen du.

Erreferentzi eta zenbatzaileen tratamendurako hobekuntzak Mellish-en deskribatutako tinkotasun- eta osotasun-erregelak aplikatzeko prozeduren bidez bideratu beharko dira. Erregelak etengabe aplikatu beharko lirateke analisia bukatu arte itxoin gabe.

Enuntziatuan bertan deskribatutako funtzioen ulerkuntza enuntziatuen egiturari buruzko ezagumendua erantsiz azaldu ahal zitekeen. Testuan ageritako objektuen errepresentazioa eraikitzeke erabiltzen den estrategia goranzkorra, osatua izango litzateke funtzio-definizioak, ekuazioak edo espresio konplexuak ulertzeko estrategia beheranzkor bat erantsiz. Espezifikazioa eraikitzeke fasean modulu tutoreak erabiltzen duen estrategia abiapuntu egokia litzateke estrategia beheranzkor hori diseinatzeko.

BIBLIOGRAFIA

BIBLIOGRAFIA

[Bates 75]

Bates, M.

The use of Syntax in a Speech Understanding System.

IEEE Transactions on Speech and Signal Processing, Vol. ASSP-3,
No 1, Feb. 1975 , 112-117.

[Bates 76]

Bates, M.

Syntax in Automatic Speech Understanding.

American Journal of Computational Linguistics, Microfiche 45,
1976

[Bates 78]

Bates, M.

The theory and practice of augmented transition network
grammars.

In L. Bolc (ed.) *Natural Language Communication with Computers*,
Berlin, Springer-Verlag 1978.

[Birnbaum and Selfridge 79]

Birnbaum, L. and Selfridge, M.

Problems in conceptual analysis of natural language.

Research Report No. 168, Dept. of Computer Science, Yale
University.

[Blackwell 81]

Blackwell, S. A.

Processing conjunctions in an ATN Parser.

M.Phil. Thesis University of Cambridge, 1981.

[Bobrow 68]

Bobrow, D. G.

Natural Language Input for a Computer Problem-Solving System,
In *Semantic Information Processing*, Minsky (ed.), MIT, 1968.

[Bobrow and Webber 80]

Bobrow, R. and Webber, B.

PSI-KLONE: Parsing and Semantic Interpretation in the BBN Natural Language Understanding System.

In Procs of the *CSCSI/CSEIO Annual Conference*, 1980.

[Boguraev 79]

Boguraev, B. K.

Automatic resolution of linguistic ambiguities.

PhD thesis. Computer Laboratory, University of Cambridge, TR 11 1979.

[Boguraev 83]

Boguraev, B. K.

Recognising conjunctions withing the ATN framework.

In Sparck-Jones, K. and Wilks, Y. editors, *Automatic Natural Language Parsing*. Ellis Horwood, 1983.

[Brown and Burton 74]

Brown, J. S. and Burton, R. R. and Bell A. G.

SOPHIE/ A Sophisticated Instructional Environment for Teaching Electronic Troubleshooting.

Report 2790, Cambridge Mas. BBN, 1974.

[Brown and Burton 75]

Brown, J. S. and Burton, R. R.

Multiple representations of knowledge for tutorial reasoning.

In Bobrow, D. C. and Collins, A. (ed.) *Representations and Understanding*. Academic press, 1975.

[Burton 76]

Burton, R. R.

Semantic grammar. An Engineereing Technique for Constructing Natural Language Understanding Systems.

Report no. 3453, Bolt, Beranek and Newman Inc., Cambridge, Mass, 1976.

[Charniak 68]

Charniak, E.

CARPS, A program which Solves Calculus Word Problems.

Report MAC-TR5, MIT, 1968

- [Charniak 72]
Charniak, E.
Towards a model of children's Story comprehension.
PhD thesis, MIT AI Laboratory. 1972.
- [Chomsky 65]
Chomsky, N.
Aspects of the theory of Syntax.
Cambridge, Mass: MIT Press,1965.
- [Christaller and Metzing 83]
Christaller, T. and Metzing, D.
Parsing interactions and a multilevel parser formalism based
on cascaded ATNs.
In Sparck-Jones, K. and Wilks, Y. editors, *Automatic Natural
Language Parsing.* Ellis Horwood, 1983.
- [Fernández et al. 88]
Fernández, I. Díaz, A. y Verdejo, M. F.
Diseño de un tutor automático mediante planificación dinámica.
Iberamia, 1988
- [Garijo and Verdejo 84]
Garijo, F.J. Verdejo, M.F.
Knowledge representation for teaching programming in an ICAI
environment.
First IEEE Conference on Artificial Intelligence Applications.
Denver. Colorado.1984.
- [Garijo et al. 86]
Garijo, F.J. Hernández, L. Intxausti, K. Verdejo, M.F.
SYNTALAB: an expert system for the synthesis of abstract
algorithms.
The 6th. Int. W. on Expert Systems & their applications. Avignon
1986.
- [Garijo et al. 87]
Garijo, F.J. Verdejo, M.F. Díaz, A. Fernández, I. Sarasola, K.
CAPRA : An intelligent System to Teach Novice Programmers.
II Congreso Mundial Vasco, 1987

[Gazdar 82]

Gazdar, G.

Phrase structure grammar.

In P. Jacobson and G. Pullum (ed.) *The Nature of Syntactic Representation*. Dordrecht: D. Reidel, 131-186, 1982

[Ginsparg 78]

Ginsparg, J.

Natural Language Processing in An Automatic Programming domain.

Stanford, Computer Science Dept., Alm-316, 1978

[Gómez 81]

Gómez, F.

Understanding Programming Problems Stated in Natural Language.

CISR-TR-81-9 Ohio State University, 1981.

[Hayes and Simon 74]

Hayes, J. R. and Simon, J. A.

Understanding Written Problem Instructions.

In P L.W. Greeno (ed.) *Knowledge and Cognition*

Lawrence Erlbaum Associates, 1974.

[Heidorn 72]

Heidorn, G. E.

Natural Language Inputs to a Simulation Programming System.

NPS-55HD72101A, Naval postgraduate School, Monterey, Ca., 1972

[Kaplan 82]

Kaplan, R.M. and Bresnan, J.

Lexical Functional grammar: a formal system for grammatical representation.

In J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press.

- [Marcus 79]
Marcus, M.
Overview of a theory of syntactic recognition for Natural Language.
Technical Report TR-531, Ai Lab, MIT, 1979.
- [Marcus 80]
Marcus, M.
A theory of syntactic recognition for natural language.
Cambridge, Mass.: MIT, 1980.
- [Mellish 83]
Mellish, C. S.
Incremental Semantic Interpretation in a Modular Parsing System.
In Sparck-Jones, K. and Wilks, Y. editors, *Automatic Natural Language Parsing.* Ellis Horwood, 1983.
- [Mellish 85]
Mellish, C. S.
Computer interpretation of Natural Language Descriptions.
Ellis Horwood Limited, 1985.
- [Novak 76]
Novak, G. S.
Computer Understanding of Physics Problems Stated in Natural Language.
AJCL Microfiche 53, 1976.
- [Riesbeck 75]
Riesbeck, C.
Conceptual analysis.
In R. C. Schank (ed.) *Conceptual Information Processing.*
Amsterdam: North Holland. 1975
- [Riesbeck and Schank 76]
Riesbeck, C. and Schank, R.
Comprehension by computer: expectation-based analysis of sentences in context.
Research Report 78, Dept. of Computer Science, Yale University, 1976.

[Schank 75]

Schank, R. C. (ed.)
Conceptual Information processing,
North Holland, 1975.

[Schank and Abelson 77]

Schank, R. C. and Abelson, R. P.
Scripts, plans, goals and understanding.
Lawrence Erlbaum Associates, 1977.

[Sebright 81]

Sebright, A. W.
Analysis and Inference for English.
PhD thesis. Cambridge, 1981.

[Tennant 81]

Tennant, H.
Natural Language Processing
Petrocelli Books. 1981

[Wilks 75]

Wilks, Y.
A Preferential Pattern-seeking, Semantics for Natural Language
Inference.
Artificial Intelligence 6: 53-74, 1975.

[Wilks 78]

Wilks, Y.
Making preferences more active.
Artificial Intelligence 11: 197-223, 1978.

[Winograd 83]

Winograd, T.
Language as a Cognitive Process. Vol I: Syntax.
Addison-Wesley 1983.

[Woods 70]

Woods, W. A.
Transition network grammars for Natural language analysis.
Communications of the ACM 13,591-606, 1970.

[Woods 72]

Woods, W. A.

The LUNAR sciences natural language information system .

Cambridge, Mass.: Bolt, Beranek and Newman, 1972.

[Woods 73]

Woods, W. A.

An experimental parsing system for transition network grammars

In R. Rustin (ed.) *Natural Language Processing.*

New York: Algorithmics Press, 1973.

[Woods 80]

Woods, W. A.

Cascaded ATN Grammars.

American Journal of Computational Linguistics 6:1-12, 1980.

ERANSKINAK

A ERANSKINA

Inplementazioaren ezaugarriak

CAPRATE sistema MICROVAX-II Q5 konputagailu batetan inplementatu da, POP11 lengoaiaz eta POPLOG programazio-ingurunean. Lengoia hau aplikazio honetarako egokia da, lista eta pattern matching-a erabiltzeko erraztasunarengatik. Gainera lengoia funtzional edo agintzaile gisa erabil daiteke, baita modu biak bilduz ere.

Enuntziatu baten espezifikazioa lortu ahal izateko, aurretik definitu beharko dira hiztegia, analisi sintaktiko-semantikoa gidatuko duen ATNen multzoa eta enuntziatu bera.

Hiztegia "hiztegia" aldagaiari asignatu behar zaio.

ATNen multzoa "network" aldagaiari asignatu behar zaio.

Problemaren enuntziatua "enunciado" aldagaiari asignatu behar zaio.

Tratamendua burutzeko behar diren funtzio guztien irakurketa eta interpretazioa hurrengo POP11-zko komandoaren bidez lortzen dira:

```
:load abiatzeko.p
```

Beharrezkoak diren funtzio guztiak sartu eta interpretatu ondoren "analisa" funtzioari hurrengo parametroekin deitzea nahikoa da:

1) Analisisatzeko enuntziatua.

2) Analisisirako abiapuntu izango den korapilunearen izena.

Adibidea: :analisa(enunciado, "txt");

Zenbait perpaus prozesatzen dituen ATNaren lehen korapilunearen izena "txt" dela jakinda. Problemaren enuntziatua

osorik idatz liteke lehenengo argumentua bezala, baina irakurgarritasunarengatik eta erroreak egiteko arriskuarengatik ez da gomendagarria. Hobe da aldagai bati edukizat asignatzea eta aldagaia bera argumentu bezala erabiltzea.

Analisiaren traza atera nahi bada ondoko aldagai global boolearren baloreak birdefinitzea nahikoa da:

traza: bere balorea "true" bada, arrakastaz prozesatutako arkuaren ezaugarrien balioak idatziko dira, arkuaren azterketa bukatutakoan. Behean azaltzen diren aldagaiak aipatutako ezaugarriekin elkartzen dira. Hauetako bat erakusteko, bai "traza" aldagaiak eta bai ezaugarriari dagokion aldagai partikularrak ere, biek "true" balioa eduki beharko dute.

trarc: prozesatutako arkuak.

trarc: egungo egoeratik prozesatzeko geratzen diren arkuak.

trhold: VIR arkuaren bidez sortutako erregistroak.

trregisters: analisatzen den ATNaren erregistroak.

trwords: prozesatzeko geratzen den enuntziatuaren zatia.

trreturns: "Push" arkuaren bidez sortutako zintzilik dauden itzulerak.

trstates: geroago prozesatzeko geratzen diren analisi-egoera baliagarriak.

trobj_bild: orain arte agertutako objektuak errepresentatzen dituen asertzio-multzoak

Erabiltzen den edozein funtzioaren traza ondoko POP11-zko komandoaz ere lortu ahal da:

```
:trace <funtzio-izena>;
```

"proces_arc" funtzioaren trazak arku bakoitzaren prozesaketa pausoz pauso ikertzea erraztatzen du, arrakastaz prozesatutakoena baita baztertuena ere. ATN berriak eraikitze-prozesuarentzat oso erabilgarria da traza hau.

B ERANSKINA

Objektuak erakusteko erabilitako predikatuak

[contenido X Y]

X Y sekuentziaren osagaia da.

[desalida X]

X objektua espezifikazioaren irteerakoa da.

[deentrada X]

X objektua espezifikazioaren sarrerakoa da.

[diferencia X Y Z]

X zenbakia Y eta Z zenbakien kendura da.

[equal X Y]

X eta Y-ren balioak berdinak dira.

[longitud X Y]

Y zenbaki osoa X sekuentziaren luzera da.

[longitud-inv Y X]

Aurreko predikatuaren kontrakoa da. Y zenbaki osoa X sekuentziaren luzera da.

[nombre X Y]

Y X objektuaren izena da.

[numelem X Y Z]

Z sekuentzia barruan Y deskribatzaileak bereizten dituen osagaien kopurua X da.

[primero_sec X Y]

X Y sekuentziaren lehenengo osagaia da.

[prop_cuadrado X Y]

X zenbakia Y zenbakiaren karratua da.

[prop_divisible X Y]

X osoa Y osoaz zatigarria da.

[prop_divisor X Y]

X osoa Y osoaren zatitzailea da.

[prop_factorial X Y]

X zenbakia Y zenbakiaren faktoriala da.

[prop_impar X]

X zenbakia bakoitia da.

[prop_máximo X Y]

X zenbakia Y sekuentziaren osagaien maximoa da.

[prop_mayor X Y]

X zenbakia Y zenbakia baino haundiagoa da.

[prop_menor X Y]

X zenbakia Y zenbakia baino txikiagoa da.

[prop_mínimo X Y]

X zenbakia Y sekuentziaren osagaien minimoa da.

[prop_múltiplo X Y]

X zenbaki osoa Y zenbaki osoaren anizkoitza da.

[prop_negativo X]

X zenbakia negatiboa da.

[prop_par X]

X zenbakia bikoitia da.

[prop_positivo X]

X zenbakia positiboa da.

[prop_primo X]

X zenbaki osoa lehena da.

[prop_productorio X Y Z T]
Y-tik Z-raino T zenbakizko espresioaren biderkaria da X zenbakia.

[prop_sumatorio X Y Z T]
Y-tik Z-raino T zenbakizko espresioaren batukaria da X zenbakia.

[prop_media_aritm X Y]
X zenbaki erreala Y sekuentziaren osagaien batezbesteko aritmetikoa da.

[prop_val_absoluto X Y]
X zenbakia Y zenbakiaren balio absolutoa da.

[suma X Y Z]
X zenbakia Y eta Z zenbakien batura da.

[sumasec X Y]
X zenbakia Y sekuentziaren osagaien batura da.

[tipo X Y]
Y X objektuaren mota da.

[t_q_elem X Y]
Y deskribatzaileak X sekuentzia definitzen du.

[t_q_elem_sec X Y Z]
X sekuentzia Y sekuentziaren azpisekuentzia da, Z deskribatzaileak definitzen dituen osagaiez osatuta dagoena.

[último_sec X Y]
X osagaia Y sekuentziaren azkena da.

[valdef X Y]
X objektuak Y balio konstantea dauka.

C ERANSKINA

Problema ebazkarrien enuntziatuak (ATNen interpretatzailea aldatu gabe)

P9.

Dada una secuencia de M caracteres, escribir el número de comas que tiene dicha secuencia.

P10

Dada una cadena de entrada en la que existe un determinado carácter "a", determinar la posición en la que aparece este determinado carácter.

P11

Dada una secuencia de N números enteros , hallar la suma de todos los números pares que se encuentren en dicha secuencia.

P12

Dada una progresión aritmética de N elementos enteros, $A_1=3$ y $R=4$, hallar la suma.

P13.

Escribir un programa que lea 10 números enteros y escriba su suma.

P15

Dada una secuencia de L dígitos que representa un número en base 8, hallar el valor entero en base 10 del número que representa la secuencia de entrada.

P16.

Sea TITULO una secuencia de 200 caracteres. Escribir un programa que cuente el número de ocurrencias de la cadena "ABCD".

P20

Escribir un programa que dada una secuencia de 50 caracteres escriba dicha secuencia sustituyendo todos los grupos de blancos por un solo blanco.

P21

Dado un número N natural, determinar cual es el K -ésimo dígito del número.

P24

Dada una secuencia de 200 enteros , ordenarlos ascendentemente.

P28.

Dada una secuencia S de 20 números enteros, obtener otra secuencia con los elementos pares de S . Determinar su longitud.

P31

Dada una secuencia de 25 enteros positivos , escriba un programa para encontrar el segundo entero mayor.

P36

Encontrar un número N de 3 cifras tal que dicho número N se pueda representar por la suma de los cubos de sus cifras.

P41.

Dadas dos secuencias de enteros A y B de 100 y 200 elementos. Escribir la secuencia resultante de cambiar en A el mayor valor de A por el mayor valor de B .

P44.

Dada una secuencia de 50 números enteros, calcular el máximo de ellos.

P47.

Dados dos números naturales mayores que 0 calcular el máximo comun divisor.

P49.

Calcular el sumatorio desde 1 hasta N de $1/(i+1)$.

P50

Calcular el producto desde 1 hasta 10 de $(i+1)*(i+2)/i$.

P51

Dada una secuencia de 25 enteros positivos, escriba un programa para encontrar el segundo entero mayor.

P52

Dado un número R entero positivo. Calcular la longitud de la circunferencia de radio R .

P54

Diseñar un algoritmo que lea un valor entero y determine cuando es o no par.

P55

Dados dos puntos en el espacio bidimensional, hallar la distancia entre ellos.

P56

Dado un número entero determinar su valor absoluto.

P57

Dados 50 números enteros, calcular el máximo de ellos.

P59

Calcular el factorial de N , siendo N un número entero positivo.

P60

Escribir un programa que calcule el número de cifras de un entero N .

P62

Dado un número real positivo, escribir dicho número redondeando a cero las cifras decimales.

P69.

Escriba un programa que tome un natural de 5 dígitos e invierta su orden.

P73.

Escribir un programa que dado un número N natural > 0 calcule las N primeras potencias negativas de 2.

P74.

Dada una secuencia de M números enteros, calcular su media aritmetica.

P80.

Sea N un número entero. Sea S una secuencia de M números enteros, en esta secuencia el primer número es distinto de N . Escribir aquellos números menores que N y que estén antes de la primera aparición del número N en la secuencia de entrada.

P85.

Calcular la diferencia entre dos enteros positivos.

P86.

Dada una secuencia de enteros, escribir los veinte primeros que no sean múltiplos de cuatro y que terminen en cuatro.

P87.

Dada una secuencia de 100 números enteros escribir dicha secuencia hasta que dos números consecutivos sean iguales, y de éstos escribir sólo uno.

P89.

Escribir aquellos números naturales cuyo cuadrado es menor que 1000.

P90.

Encontrar los divisores de un número natural no primo.

