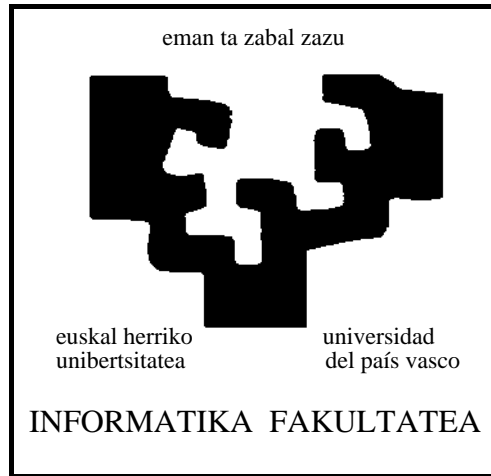


LENGOAIA ETA SISTEMA INFORMATIKOEN SAILA



## **HIZTSUA: HIZTEGI-SISTEMA URGAZLE ADIMENDUNAREN SORKUNTZA ETA ERAIKUNTZA**

**Hiztegi-ezagutzaren eskuratze eta errepresentazioa,  
dedukzio-mekanismoen ezarrera eta oinarrizko  
funtzionalitatearen zehaztapena.**

**Xabier Artola Zubillagak**

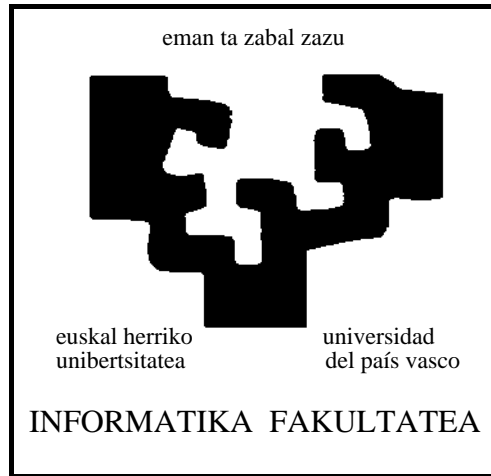
Informatikan Doktore titulua eskuratzeko aurkezturiko

**TESI-TXOSTENA**

Donostia, 1993ko urtarrila.



## LENGOAIA ETA SISTEMA INFORMATIKOEN SAILA



## **HIZTSUA: HIZTEGI-SISTEMA URGAZLE ADIMENDUNAREN SORKUNTZA ETA ERAIKUNTZA**

**Hiztegi-ezagutzaren eskuratze eta errepresentazioa,  
dedukzio-mekanismoen ezarrera eta oinarritzko  
funtzionalitatearen zehaztapena.**

Xabier Artola Zubillagak Fabrice Evrard-en zuzendaritzapean egindako tesiaren txostena, Euskal Herriko Unibertsitatean Informatikan Doktore titulua eskuratzeko aurkeztua.

Donostia, 1993ko urtarrila.



*“Hitz bat aurkitzen duzun bakoitzean mundu zati berri bat azaltzen zaizu. Ez du axola hitz hori hiztegiatan dagoen edo ez. Hiztegian, insektoa kolekzio batetan bezala dago, lehortua eta orratzez zeharkatua, betirako lotua, bere esanahiarekin eta bere etimologiarekin, baina hila. Baina hitzak bere bizia du, hitza esaten duen pertsona bakoitzak arnasa ematen dio.” (J. Sarrionaindia, Ainhoari Gutunak. Donostia: Elkar, 1990).*



## *eskerrak ematen*

*mila be(de)ratzi ehun kunplitu  
ta larogei ta hamabian  
urte ditxoso honetan eta  
inoxente egunean  
plazako lanak ere oraintxe  
dira eta ia-ian  
erramu adarra hor jarri dute  
iritsi gara batean  
neuk ere ez dut sinesten baina  
bukatu dut ezarian*

*eskerrik anitz eman ohi denez  
holakoak amaitzean  
ni ere hortan hasiko naiz ba  
badut eta nori eman  
lan luze honen azkenekoak  
hauexek direlakoan  
aurkibidea egin dut eta  
ondorioak atzenean  
bertso txar batzuk jarriko ditut  
inork irakurri ditzan*

*aurren-aurrenik gogoan dut nik  
fabrice evrard bizarduna  
tesi-lan honen gidari gisa  
ezinbesteko laguna  
ezin ahaztu hango guztien  
laguntzeko prestasuna  
jatortasuna eta litekeen  
lan-girorik alaiena  
agur lagunak ikusi arte  
donostialdean hurrena*

*etxe ondoko tolosatikan  
auzokorako bidean  
usurbildar fin arantza nuen  
tren hartako lagun izan  
han egindako adiskideak  
ere baditut gogoan  
xabier, iñaki ta gainerako  
euskaldunak okzitaniar  
hango bazterrak erakutsi ta  
jarri ninduten abian*

*frantsesik jakin ez zeki(e)n baina  
txintikan ateratzeke  
anak hiztegi hura harturik  
kopiatu zuen jo ta ke  
blanka, joserra ta miguel angel  
langile fin eta trebe  
lata ematera sarritan joan naiz  
hitz txarrikan jaso gabe  
pazientzi ederra duzue eta  
tori eskerrak dealde*

*ta ezin utzi aipatu gabe  
taldeko kepa eta xabier  
eneko etorri berriarekin  
hara zer hirukote eder  
adarjotzaile aparta bata  
nahasten du hainbat bazter  
besteak bajak maite ditu ta  
puskatzen du belaun ta izter  
ta eibartarrak oraintxe ez daki  
nola konpondu hainbat oker*

*fakultateko eta saileko  
kide(e)k badute bertute  
urte hauetan lagundu eta  
agoantatu izan naute  
noizko ta noizko, ez denek baina  
hori lata eman dute!  
eskerrak denoi eta barkatu  
egin badut inon kalte  
aurrerantzean saiatuko naiz  
ez izaten etxekalte*

*eta hurrena zuk badakizu  
zeinengatik ari naizen  
arantza baino zu izan zara  
akuilu finetan finen  
batzutan aupa bestetan aida  
aurrera beti bultzatzen  
ez baita erraza inondik ere  
ni horrela eramaten  
merezi duzu dudarik gabe  
mila esker eta omen*

*txosten luze ta astun honeri  
inork ekingo balio  
aholkutxo bat eman nahirikan  
egileak hauxe dio  
patxadaz hasi tamaina hartu  
ez presaka esan adio  
testutzar honen goxoarekin  
ez naizenez oso fio  
hasi eta buka leitzen duenak  
ez du ez gutxi balio*





*Egilea baliatu da, lan honetan, Gipuzkoako Kutzaren diru-laguntzaz.*



# AURKIBIDEA

SARRERA ETA AURKEZPEN OROKORRA	11
<b>I. Proiektuaren nondik norakoak eta aurkezpen orokorra.</b>	<b>11</b>
I.1 Sarrera gisa.....	11
I.2 Sistema urgazleak.....	14
I.2.1 Sistema urgazle arruntak.....	14
I.2.2 Sistema urgazle adimendunak.....	15
I.3 Hiztegia sistema urgazle. ....	17
I.3.1 Hiztegi klasikoa. ....	19
I.3.2 Hiztegi elektronikoak (automatizatuak).....	20
I.4 HIZTSUA: Hiztegi-Sistema Urgazle Adimenduna.....	23
I.4.1 Arkitekturaren diseinua.....	27
I.4.1.1 Sistemaren osagaiak. ....	27
I.4.1.2 Moduluen arteko lankidetzak.....	29
I.4.2 HIZTSUAren atzipen-bideak.....	30
I.5 Ebaluazioa. ....	31
I.5.1 Implementazioaren emaitzak. ....	31
I.6 Txostenaren eskema. ....	33
LEHEN PARTEA: HIZTSUA ERAIKITZEKO PROZESUA	35
<b>II. Hiztegi arruntetik Hiztegi-Ezagutza Baserantz.</b>	<b>35</b>
II.1 Lexikografia konputazionala: sarrera gisakoa.....	36
II.2 MRDetako ezagutzaren eskuratzea. ....	39
II.2.1 Definizio-esaldien egitura. ....	41
II.2.2 Definizioetan gordetako kontzeptuarteko erlazioak. ....	46
II.3 Le Plus Petit Larousse (LPPL): hiztegiaren ezaugarriak. Hiztegi Datu-Basea.....	49
II.3.1 Hiztegi Datu-Basearen eraketa. ....	50
II.4 Definizioen azterketarako metodoa.....	52
II.4.1 Maiztasun-azterketa.....	54
II.4.2 Desanbiguatze lexikala. Etiketatzeko edo "tagging"a. ....	57
II.4.3 Definizio-eskemak. ....	60
II.4.4 Behetik gorako analisi sintaktikoa: kategoria sintagmatikodun eskemen lorpena.....	62
II.4.5 Azterketa estatistikoa eta definizio-formulei buruzko ikerketa empirikoa. Erlatore bereziak. ....	65
II.4.6 LPPLko definizioetatik erazutako erlazio lexiko-semantikoak.....	68
II.5 Definizioen analisi sintaktiko-semantikoa. ....	77
II.5.1 Alshawiren analizatzailearen implementazioa.....	78
II.5.2 Patroi-hierarkiak. ....	80
II.5.3 LPPLko definizioen analisiaren emaitzak. ....	82
II.5.4 Egitura semantikoaren eraketa: HEBaren sorkuntza. ....	85
II.6 Hiztegi Ezagutza-Basearen prototipoaren eraikuntza. ....	87

II.6.1 LPPLtik aukeratutako azpimultzoa: Le Français Fondamental. ....	87
II.6.2 Prototipoaren erabidea. ....	88
II.7 Erabilitako prozeduren mugak. ....	92
II.7.1 Anlisi sintaktiko-semantikoaren mugak eta arazoak. ....	92
II.7.2 Modu honetan eraikitako HEBaren mugak. ....	96

### **III. Ezagutzaren errepresentazioa. Aberasketa dinamikoa eta alderdi**

#### **inferentziala HEBan.**

**99**

III.1 Ezagutzaren errepresentazioa. ....	99
III.1.1 Sare semantikoak eta frame-sistemak. ....	103
III.1.2 Hiztegi-ezagutzaren errepresentazioa. ....	107
III.1.2.1 Ezagutza lexikalaren errepresentazio semantikoa. ....	113
III.1.3 Ezagutzaren errepresentazioa HIZTSUAren HEBan. ....	115
III.1.3.1 Ezagutzaren egituraketa. ....	116
III.1.3.2 Metaezagutza: STRUCTURES azpi-EBa. ....	118
III.1.3.2.1 Hiztegi eta thesauruseko egituren klasea: STRUCTURES-DICO. ....	119
III.1.3.2.2 Atributuen sailkapena: ATTRIBUTS. Erlazio lexiko-semantikoei buruzko metaezagutza. ....	126
III.1.3.2.3 Iratxoak: VALEURS-ACTIVES. ....	133
III.1.3.2.4 Erregelaren sailkapena: REGLES. ....	133
III.1.3.3 DICTIONNAIRE azpi-EBa. ....	133
III.1.3.4 Sarea: THESAURUS. ....	135
III.1.3.5 Asmatutako errepresentazio-ereduari buruzko zenbait gogoeta. Beste errepresentazio batzuekiko alderaketa. ....	146
III.1.4 HEBaren prototipoaren ikuspegi orokorra: zenbait datu. ....	157
III.2 HEBaren aberasketa dinamiko erdiautomatikoa. ....	158
III.2.1 HEBaren hasierako eraikuntzan lortutako aberasketa. ....	159
III.2.2 Bigarren urratsa HEBaren aberasketan. ....	164
III.2.2.1 Sinonimiaren propietateen ustiapena. ....	165
III.2.2.2 Egitura taxonomikoaren hedakuntza. ....	166
III.2.2.3 Desanbiguaketa HEBan. ....	166
III.2.2.3.1 Desanbiguaketa lexikala. ....	167
III.2.2.3.2 Desanbiguaketa sintaktikoa. ....	171
III.3 Alderdi inferentzialaren ikuspegia. ....	173
III.3.1 Erlazio lexiko-semantikoen konposaketa. ....	175
III.3.2 Propietateen herentzia. ....	180
III.3.3 Erlazioen eta THESAURUSeko kontzeptuen arteko loturak. ....	182
III.3.4 Ziurtasun-ezaren tratamendua: HEBan dirauen anbiguetatea. ....	183
III.4 Prototipoaren egungo egoera. ....	185

### **BIGARREN PARTEA: HIZTEGI-SISTEMA URGAZLEAREN ALDERDI FUNTZIONALA**

**187**

#### **IV. HIZTSUAren funtzionalitatea.**

**187**

IV.1 Hiztegietako informazioaren atzipenerako ohizko bideak. ....	188
IV.1.1 Paperezko hiztegia. ....	188
IV.1.2 Hiztegia euskarri informatikoan. ....	191
IV.2 Oinarrizko funtzionalitatea. ....	194
IV.2.1 Eragiketa primitiboak. ....	196

IV.2.2 Oinarrizko funtzioen zehaztapena. Inplementaziorako estrategiak. ....	198
IV.2.2.1 Oinarrizko datu-egiturak: deskribapena eta sintaxia.....	200
IV.2.2.2 Galde-funtzioak.....	205
IV.2.2.2.1 DDEF: Definizio-galdea. ....	205
IV.2.2.2.2 RDEF: Definizio-birformulaketa. ....	209
IV.2.2.2.3 DPRO: Kontzeptu baten propietateen galdea. ....	210
IV.2.2.2.4 DRAP: Bi kontzepturen arteko elkarbidearen bilaketa. ....	212
IV.2.2.2.5 DDIF: Bi kontzepturen arteko diferentzien galdea.....	213
IV.2.2.2.6 RTHS: Kontzeptu-bilaketa tesaurikoa. ....	215
IV.2.2.3 Egiaztapen-funtzioak.....	217
IV.2.2.3.1 VDEF: Definizioaren egiaztapena. ....	217
IV.2.2.3.2 VPRO: Kontzeptu baten propietateen egiaztapena.....	219
IV.2.2.3.3 VRAP: Bi kontzepturen arteko elkarbidearen egiaztapena. ....	220
IV.2.2.4 Oinarrizko beste zenbait funtzio. ....	221
IV.2.2.5 RTHS funtzioaren inplementazioari buruzko zenbait ohar.....	222
IV.2.3 Oinarrizko funtzioen konposagarritasuna. Kontsulta- eta eskari-lengoaia. ....	224
IV.2.3.1 Funtzio-konposaketa HIZTSUAren baitan. ....	225
IV.2.4 HIZTSUAren erantzunak. ....	228
<b>V. Erabiltzailearekiko elkarrizketa.</b>	<b>231</b>
V.1 Erabiltzaile-sistema elkarrekintzaren ezaugarriak.....	232
V.2 HIZTSUAren interfazearen zirriborroa.....	236
V.3 Erabiltzaile-sistema komunikazio-prozesua eta moduluen arteko elkarrekintza.....	241
 <b>AURRERA BEGIRAKOAK ETA ONDORIOAK</b>	<b>246</b>
 <b>VI. Zabaldutako ikerlerroak eta perspektibak.</b>	<b>246</b>
VI.1 HIZTSUA erabilinguru eta erabiltzailearekin integratzea: sistema "aktiboago" eta "adimendutsuago" baterantz.....	246
VI.1.1 Lexikografoa erabiltzaile. Hiztegitzarako laguntza. ....	247
VI.2 HEBaren aberasketa: dedukzio-mekanismo berrien ezarrera.....	247
VI.3 Definizioen errepresentazio-ereduaren baliozkotasuna. Sistema hiztegianiztunak. ....	248
VI.4 HIZTSUA bestelako inguruneetan erabiltzeko aukerak.....	250
VI.4.1 Lengoaia naturalaren tratamendurako sistemen osagai lexikalak eraikitzeke iturburua.....	251
 <b>VII. Ondorio gisa.</b>	<b>253</b>
 <b>BIBLIOGRAFIA</b>	<b>255</b>
 <b>Orokorra</b>	<b>255</b>
 <b>Hiztegiak</b>	<b>265</b>



# SARRERA ETA AURKEZPEN OROKORRA

## I. Proiektuaren nondik norakoak eta aurkezpen orokorra.

### I.1 Sarrera gisa.

Lengoaia naturalaren prozesamenduko (LNP) ikerkuntzan, lexikoaren gaia gero eta garrantzitsuago bilakatzen ari da. Azken hamarkadan sortu eta garatutako analisi-formalismo eta sistemek sintaxiaren arlotik lexikoarenera lerrarazi dute interesa. Lexikoa bihurtu da hizkuntz ezagutzaren gordailu nagusia, eta sarri baita hizkuntzatik kanpoko jakintzarena ere. Horrela bada, bistan da egungo joera hizkuntz ezagutza gramatikaren arlotik lexikoarenera lerratzearen aldetik ari dela.

Zenbait ikerlarik aspalditik aldarrikatu du lexikoaren lehentasun hau LNPko lanetarako. LNPko sistema gehienek hiztegiari oso garrantzi txikia ematen dioten bitartean, benetako garrantzia hizkuntzaren analisi sintaktiko eta semantikoak hartzen dute. Aitzitik, hiztegiak osagairik behinena izan behar du LNPko sistema orotan. Melçûk-ek (1978a:3) dioen bezala:

. . . a dictionary is the most important component of any meaningful automatic text processing system.

Horren ondoren, egile honek dio gramatikako erregelak hiztegia laburbiltzeko bide bat besterik ez direla, eta bukatzeko esaten du, benetako informazio linguistikoaren gordailu bakarra den heinean, hiztegiari erabateko lehentasuna zor zaiola.

## *I. kapitulua*

Berrikiago, lexikografia konputazionalaren arloko zenbait ikertzailek ere —horien artean (Boguraev, 86) eta (Calzolari & Picchi, 88)— aitortzen dio lexikoari lehentasun hori hizkuntz tratamenduko lanetarako.

Bestalde, esan behar da hizkuntz problema erreala bati ezin zaiola aurre egin 25 hitzeko (batez-beste) hiztegia duen sistema batekin, eta hori da nonbait (Boguraev & Briscoe eds., 89a) sistema askoren kasua. LNPa neurri errealeko osagai lexikalak dituzten "benetako" sistemak eraiki nahi direnean, eta egile gehienek aitortzen dute lexiko handi horien premia (Smith & Maxwell, 80; Boguraev & Briscoe arg., 89a), ezagutza lexikalaren eskuratze-prozesua arazo larria bihurtzen da. Neurri errealeko osagai lexikal horiek eskuz egitea hain da lan handia ezinezko baita ia.

Gauzak horrela, hizkuntz hiztegi arruntek berebiziko garrantzia hartzen dute. Bistakoa da hiztegi-argitarapen arruntetan aurki daitekeen informazio pila behar den bezala tratatuz gero, osagai lexikal horien eraikuntza automatiko edo erdiautomatikorako abiapuntu inportantea izan daitezkeela.

(Wilks *et al.*, 90)en, egileek bi interes mota ezagutzen dizkiote hiztegiari —hizkuntza gaitzat duen testua dela esanez, baina munduari buruzko jakintzaren bilgune bezala ere hartuz—, informatika linguistiko eta adimen artifizialeko ikertzaileentzat: lehenik, interes teorikoa, hiztegia hizkuntzaren egitura semantikoaren ikerketarako bide den neurrian; baina baita interes praktikoa ere, hiztegiak adimen artifizialean arazo larriena den jakintzaren eskuratzean izan dezakeen baliagarritasunagatik.

Hizkuntz hiztegiak —hiztegi esplikatzaile elebakarraz ari gara hemen batik bat— bere baitan gordetzen duen ezagutza ortografiko, morfologiko, lexiko-semantiko eta entziklopedikoen multzoa haintxe da handia ezen bertatik informazio guzti hori erauzi eta ondoren tratamendu automatikoko sistema batek erabil dezan prestatzea erronka galanta bait da.

Ezagutza guzti horren errepresentazioa da kontua, hizkuntzaren tratamendu automatikoko sistema baten osagai nagusi gisa erabiltzea posible egingo duen eran alegia. Honi buruz, hara zer azpimarratzen duen Calzolari-k (1983:49):

Our basic theoretical point has been this: the dictionary should be conceived as a central, core component in natural language analysis and application. Our aim finally is to transform it into a dynamic system of representation rather than a static object for consultation.

Hiztegi arruntak euskarri magnetikoan edukitzeko aukerak (Machine Readable Dictionaries edo MRDak) ikerrarlo honi ekiteko bideak zabaldu egin ditu. Zoritxarrez ordea, hiztegien makina-bertsio horiek ez dira, maiz, paperezko bertsioen iturri izateko bereziki grabatutako banda magnetiko hutsak baino. "Datu-base lexikala", hots, hiztegiak daukan



informazio guztiaren biltegitze eta kontsultarako espreski diseinatu eta prestatutakoa, ez da oso sarri suertatzen. Horiek horrela, MRDak ezagutza-iturburu gisa automatikoki tratatatu nahia ez da txantxetakoa. Boguraev eta Briscoe-k (1989a) argitaratutako liburuan, lexikografia konputazionalaren aurkezpen modukoarekin batera ikerrarlo honetan egindako edota egun burutzen ari den zenbait lanen azalpena ematen da.

Alde batetik begiratuta, txosten honetako tesi-proiektua testuinguru honetan koka daiteke. Proiektuaren nondik norakoak aipatzen hasita bada, ondokoak azpimarra genitzake:

- *Ezagutza lexiko-semanticoren erauzketa* hiztegi arruntetik. Hiztegi elebakar bateko definizioetan dagoen ezagutzaz dihardugu hemen batik bat.
- Ezagutza horren *errepresentaziorako proposamena*.
- Hiztegi-egituretan (definizioetan batez ere) dagoen ezagutza inplizitua esplizitu eta atzigarri bihurtzeko beharrezko diren *esplotazio-mekanismoen ezarrera*.

Gure proiektua gorago aipaturiko problematikatik bereizten duena sistemaren helburu diren erabiltzaileen baitan datza: gurean giza erabiltzailea izan baitugu gogoan. Byrd-ek (1990:117, 128) honako hau dio:

Lexicological analysis techniques which produce lexical information for natural language processing systems also yield data that are useful for human users of on-line dictionaries.

. . . lexical information derived by lexicological analysis for a variety of applications is also useful for on-line reference. We expect that the converse is also true: information developed specifically for on-line reference will, when available, prove useful for other purposes.

Informazioa biltegitzeko sistemek gero eta ahalmen handiagoa (CD-ROM, etab.) eskaintzen dute, eta giza erabiltzailea helburu duten hiztegi elektronikoak merkatuan agertzen hasiak dira dagoeneko.

Honela bada, gure proiektuaren motibazio eta zeregin nagusia *giza erabiltzailearentzat hiztegi-laguntzako sistema diseinatu eta eraikitzea* izan da. Horretarako, sistema horrek eskaini beharreko funtzionalitatea zehaztu behar izan da, ondoko alderdiok kontuan harturik:

- *Laguntza-egoera batean integratua*, ulerkuntza zein idazketa burutzen ari den ingurunean.
- Funtzionalitatearen aldetik, erabiltzaileari hiztegi datu-base arrunt batek baino "zerbait gehiago" eskainiko liokeen *funtzionalitate-multzoa*.
- *Erabiltzaile desberdinen tipologia*: erabiltzaile arrunta, ikaslea, itzultzailea, hiztegi-gilea, etab.

## **I.2 Sistema urgazleak.**

Ikuspuntu informatikotik, laguntza-sistema edo sistema urgazlea ("help system") tresna informatiko baten —jeneralean— erabileraz kontsultak egiteko aukerak eskaintzen dituen sistema izan ohi da. Sistema eragile orok, programazio-lengoaiek edota testu-tratamendurako programek beren erabilera erosoagoa egitearren laguntza-sistemaren bat —sinpleenetatik hasi eta sofistikatuenetaraino— jarri ohi dute erabiltzailearen eskuetan. Kasu askotan, programa horien erreferentzi eskuliburuaren makina-bertsioa —erabiltzailearekiko elkarrekintza barne dutela— izaten dira laguntza-sistemok.

Hiztegia erreferentzi lan erraldoi bat bezala har daiteke (ikus § I.3). Hiztegian ordena alfabetikoz sailkatutako hitz-zerrenda bat aurkituko dugu, hitz bakoitzari buruz honakoa izango dugularik: ortografia, hitzaren esanahien banaketa adiera desberdinetan, adiera bakoitzari dagokion kategoria gramatikala, erabileremuak, adibideak, beste hitz batzuen erreferentziak, etab. Kontua da beraz, hiztegia hizkuntzaren lexikoari buruzko "erreferentzi eskuliburu" erraldoi bat bezala ikustea.

Hizkuntzaren erreferentzia gisako hiztegiaren ikuspegi hori eta berorren baliagarritasuna hitzen ulerkuntza eta erabilerarako laguntza-tresna bezala izan ditugu hausnargai proiektu honetan.

### **I.2.1 Sistema urgazle arruntak.**

Aipatutako programek eskainitako sistema urgazle arruntan gunea, lokarri desberdinen bitartez lotutako sarrerekin osatutako thesaurus delako bat izan ohi da. Sarrera horiek programa edo aplikazioari berez dagozkion kontzeptuak izaten dira, hala nola aukera posibleak, komandoak, erabilitako objektuak, erabilera desberdinak, etab. LISP inguruneetan egon ohi diren laguntza-sistemak, UNIX sistema eragilearena, etab. dira tresna hauetako batzuk.

Jeneralean, thesaurus horiek kontzeptuak taxonomiatan sailkatzen dituzten lokarri hierarkikoak dauzkate. Sistemarekiko elkarrekintza gertatzen denean, sistemak kontzeptu-zerrenda bat aurkezten dio erabiltzaileari —hierarkiako leku gorenean daudenena, hain zuzen—, kontzeptu horiei buruz informazioa duela adieraziz. Erabiltzaileak kontzeptu horietako bat aukeratu eta sistemak, hierarkian bilatu ondoren, kontzeptu horretaz dakiena kanporatzen du —aurretik prestatutako testuak erabiliz— edo kontzeptu horren azpian dauden aukera desberdinak aurkezten dizkio, informazio gehiago zeinetaz nahi duen esan dezan eskatuz. Prozedura hau behin eta berriro errepikatuko da erabiltzailea pozik geratu arte.

Sistema urgazle hauetan sofistikazio maila altuxeagoa ere aurki daiteke, sistema kontzeptu bat sinonimo baten bitartez aurkitzeko gauza denean. Baina oro har, horixe da ohizko sistema urgazleok eskaintzen duten guztia.

Lan-estazio pertsonalak eraginkorrago eta ahalmentsuago bilakatzen ari diren honetan, lineango erreferentzi sistemak gero eta erabiliago izango direla esan daiteke (Byrd, 90). Aukera honek berekin behar ditu informazio pila handi horiek antolatzeke era berriak eta, horrekin batera, erabiltzailearen eskumenean atzipen-bide desberdin eta sofistikatuagoak jartzea.

Sistema urgazle hauen pasibotasuna da beste alderdi inportante bat hemen. Hartley eta Pilkington-ek (1988), ohizko sistema hauen mugak adieraztearekin bat, laguntza-sistemak erabiltzailearen planak eta eginbeharrak interpretatu eta plan horiek burutzerakoan gidari izan behar duela diote, erabiltzaileak egindako hutsak ere azaleratuz. Egile horien ustetan, sistema urgazleak jarrera "aktiboagoa" hartu behar luke erabiltzaileari bere laguntza eskaintzerakoan. Honela, sistema urgazle adimendunak aipatzen dira, irakaskuntza-sistema adimendunetatik (Intelligent Tutoring Systems, ITS) gero eta hurbilago diren sistemak alegia.

## **I.2.2 Sistema urgazle adimendunak.**

Esan bezala, azken aldi honetan sistema urgazle adimendunak (IHS, Intelligent Help Systems), hasi dira plazaratzen. Europako Komunitateko ESPRIT P280 proiektuan, IHS direlakoak diseinatu eta eraikitzen dihardute (Breuker arg., 90). Testuinguru honetan, helburua da informazio-tratamendurako ohizko programetan (softwarea) linean laguntza eskainiko duten sistemak eraikitzea. Sistema horien ezaugarri nagusia bi laguntza mota desberdin eskaintzea litzateke:

- *Laguntza pasiboa*, erabiltzailearen eskariz egindakoa. Laguntza mota hau litzateke arestian aipatu dugun ohizko laguntza.
- *Laguntza aktiboa*. Sistemak erabiltzailea burutzen ari den lana gainbegiratu egiten du etengabe —"bizkar gainetik" erreparatzen du— eta gai da honek laguntza premia behar duenentz igartzeko.

Txosten horren egileek ematen dute, sistema hauek, "adimenduntzat" hartuko baditugu, behar lituzketen gaitasunen zerrenda:

- Planifikazioa: sistemak jakin egin behar du erabiltzaileak zerbait burutzeko egin beharko lukeena. Gaitasun hau egokia da ingurune "operazionalerako" sistema

## *I. kapitulua*

urgazleetan, hau da, erabiltzaileak —komando-sekuentzia bat edota menu-aukera jakin batzuen bitartez— lan konkretu bat burutu behar duen sistemenetan.

- Plan-ezagutza: sistemak jakin egin behar du erabiltzaileak zer egin nahi duen edozein momentutan eta nola.
- Erabiltzailearen diagnostikoa eta eredia. ITS delakoetan egiten denaren antzera, sistemak erabiltzailearen portaera desegoki baten arrazoia bilatu egin behar du, beronen hutsuneak edota kontzeptu okerrak zein diren atzematearren; diagnostikorako zein plan-ezagutzarako, sistemak erabiltzailearen eredia behar du, erabiltzaile desberdinek uneoro duten kontzeptuen ulertze maila diferenteak errepresentatzearren.
- Aplikazioaren modelizazioa: sistema urgazlearen aplikazio-eremuaren errepresentazioa.
- Aplikazioaren emulazioa: aplikazioaren uneoroko egoeraren modelizazio dinamikoa.
- Sistemaren erantzunen sorkuntza lengoia naturalean.

Hartley et Pilkington-ek (1988) beste alderdi interesgarri bat ere aipatzen dute IHS hauek direla eta: ikastearen aspektua hain zuzen, sistema hauek erabiltzaileari proposatu eta eskaini beharko lioketen "aholku didaktikoari" dagokiona.

Gorago aipatu bezala, sistema urgazlearen kontzeptu hau programa informatikoen azpisistema lagungarri baten ideiatik dator, programa horiekin dabilenean erabiltzailea prozesu operazional batean murgildurik —hau da, lan jakin bat burutzekotan eragiketa batzuk egin beharrean— egon ohi delarik. Egoera honetan, sistema urgazleak funtzionalitate multzo bat eskainiko dio erabiltzaileari bere ihardueran laguntzeko asmotan. Ezaugarri operazional hori dela medio, aipatu berri ditugun sistema urgazleek planifikazio edota plan-ezagutzazko baliapideak izan behar dituzte. Alabaina, testu baten irakurketa eta ulerkuntza, edota idazketa-prozesua, prozesu operazional horien artean sailkatu behar ote ditugu?

Erantzunak direla eta berriz, edo hobe esanda, sistema hauek behar duten pertsona-makina interfazeaz ari garelarik, lengoia naturala aipatzen da lan anitzetan (Breuker arg., 90; Hartley & Smith, 88; Hartley & Pilkington, 88). Lengoia naturala ez da, ordea, sistema urgazle hauen interfazea eratzeko modu bakarra: interfaze-sistema "klasikoagoekin" ere, era inteligente batez diseinatu eta erabiliz gero, egoki ebatz baitaiteke arazo hori, interfazeon "zurruntasunean" erori gabe. Esandako horren adibidetzat, (Hartley & Pilkington, 88)n aurretik finkatu gabeko menuz egindako interfazea aipatzen da: menu horiek erabiltzaileak une bakoitzean dituen premien arabera aukeratu eta aurkezten dira.

Bukatzeko azpimarratu, gure proiektuan *aplikazio-eremuaren errepresentazioari lotu gaitzaizkiola batik bat*, aplikazio-eremu bezala hiztegi-ezagutza hartuz. Aukeratutako adierazpidea eta bertan ezarritako inferentzi gaitasuna dira, hiztegia laguntza-tresna "adimenduna" bihurtu behar dutenak.

### **I.3 Hiztegia sistema urgazle.**

Hiztegia laguntza-sistema da batik bat. Bertara jo ohi dugu hitz baten esanahia jakin nahi dugunean, hitz baten sinonimo edo antzekoen bila gabiltzanean, hitz baten idazkerari buruzko susmo bat dugunean, etab. Susmoen egiaztapenaren alderdi horrek badu bere garrantzia hiztegien erabilerari dagokionean eta ez dago inola ere ortografi egiaztapenera mugatua: sendotu beharrean daukagun ezagutzaren bat baieztarren maiz erabili ohi dugun tresna da hiztegia.

Hiztegi baten helburu nagusia erabiltzaileei *laguntza eskaintzea* da, hizkuntza idatzia — ahozkoa ere bai noski, nahiz eta gure proiektuan ez horretaz ihardun— erabiltzen den prozesu desberdinetan, ondokoak nagusi direlarik:

- *Hizkuntz eskurapena (irakurketa)*: hitzen ulerkuntza eta azalpena maila desberdinetan eta bitarteko desberdinekin, erabiltzailearen ezagutzen egiaztapena, hitzen arteko harremanen bilaketa, kontzeptu bati dagozkion ezaugarri edota tasunen bilaketa, etab.
- *Hizkuntz sorkuntza (idazketa)*: hizkuntz eskurapenean aipatutakoez gain, idazketan hiztegia erabili ohi da hitz, esamolde edota sintagmen parafraseaketarako, sinonimo edo antzeko moten bilaketan, hitzen arteko bateragarritasunaren egiaztapenean (agerkidetza lexikaleko murriztapenak), erabilerako zailtasunak ebazteko, eta beharrezko den hitz egokiaren ("mot juste") bilaketan, besteak beste.

Ohizko hiztegi gehienak —elebakarrak, definiziodunak— hitzak ulertzen laguntzeko eginak daude. Hala ere, ia hiztegi guztietan aurki ditzakegun erabilera-adibideek sorkuntzaranzko joera erakusten digute. Thesaurus delakoek, hiztegi analogikoek eta baita sinonimo-hiztegiek ere laguntza handiagoa eskaintzen diote erabiltzaileari hizkuntz sorkuntzako lanetan.

*Dictionnaire Explicatif et Combinatoire du Français Contemporain* (Melçûk, 78a; 82; 84) hiztegi ezaguna sorkuntzari begira egindako hiztegia da. Hiztegi mota berri hau hiztegi-informazioaren sistematizazio eta uniformizazio orokorra lortzeko asmoan egindako tresna da. Hiztegi honek ematen dio —sistematikoki— erabiltzaileari bilatutako sarrerari dagokion

## I. kapitulua

informazio semantiko, sintaktiko, konbinatoria lexikaleko, eta abarreko guztia. Sorkuntz hiztegia da, "nola esan X" delakoari begira egina, "zer esan nahi du X-ek" delakoari baino.

Hiztegiaren helburua bada, besteak beste, honakoetan datza:

- Hiztegian *kontzeptuak bereiztu egiten dira*. Bereizketa hori ez da era bakarrean egiten hizkuntza bateko hiztegi guztietan: zenbat eta landuago hiztegia orduan eta finago joko du hitz bakoitzaren adierak eta ñabardurak bereizterakoan.
- Hiztegiak *kontzeptuak esplikatzen ditu*, horretarako baliapideok erabiliz:

- a) Definizio-testua: hiztegiak definitutako kontzeptua —"definiendum"— azaldu egiten du definizio-testuaren bitartez —"definiens"—, beste kontzeptu batzuekin harremanetan jarritz (ikus § II.2.1). Sinclair-ek (1987:77), Murray aipatuz, ondokoa dio:

Murray (1888) declares: "A Dictionary ... is not a Cyclopaedia: the Cyclopaedia describes things; the Dictionary explains words, and deals with the description of things only so far as it is necessary in order to fix the exact significations and uses of words."

- b) Erabilera-adibideak. (Sinclair, 87:137)n ondoko hau irakur daiteke hiztegiak adibideei buruz:

This should help to reinforce the meaning —not by acting as a reformulation of the definition, but showing how the word is actually used in an appropriate context, a typical grammatical structure, and with words that are normally associated with it.

- c) Beste kontzeptu batzuen erreferentziak: sinonimoak bereizki baina baita bestelakoak ere, zenbait hiztegitan behintzat.

- Hiztegia *kontzeptuak esplikatzen munduaz baliatzen da* hizkuntzaz baino gehiago: hizkuntz hiztegia ez da entziklopedia, baina mundu errealeko kontzeptuak erabiltzen ditu —azal-azalek bada ere— hizkuntza horretako hitzen esanahi desberdinak azaldu eta finkatzeko. Autore askok aipatzen du hiztegia mundu errealeari zein hizkuntzari buruzko informazioa daukan biltegi handia bezala (Litkowsky, 78; Boguraev, 86; Sinclair, 87). (Calzolari & Picchi, 88:87)n esaterako, hau diote:

The dictionary is now considered as a primary source not only of lexical knowledge but also of basic general knowledge (ranging over the entire "world"), and some of the dictionary systems which are being developed have knowledge acquisition and knowledge organization as their principal objectives. . .

- Hiztegia *arau-emailea* da (hizkuntza idatziaren mailan batez ere). Lehenago ere aipatu dugunez, hiztegiaren funtzio inportanteenetako bat hitz baten idazkera

zuzena egiaztatzea da: askotan jo ohi da hiztegiara —hitzaren forma grafikoa fonetikotik urrun samar dagoen hizkuntzetan batez ere, ingelesez edo frantsesez esaterako— hitz bat xuxen nola idatzi behar den jakiteko. Izan ere, testu-tratamendurako programa orotan egun aurkitzen dugun egiaztapen eta zuzenketa ortografikoa funtzio hori betetzeko bereziki prestatutako hiztegiak besterik ez dira. Gainera, hiztegiaren funtzio arau-emaile hori ez dagokio bakarrik ortografiari baizik eta harantzago doan zerbait da, lexikoaren mailan neologia edota mailegutzan, eta sintaxiaren aldetik aditzen erregimena bezalako arloetan ere araua finkatzen duelarik.

Proiektu honetan interesatzen zaiguna berriz, hiztegiaren alderdi semantikoa da batez ere. Hizkuntz ikasleak hitz baten edo esaldi baten esanahia jakin nahi duenean erabili ohi duen hiztegia. Itzulpengintzan diharduen itzultzaileak iturburu-hizkuntzako zenbait elementuri buruz duen ezagutza zehaztu behar duenean darabilen hiztegi elebakarra. Eskoletako umeei irakurketarako, testu-ulerkuntzarako edota idazketa-lanetarako tresna lagungarritzat hartu ohi duten hiztegia. Hizkuntzako hitzen esanahia "azaltzen" duena, labur esatearren.

### **I.3.1 Hiztegi klasikoa.**

Esan dugun bezala, hiztegia erreferentzi lan erraldoia eta oso baliagarria den zerbait da (Miller, 84); Marchionini-k (1989) entziklopediaren antzeko zeregina aitortzen dio hiztegiari, entziklopedia "oinarrizko erreferentzi tresnatzat" hartuz.

Hiztegian, interesatzen zaigun ikuspuntu semantikotik begiratuta, askotariko ezagutza pila handia aurki daiteke. Hitzak azaltzeko, definizioak idazten dituen lexikografoak mundu errealeko erreferentziak hartu ohi ditu maiz. Definizio horiek hizkuntzari buruzko informazioa ere badaukate beren baitan, hitzen esanahia dela, hitz horiei dagokien zenbait murriztapen semantiko dela, etab. Hiztegi-sarreretan emandako adibideek ere mundu erreal horretara eramaten gaituzte, horrekin batera hitz horiek testuinguru desberdinetan erabiltzeko "arauak" emanaz: hizkuntz ezagutza da beraz. Orobat, hitzen agerkidetzari buruzko ezagutza kodeturik dago hiztegiko sarreretan. Bestalde, hizkuntz kontzeptuen arteko erlazio multzo polita ere badago ezkutuan —batzuetan ez hain gorderik, erreferentzia esplizituei esker— hiztegi-egitura desberdinetan.

Azkena aipatutako ikuspegi hori, hiztegia erlazio lexiko-semantiko desberdinen bitartez lotutako kontzeptu-multzo bezala ikusten duena alegia, izan da ikerketa-proiektu honetan gehien interesatu zaiguna.

Hiztegia bada, mundu errealari eta hizkuntzari buruz potentzialki informazio multzo handia eskain dezakeen lana dugu. Hala eta guztiz ere, informazio multzo handi horren parte

## *I. kapitulua*

bat atziezina da, neurri handi batean eskainitako atzipen-bideen eskastasunagatik. Izan ere, hiztegi arruntean atzipen-bide bakarra izan ohi dugu: atzipen "alfabetikoa", hain zuzen. Hitza ezagutu egin behar da —nola idazten den— hitz horretaz hiztegiak daukan informazioa atzemango badugu.

Hiztegi arruntaren atzigarritasun "eskasa" erakusten duen adibide tipikoa dugu gurutzegramak egiten dituenarena. Izan ere, gurutzegramak betetzerakoan hitzen atzipena murriztapen orto-grafemikoen bidez —hitzen hasiera ez du ezagun sarritan!— posible izatea gustatuko bailitzaioke, edota "esanahiaz"; gogoko luke, inondik ere, definiziotik —aldeztatik aurretik ezagun— hitzerainoko bidea hiztegia erabiliz egiteko aukera izango balu. Bistakoa da bide hori ezin dela ibili hiztegi klasikoekin.

Beste alderdi interesgarri bat hiztegien erabilerari dagokio. Askotan pentsatu gabe egin ohi dugun erabilera horrek oso gaitasun intelektual inportanteak eskatzen baititu. Hori dela eta, hara zer dioen Miller-ek (1984:462):

A dictionary is an extremely valuable reference book, but its familiarity tends to blind adults to the high level of intelligence required to read it. This aspect becomes apparent, however, when school children are observed learning dictionary skills.

### **I.3.2 Hiztegi elektronikoak (automatizatuak).**

Baliapide informatikoen eskaintzen duten ahalmena benetan aprobeztatu nahi baldin badugu, hiztegi elektronikoa ezin daiteke MRD sinple bat izan (Miller, 84). Hiztegiaren informatizazioak euskarri magnetiko edo bestelako batean papereko hiztegi-testuaren grabatze soila baino askoz gehiago izan beharra dauka; hori ez ezik, ohizko datu-baseek eskaintzen digutena ere gainditu egin behar du. Honi buruz, Byrd-ek (1990:134) argi asko dio:

Developers of on-line reference systems will need to avoid the temptation of making the computer a mere surrogate for printed reference books. New ways of organizing and accessing reference information are possible with computers and should be pursued.

Hiztegien informatizazioari buruz ari garenean ordea, oso gauza desberdinei buruz aritzen gara.

Argitaldarien gero eta gehiago erabiltzen dute ordenadorea inprimazteko diren testuen konposaketarako. Hiztegi-argitaletxe handietan tresna informatikoak ezinbestekoak dira lan desberdinetarako oinarri diren informazio lexikografikoko datu-banku handien biltegitze eta kudeaketarako. Gaur egun lexikografoek laguntza informatiko handia eskura dezakete: KWIC delako konkordantzi programak, informazio-eskurapenerako sistemak linean, "lexicographer's workbench" (Lenders, 90) delakoa, etab. dira, besteak beste, zenbait kasu.



Ordenadorez lagundutako lexikografiaren arlo honetako zenbait lan aipatzen hasita, Collins argitaletxearen COBUILD hiztegia (Sinclair, 87) edo 1984ean hasitako New Oxford English Dictionary (OED) proiektuaren inguruko hainbat eta hainbat lan aipa daitezke (Simpson, 85; Weiner, 89); azken horren inguruan urtero egiten da kongresu bat Kanadako Waterlooko Unibertsitateko UW Centre for the New OED delakoan.

CD-ROMean argitaratu eta merkaturatutako hiztegiak edo telematikari<sup>1</sup> esker jendearen eskueran ezarritako hiztegi datu-baseak lirateke hiztegien informatizazioaren beste ikuspegi bat. Kasu hauetan, ordena alfabetiko tradizionalaz at bestelako kontsulta-aukerak ere eskaini ohi zaizkio erabiltzaileari. Gai berezitueta terminologi bankuak ere oso hiztegi-tresna inportanteak dira itzultzaileentzat, irakasleentzat, etab., eta euskarri informatikoaz eta erabileraz bestelako ikuspegi bat erakusten digute. Horien artean aipa ditzakegu TERMIUM (Kanadako terminologi bankua), UZEIk prestatutako EUSKALTERM, etab.

Badira harantzago doazen software-egileak ere. MEMODATAk (Caen, Frantzia) iragarria du DICOLOGIQUE™ izeneko produktua Hiztegi Interaktibo bezala. Kontsultarako tresna gisa, DICOLOGIQUE™ honek aukera emango du, egileen esanetan, hitza ideiatik abiatuz bilatzeko eta alderantziz, “. . . *comprenant des fonctions de recherche puissantes intégrant les opérateurs logiques.*”. Datu-banku lexikal bezala berriz, DICOLOGIQUEk bilaketa automatizatuak egiteko aukera eskaintzen dio LNPko programa bati: “. . . *permet d'envisager des recherches automatisées commandées par un programme de traitement du langage naturel.*”.

Hain zuzen ere, LNPko sistemetarako neurri errealeko osagai lexikalen eraikuntza litzateke hiztegien informatizazioaren beste aspektu bat, nahiz eta kasu honetan helburua oso bestelakoa izan. Izan ere, kasu horretan, lengoiaia naturaleko ulerkuntza- zein sorkuntza-sistemetarako beharrezko den ezagutza lexikal guztia sistematizatu eta prestatzea baita lana, erabiltzailea programa bat izango delarik. Kontuan hartzekoa, dudarik gabe, testuinguru honetan, munduari buruzko eta "zentzu komuneko" informazio kopuru handia esplizitu egitearen premia, erabiltzailea pertsona denean albo batera utzi edo egiten dugun informazioa.

Hiztegia informazioa eskuratzeko sistema "elektroniko" bezala ikusteak paperezkoaren aldean dituen abantailez pentsatzeko bide ematen digu: ezagutzaren atzipen hobe, interfaze egokia, dinamikotasuna, bestelako tresna informatikoekiko integrazioa, etab.

Atzipena dela eta, ez da zaila paperezkoak ukatzen digun baina tresna informatikoan erraz ezar daitekeen zenbait aukera imajinatzea: pattern-matching-erako teknikak sarreren

---

<sup>1</sup> Frantziako Miniteltek aurki daitezke dagoeneko Interrosiences (3615 RF) —“*encyclopédie télématique interactive accessible au public*”— entziklopedia bezalako aukerak.

## I. kapitulua

bilaketarako —edota sarreron edukinaren bilaketarako—, bilaketa baldintzak adierazteko eragile logikoak erabiltzeko posibilitatea, etab. (Marchionini, 89).

Marchionini-k (1989), entziklopedia elektronikoez diharduela, antolakuntzazko ezaugarri interaktiboak ere aipatzen ditu tresnok eduki behar dituzten ezugarrien artean. Honi buruz, bilaketa-estrategia elkarreragileak eskainiko dituzten sistema nabigazionalak sortu behar direla dio, lan entziklopedikoak informatizatzen hasiz gero. Informazio-bilaketarako sistema elektronikoen, paperezko beren homonimoekiko izan behar duten aldea aipatzen hasita, erabiltzaileari laguntza ematerakoan dinamikoak izan behar dutela esaten du —hitz-zerrendak aurkezteko gaitasunarekin, adibidez—; ezaugarri dinamiko horiek sistema hauek tresna malguagoak bihurtuko dituzte, feed-back mekanismodunak, eta erabiltzailearekiko "benetako elkarrizketa" garatzeko interfaze egokiak izango dituztenak.

Hiztegia automatizatzerakoan kontuan eduki beharreko beste alderdi bat da hiztegi hori hizkuntzarekin loturiko bestelako ingurune operatiboetan integratzeko aukerena. Hortxe ditugu, gaur egun, testu-tramenduko programetan integratutako hiztegiak, baina gehienetan berorien erabilera ortografi egiaztapen eta zuzenketa hutsetik harantzago ez doa. Zenbait programak sinonimo- eta antonimo-hiztegia ere eskaintzen ditu linean. "Itzultzailearen lanpostua" delakoak gehiago eskaini behar luke: hiztegi elebakar zein eleanitzekoak, orokorrak zein berezituak, sinonimo-hiztegiak, analogikoak, etab., itzultzailearen eskueran jarrita bere lanean diharduelarik. Arlo honetan erabat egingarritzat jotzen dira testu-idazketarako sistema urgazleak izango diren lineango hiztegiak (Webber, 84).

Hiztegi automatizatuok, beren ahalmen funtzionalei esker, hizkuntzaren ikasketarako ingurune elkarreragileak bihurtzen dira (Fox *et al.*, 80). Hain zuzen, hizkuntz ikasketarako inguruneetan ere hiztegia berebiziko tresna bait da, kasu honetan hizkuntza ikasten duenari zuzendua.

Baina, zer-nolako informazioa eskaini behar du horrelako tresna batek? Bistakoa da informazio ortografikoa —eta ezta ere hitzen sinonimo edota antonimoei buruzkoa— ez dela aski. Gorago esan dugun bezala —zenbait autoreren aipuz—, nahiz eta atzibide sofistikatuagoak gehitu inprimatutako hiztegien informatizazio soila ez da nahikoa. Gure iritziz, hiztegi urgazle integrala diseinatu eta sortu beharra dago, hizkuntz lexikoari buruz orotariko laguntza emango duena, erabiltzaileari atzipen-bide desberdinak eskainiko dizkiona —"edukinaren edo esanahiaren bitartezkoa" barne—, nolabait hizkuntzaren eredu bat (Webber, 84:463) islatuko duen tresna: “. . . *the complete interrelated electronic dictionary is in some sense the model of the language . . .*”. Azkeneko ideia hori dago ikerketa-lan hau motibatu duten ideia nagusien artean. Kay-ren (1984) "dictionary server" delakoaren zenbait aspektu ere egoki letorke testuinguru honetan.

## I.4 HIZTSUA: Hiztegi-Sistema Urgazle Adimenduna.

Gure sistema zertan den azaldu aurretik, ikus dezagun Kay-ren "hiztegi-zerbitzariaren" ideia nondik nora doan.

M. Kay-rentzat (1984) "dictionary server" a hizkuntz tratamendu automatikora zuzendutako sistema lexikalaren eta kontsulta pertsonalerako sortutako hiztegi-ingurunearen arteko bilgunea litzateke. Hiztegi-zerbitzari hori *American Heritage Dictionary* delako hiztegia oinarri hartuta diseinatu eta inplementatu du Kay-k, atzipena arintzearen zenbait indize ezarri dituelarik bertan. Sistemari eusten dion makina ordenadore multzo batekin lotuta dago pakete-konmutazioko sare baten bitartez, ordenadore horien informazioa aise atzitzeko moduan. Bere hitzaldiaren erreferentzia laburrean —1984eko COLINGekoa—, Kay-k bere sistemak garai hartan zituen muga estuak aitortzen ditu (estuak, zeren atzipen-bideek hitzen ortografia —forma kanoniko zein flexiodunetan—, informazio etimologikoa, definizio-testuak, sinonimoak, etab. besterik ez baitzuten eskaintzen), baina, aldi berean, sistema klase honek thesaurus erako atzibideak eskaini beharko lituzkeela azpimarratuz, hots, erabiltzaileak une batean behar duen "*le mot juste*" hori topatzeko erraztasunak hain zuzen ere. Roget-ek bere Thesaurus (1852) delakoarentzat pentsatutako funtzionalitatearen parekoa den atzipen era hori, hiztegi-zerbitzariaren tankerako sistema batekin erdiets liteke, eta lehenbiziko aldiz hobetu. Bukatzeko, hona zer dioen Kay-k bere idatzian (461. orr.):

What stands in the way of dictionary services of far greater utility than even the largest currently available books is not technological inadequacies, or even shortcomings of linguistic or lexicological theory, so much as the courage and foresight to invest in lexicographic databases of radical design.

Lenders-ek berriz (1990:48), Kay-ren hiztegi-zerbitzari horren ideiarekin loturik, hiztegia sistema aditutzat hartzen du:

Dictionaries are a certain kind of expert system. They contain knowledge either about the general vocabulary of a language, or about the terminology of a specific field. Since the normal usage of a dictionary is a question/answering process, an obvious question is whether it is possible to produce dictionaries in the form of expert systems. There is a tendency to understand a dictionary not in the traditional way as a "data base" but as a "knowledge base", which is a part of a lexical expert system; however, it is clear, that this is not a fundamentally different object but a special point of view and a special way of access. Martin Kay (1984) called such an instrument a "dictionary server", and there are already several approaches discussing and constructing "lexical knowledge bases".

Gurean garatutako sistemak aipatu berri ditugun ideia horietatik edan nahi luke. Hizkuntz erabilerako ingurune batean ezarritako "hiztegi-zerbitzari" modukoa edo bailitzateke gurea.

## I. kapitulua

Hiztegi-laguntzarako sistema adimendun baten ideia da ikerketa hau eragin dutenetan nagusiena<sup>1</sup>.

HIZTSUA, **Hiztegi-Sistema Urgazle Adimenduna**, giza erabilerarako hiztegi-sistema elebakarra da. HIZTSUAREN helburua, hizkuntza erabiltzen den egoera batean hiztegiaren erabilera eta kontsulta erabiltzaileari erraztea litzateke. Kontzeptuen esanahiazko ezagutza eskuratzeko, kontzeptuen ulerkuntzan azken finean, erabiltzaileari laguntza ematea alegia.

Halako sistema bat definitzen saiatu gara proiektu honetan —prototipo bat ere eraikiz—, zer-nolako ezaugarriak izan beharko lituzkeen eta zein funtzionalitate eskaini beharko liokeen erabiltzaileari zehaztuz.

Sistemaren prototipoa eraikitzearen, ikerketa-lan enpirikoaren oinarritzat hartu dugun hiztegi errealetik abiatu nahi izan dugu. Frantsesezko<sup>2</sup> hiztegitxo elebakar baten gainean burututako ikerketa-lan enpiriko horren emaitza eta sistemak eskaini beharreko funtzionalitea izan dira hiztegi-ezagutzaren errepresentazio-ereduaren hautaketan kontuan edukitako alderdi nagusiak.

HIZTSUAREN diseinu eta inplementazio informatikoan, honako aldeok eduki behar izan ditugu kontuan:

- *Ezagutza lexiko-semantikoaren eskuratzea*, hiztegi arrunt batetik abiatuz. Hiztegi arrunt batean dagoen ezagutza lexiko-semantikoa erazteko prozesuaren automatizaziorako lan-metodologia bat hartu dugu proiektu honetan. Frantsesezko definiziodun hiztegi elebakar bateko definizioetan dagoen ezagutza edo jakintzaren erazketara jo dugu batez ere gure proiektuan (hemendik aurrera, hiztegi elebakar bateko definizioetan dagoen jakintzaz arituko gara batik bat txosten honetan), erazketa horretarako LNPko ohizko eta ez hain ohizko lan-prozedurak erabiliz. Paperezko hiztegi bat izan dugu bada, egindako ikerketaren oinarri: ikerketa honek iturburuko informazioaren interpretazio- eta birformulatze-lan handia eskatu du, eskuz egitea pentsa ezina dena zalantzarik gabe. Horiek horrela, automatizagarri bilatu duguna automatizatzen saiatu gara —egile askok dioen legez dena ez baita automatizagarri arlo honetan—, gainerantzekoetan prozedura erdiautomatiko edo eskuzkoetara jo beharrean aurkitu garelarik.
- *Ezagutzaren errepresentazioa*. Hiztegi-ezagutzaren errepresentazioaren arazoa inportantzia handikoa dugu. Zein da adierazpiderik egokiena jakintza guzti hori

---

<sup>1</sup> (Artola & Evrad, 90)eko "hiztegi adimenduna" ("dictionnaire intelligent").

<sup>2</sup> Tesi-lan hau Toulouseko ENSEEIHT ikastetxeko F. Evrad irakaslearekiko lankidetzan egina da.

errepresentatzeko? Hiztegien dauden kontzeptuak, beren ezaugarriak, beren arteko erlazioak, egoki errepresentatu beharra dago.

Jakina, ezagutza horrekin egingo den erabilera ere kontuan hartu beharrekoa da. Halako hiztegi baten erabiltzaile-irakurle izango denak zer-nolako jarrera izango duen aztertu eta sistemaren errepresentazio- eta erabilmoduekin egokitu egin beharko da.

Hiztegien eta, batez ere, hiztegi-sarreraren definizio-testuen gainean egindako ikerlan enpirikoak eraman gaitu ezagutza horren errepresentaziorako proposamen bat egitera. Horretarako, lexikografoek definizioak idazterakoan erabilitako formula tipikoak —eta ez hain tipikoak— aztertu ditugu aukerako diren kontzeptuarteko erlazioak ulertuz, erabiltzailearen ohizko arrazoibidea eta kontsulta-modua posible egingo dituen errepresentazio egokia ematearren.

Adierazpide edo errepresentazio hori erabili dugu *hiztegi ezagutza-basearen (HEB)* prototipoa eraikitzeke, horretarako adimen artifizialeko teknikak erabiliz. Hiztegiaren isla den kontzeptuarteko erlazio lexiko-semantikoen multzoa errepresentatzeko aukeratutako adierazpidea frame-sistema edo herentzia egituratundun sare semantikoen (SI-Net; Brachman & Schmolze, 1985) hurbila da.

- *Alderdi deduktiboa*: eskuratutako ezagutzaren ustiatze adimenduna. Esan bezala, hiztegia askotariko ezagutza daukan "ontzia" dugu. Ezagutza horren parte handi bat esplizitua dugu hiztegi-egituretan eta erabiltzailearentzat ikusgai, baina beste parte bat, agian aurrekoa baino handiagoa, *ezagutza inplizitua* dugu (Boguraev & Levin, 88). Adimendun izeneko hiztegi-tresna egin nahi badugu, ezagutza inplizitu guzti hori argitaratzeko moduan ipini behar da. Beraz, iturburu-hiztegitik datorren ezagutza eskuratu eta nolabait errepresentatu ondoren, "ezkutuko" ezagutza hori esplizitu bihurtzeko gaitasuna —aldez aurretik nahiz erabiltzailearen galdera baten kariaz— izango duen dedukzio-ahalmena eman behar zaio sistemari. Aldez aurretik egingarri diren dedukzioak direla eta, *HEBaren aberasketa dinamikoa* delakoa burutzen duen prozedura multzoa zehaztu eta inplementatu dugu; aberasketa hori kontzeptuarteko zenbait erlazio esplizitu egitean datza, honela ondorengo prozesuak erraztu eta azkartuz.

Azken finean, sistemak gauza izan beharko luke bere funtzionamenduan erabiltzailearen antzera jokatzeko, erabiltzaile hori hiztegi azimentsuki baliatzean esplizituki ez dauden baina inferi daitezkeen informazioak eskuratzeko gauza baita. Horretarako, atzipen-bide desberdinak eta inferentzi mekanismo ahaltsuak eman behar zaizkio, sistema benetako sistema aditua bihurtuko dutenak.

## I. kapitulua

- *Prototipoa*: hiztegi-sistema honen sorkuntza eta garapena helburu, software-arkitektura diseinatu eta prototipo bat inplementatu dugu. HIZTSUAren gune den prototipo horren eraikuntzarako abiapuntu iturburu-hiztegiarekin eratutako hiztegi datu-basea izan dugu; prototiporako hiztegi horren azpimultzo bat aukeratu beharrean izan gara.
- Kontzeptuen *ulerkuntzarako sistema urgazlearen diseinua*. Psikolinguistikan oraindik indarrean dirauen ikuspuntua emanez, Quillian-ek (1968) hizkuntzaren lexiko-sistema —edo hobe, giza memoria semantikoa— elkarrekin erlazionatutako kontzeptuen multzoa dela erakutsi digu. Sare semantikoen sorreran garrantzi handia izan zuen ideia hori dago gure sistemako ezagutzaren errepresentaziorako aukeratu dugun ereduaren azpian ere. Hori alde batera utziz, egile beraren beste lan bat interesatzen zaigu orain. Izan ere, Quillian-ek (1969) bere "Teachable Language Comprehender" lan ezagunean ingelesezko testuak "ulertzen" ikasi duen programa deskribatzen baitu. Programa hori bere tesi-lanean (Quillian, 68) garatutako memoria semantikoaren ideiarekin erabat lotua dugu, zeren testuaren ulerkuntza-prozesua —pertsonen hizkuntz ulertzeari buruz zuen teoriaren arabera— "ulerkizun den testuan egindako mundu errealari buruzko baiezteak" sistemaren "memoria semantiko" zabaleko mota-kontzeptuekin harremanetan jartzean bailitzake. Horrela, "Teachable Language Comprehender" lanean, ongi ulertutako testuen esanahia kodetzeko memoria semantikoa kodetzeko erabilitako formatu berbera erabiltzen du.

HIZTSUAK ez du hizkuntza ulertuko duen sistema izan nahi, testua bere kasa ulertuko duen erabiltzaileari laguntza eskainiko dion sistema baizik. Erabiltzailearen testu ulertze hori erraz liteke sistemarekiko elkarrekintza baten bitartez; elkarrekintza horretan erabiltzaile-sistema elkarrizketa burutuko litzateke, non erabiltzailearen eskariari sistemaren ekintza jarraituko bailitzaiokie. Sistema bere hiztegi-ezagutza eta erabiltzailearen ulertze-egoera integratzen saiatu behar litzateke, erabiltzaileak duen ulertze-zailtasunari buruz arrazoitu eta ebazteko helburuarekin. Sistemaren komunikazio-moduluak elkarrizketa hori kudeatzeko gai izan behar du, erabiltzaileari, bere galdera zehatz dezan, oinarritzko eskari-aukera eta eskari-konbinazio desberdinak eskainiz. Sistemak, bere ezagutzari esker, lagundu egin diezaiokie erabiltzaileari eskaria zehazte horretan, anbiguotasun lexikaleko kasuaz bereiziz, erabiltzailearen ortografi hutsak igarritz, sinonimo-zerrendak eskainiz, etab..

Horrexegatik da inportantea erabiltzaile eta sistemaren arteko komunikazio-modulu hori benetan ahalmentsua izatea; berorren ardura baita, aipatutako elkarrizketa

horren kudeaketaz gain, sistemaren ezagutzaren ustiapena askotara, hitz batean esateko, adimentsu azaltzea erabiltzaileak hiztegiara jo dezanean.

- *Oinarrizko eskari-funtzionalitatea.* Aipatu ditugun oinarrizko eskariei dagokienean, sistemak erabiltzaileari eskaini beharreko funtzionalitate multzoa definitu eta inplementatu dugu. Funtzionalitate hori egoki definitzearen, hiztegien giza erabilera ingurune elebakarrean zertan den aztertuko duen ikerketa enpirikoa beharrezkoa da. Gure proiektuan tankera horretako lan bat egin da, oso maila txikian alabaina, eta lantxo horren fruitu ditugu definitu eta zehaztutako oinarrizko funtzionalitateak.

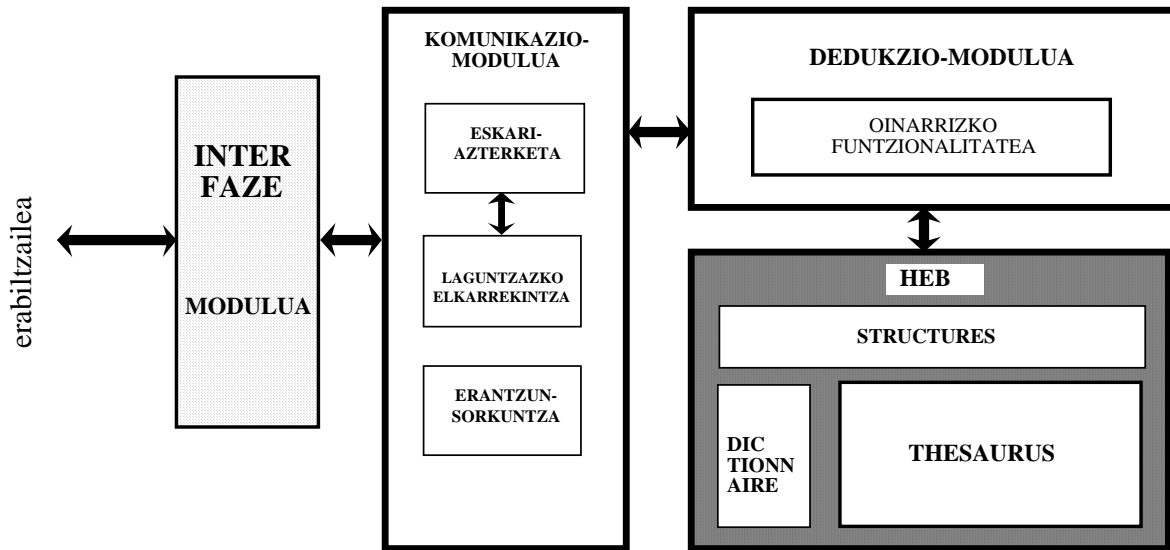
Lehen ere esana dugu, HIZTSUAK izan behar duen funtzionalitate horrek baldintzatzen du nolabait aukeratutako errepresentazio-eredua. Bestalde, erabiltzailearen eskueran ezarritako funtzionalitate hori hiztegi datu-base arrunt batek eskainitakoa baino "zerbait gehiago" izatea nahi dugu. Gure iritziz, sortutako funtzionalitate horri esker, sistemak ematen dio erabiltzaileari ezagutzara iristeko zenbait bide, sistemak duen hiztegi-ezagutza horretan inplizitu legokeena ere eskuratzeko aukera emanez.

- *Erabiltzaile mota desberdinak:* proiektuan garatutako sisteman hiztegi-erabiltzaile arrunta izan dugu gogoan batik bat, hizkuntza ongi ez dakien eta bere ikaste-prozesuan hiztegiak baliatzen den erabiltzailea zehatzago esanda. Alabaina, sistemaren sorkuntzan, eta batez ere ezagutzaren adierazpidea aukeratzekoan, ez ditugu bestelako erabiltzaileak ahaztu nahi izan —lexikografoa bereziki—, zeinentzat sistemak oso funtzionalitate desberdina eskaini behar bailuke. Bestelako funtzionalitate hori aztertu, eta HIZTSUAREN gunea hiztegi-gintzan ari den lexikografoari baliagarri lekiokoenentz —sarrerak idatzi eta egiaztatzeko, esate baterako— ikustea litzateke proiektu honen ondoren ikerkizun geratzen den bideetariko bat.

## **I.4.1 Arkitekturaren diseinua.**

### **I.4.1.1 Sistemaren osagaiak.**

HIZTSUARENTZAT asmatutako software-arkitektura ikus daiteke I.1 irudian. Paragrafo honetan eta hurrengoan arkitektura horren ikuspegi orokorra eman nahi genuke, sistemaren osagai bakoitza deskribatuz eta, labur bada ere, hizkuntz erabilerako egoera batean sistema-erabiltzaile elkarrekintza suertatzen denean modulu desberdinen arteko lankidetzaz zertan den azalduz.



I.1 irudia.- HIZTSUaren software-arkitektura.

Irudian HIZTSUaren prototiporako bururatu dugun software-arkitektura osatzen duten lau moduluak ikus daitezke: Hiztegi Ezagutza-Basea, Dedukzio-Modulua, Komunikazio-Modulua eta Erabiltzaile-Sistema Interfazea.

- *Hiztegi Ezagutza-Basea* (HEB) da lehenago aipatu dugun hiztegia, nolabait moldatua. Sistemarako hautatu dugun ezagutza-errepresentazioaren eredua ezagutza-base honetan dago eratua. HEBa hiru ezagutza-azpibase desberdinetan antolatu dugu. Batetik THESAURUS izenekoa, zeinek § I.4eko sarreran aipatutako lexikoaren eredu erlazionala islatzen baitu; ezagutza-base honen itxura frame-sare batena da, non kontzeptu bakoitza errepresentatzeko frame bat erabiltzen den eta frame horien atributuek kontzeptuak sareko beste kontzeptuekiko dituen erlazio lexiko-semantikoak errepresentatzen baitituzte. Bigarrenik DICTIONNAIRE<sup>1</sup> delako azpibasea daukagu, hiztegi-sarreretatik dagozkien kontzeptuetarako aplikazioa dena, erabiltzaileari hitzetik kontzepturako sarbidea eskaintzeko helburuarekin eratua. Eta azkenik STRUCTURES dugu, HEBko egitura desberdinen definizioen gordailu —beste baseetako unitateen atributuen definizioa, atributu horietako balioen herentzi moduaren espresioak, etab.—, HEBan bildutako ezagutza maneiatzeko beharrezko den metaezagutza multzoa edo osatzen duena.
- Sistemaren arrazoiketa posible egiten duen dedukzio-mekanismoen multzoa da *Dedukzio-Modulua*. Modulu horretan sartu ditugu sistemak eskainitako *oinarrizko funtzionalitatea* zein haren portaera adimentsua modelizatzeko ezarritako inferentzi mekanismoak. Aipa ditzagun hemen, eskuarki STRUCTURES azpibasean ezarri

<sup>1</sup> Hemendik aurrera, eta txosten osoan zehar, egindako inplementazioan nolabaiteko "gauzatzea" duten entitateen deiturak eta frantsesez utziko ditugu.



diren mekanismo horien adibide gisa, *herentzia* eta *erlazio lexikalen konposaketa* (ikus III. kapituluan).

Gainera, prototipoaren hasierako eraikuntzan soilik —edo eraiki eta berehala— gaitu diren prozedura deduktiboen multzo bat bada sisteman: arrazoiketa-mekanismo horiei esker burutu ahal izan dugu *HEBaren aberasketa dinamikoa* (ikus § III.2) dioguna; aberaste-prozesu hori eraikuntza taxonomikoan eta analizatutako definizio-testuen desanbiguatze lexikal eta sintaktikoan datza gehienbat.

- *Komunikazio-Modulua* hiru azpimoduluk osatzen dute: *Eskari-Azterketa*, *Laguntzazko Elkarrekintza* eta *Erantzun-Sorkuntza*. Eskari-Azterketako moduluak erabiltzailearen eskariak aztertu eta Dedukzio-Moduluari bidaltzen dizkio. Erabiltzailearen galderak, normalean, hitzen funtzioan adieraziak etorriko dira, baina HEBan onartzen diren eskariak adiera edo zentzuen funtzioan adieraziak izan behar dute. Eskari-Azterketako modulu hori izango da, HEBko Dictionnaire azpibaseko informazioaz baliatuz, arazo hori ebatziko duena. Halaber, modulu honi dagokio erabiltzailea ari den erabilegoera eta sistemak daukan hiztegi-ezagutza nolabait integratzea. Beraz, eskari-azterketako modulu horrek erabiltzailearekiko komunikazioa bideratu beharko du sarri, erabiltzaileari bere galdera zehatz dezan eskatzeko hain zuzen ere. Komunikazio hori Laguntzazko Elkarrekintzako moduluari esker bideratuko da; erantzun ahal izateko erabiltzailearen laguntza behar duenean ere eskatuko du sistemak modulu horren laguntza. Erantzun-Sorkuntzako modulua azkenik, sistemak erabiltzailearentzako erantzunak sortu behar dituen erabiltzen da.
- Bukatzeko *Interfaze-Modulua* aipatu behar dugu. Modulu honi esker gauzatuko da erabiltzailea —eta bere erabilegoera— eta sistemaren arteko interfaze fisikoa. Sistemak leiho-mekanismo adiskidetsu eta erabilkorra beharko du, lengoia naturalazko komunikazioaren ezean erabiltzaileari bere kontsultak adierazteko bide egokia eskainiko diona. Horretarako bururatu dugun menu-sistemak —Komunikazio-Moduluaren manupean— erabiltzailearen ekintzen aurrean dinamikoki erantzun beharko du, eta erabiltzaileari bere kontsultak atseginkiro egiteko bidea eman (ikus V. kapitulua).

#### **I.4.1.2 Moduluen arteko lankidetzak.**

HIZTSUArekiko elkarrekintza erabiltzaileak bere galdera egiten duenean hasi ohi da. Horretarako Interfazeak erabiltzailea dagoen erabilegoerari dagozkion aukera egokiak eskaini behar izan dizkio aurretik. Aukera eskaintza hori egiteko sistemak kontuan izan

## *I. kapitulua*

beharko du erabiltzailea zein eginkizunetan ari den eta zer-nolako ekintzak egin dituen bere eskaria egiterakoan.

Behin eskaria adierazi eta gero, Eskari-Azterketako Moduluak izango da eskari hori oinarrizko funtzionalitateen sintaxia beteko duen adierazpena bihurtu eta Dedukzio-Moduluari bidaliko diona, erantzuna lortzeko asmotan. Horretarako badaiteke Eskari-Azterketak Hiztegi-Ezagutza Baseko informazioa atzitu beharra izatea: esaterako, adiera konkretu bat zehazteko, erabiltzailearen errore edota uste okerrak zuzentzeko, laguntza hobea eskaintzarren hitz-zerrendak eratzeko, eta beste. Hortaz, Eskari-Azterketa izango da hor sor daitekeen erabiltzailearekiko elkarrizketa kudeatuko duena. Lanbide horretan Eskari-Azterketa Laguntzako Elkarrekintzako Moduluaz baliatuko da, bai erabiltzaileari lortutako emaitzak aurkezteko eta bai haren eskaria hobeki zehazteko eskatzeko ere.

Ekinbide guzti horren ondorio gisa Eskari-Azterketako moduluak eskari bat lortu — oinarrizko funtzionalitateen funtzioan adierazia— eta Dedukzio-Moduluari igorri behar dio. Horrela, Dedukzio-Moduluaren barruan prozesu deduktiboa jarriko da martxan eta, printzipioz, erantzun batera iritsi beharko du. Prozesu horrek, HEBan ezarritako dedukzio-mekanismoei esker —oinarrizko eskari-funtzioak, herentzia, erlazio lexikalen konposaketa, etab.—, DICTIONNAIREko eta, batez ere, THESAURUSeko unitateak atzitu dituzten prozedura desberdinak jarriko ditu jokoan. Atzibide horretan STRUCTURES ezagutza-azpibasea izango da gidari.

Dedukzio-Moduluak, erantzuna lortzeko erabiltzailearen laguntza behar izanez gero, Laguntzako Elkarrekintzakoari egin diezaioke dei eta erabiltzailearekiko elkarrizketa-prozesu bati ekin.

Erantzuna prest dagoenean Erantzun-Sorkuntzako moduluari bidaliko zaio eta honek, Interfazea bitarteko, erabiltzaileari aurkeztuko dio —atseginagoa izan dadin beharrezko itzulketak edo birmoldaketak eginik—, ondoren suerta daitekeen elkarrizketari jarraitzeko erabilegoera, emandako erantzuna eta abar kontuan hartu beharko dituelarik.

### **I.4.2 HIZTSUAren atzipen-bideak.**

Oraingoz HIZTSUArentzat asmatu ditugun atzibideak menu eta leihozko interfazea eta eskari-lengoaia xume bat lirateke.

Txosteneko V. kapituluan azalduko dugu erabiltzaile arrunt batentzat asmatutako HIZTSUAren interfaze atseginaren diseinua. Esan dugun bezala Komunikazio-Moduluak kudeaturiko leiho- eta menu-sistema dinamiko bat litzateke. HIZTSUAren erabilegoera eta erabilegoera horren gorabehera guztiak hartu beharko lituzke kontuan sistemak erabiltzailearekiko elkarrekintza horretan.

Interfaze honen azpian dago hiztegi-sistemaren eskari-lengoaia espezifikoaren mamia osatzen duen eragiketa primitibo eta oinarrizko funtzioen multzoa. Oinarrizko funtzionalitate horren zehaztapenari eskari-konposaketarako zenbait aukera erantsi zaio (IV. kapitulua), halako lengoaia baten adierazkortasuna eta ahalmena handitzearren.

HIZTSUA, bestelako testuinguruetan edota beste erabiltzaile batzuek erabiltzekotan, funtzionalitate berrien premian legoke. Gogoan ditugu, batez ere, erabiltzaile lexikografoa edo HIZTSUAren erabilera LNPko programa batean.

## **I.5 Ebaluazioa.**

### **I.5.1 Implementazioaren emaitzak.**

HIZTSUA inplementatzeko adimen artifizialerako sortutako ingurune batez bailatu gara: KEE™ (Intellicorp). LISPen gainean antolaturik dagoen software horrek eskaintzen ditu, besteak beste, frame erako egituretan oinarritutako ezagutzaren errepresentazioarako ingurune aberatsa, klase eta instantzia bezala erazagututako objetuen arteko hierarkiak eratze bidea (goragoko mailetatik beheragokoetarako atributu-herentziarekin), eta abar. Arrazoiketa eta problema-ebazpenerako tresnen aldetik berriz, KEE™k aurreranzko eta atzeranzko inferentzi mekanismoak dituen erregela-sistema, objetuei zuzendutako programazioa, iratxo ("demon") edo balio aktiboen esleipena, eta ezagutza-baseen konsistentzia eta "mundu" desberdinetan arrazoiketa egiteko aukera ematen duen egiaren mantenimendurako sistema (TMS). Berebat, KEE™k interfazeak eraikitze —sistemaren garapenerako zein azken erabiltzaileentzako— eta irudiak egiteko erraztasunak ere ematen ditu.

Ingurune hori hautatzerako antzeko beste zenbaitekin erkatu genuen —Knowledge Craft™-ekin, batez ere—; ingurune horietako batzuek ezagutzaren errepresentazioaren aldetik adierazkortasun handiagoa eskaintzen badute ere, ez dute, gure ustez, KEE™k sistema gisa duen batasuna eta osotasuna. Osotasun hori eta KEEren malgutasuna izan dira ingurune hau aukeratzera eraman gaituzten ezaugarri nagusiak.

Eraikitako prototipoa dela eta, alderdi batzuk beste batzuk baino landuago daudela esan behar. HEBa eta Dedukzio-Modulua lirateke prototipoan "landuen" dauden atalak. Prototipo honen parte sendoenak beraz, analizatutako hiztegitxotik eskuratu informazioaren errepresentazioa, hasierako eraikuntzaren fasean —eta ondoren— ezagutza horren gainean ezarri mekanismo deduktiboak, eta sistemak eskaintzen dituen oinarrizko funtzionalitateen zehaztapen eta inplementazioa ditugu. Komunikazio-Moduluari dagokionez berriz, gaur egun dugun prototipoan oso maila xumean ezarria dagoela esan behar.

Proiektuaren garapenean erdietsitako xedeen artean honakoak aipatuko ditugu:

- *Hiztegitxo elebakar baten azterketa*, bertako definizio-testuetan gorderik dagoen ezagutza lexiko-semantikoa erauzi eta definizioetan inplizitua dena esplizitu bihurtzeko —beharrezko izanez gero— gai den moduan kodetuz.

Hiztegitiko ezagutza-eskuratze prozesu horietarako *lan-metodologia* bat diseinatu dugu. Metodologia horretan zenbait urrats bereizten dira: definizioetako hitzen maiztasun-azterketa, kategori etiketatzea eta desanbiguaketa lexikala, kategori eskemen zerrenden lorpena, zerrenda horien behetik gorako analisia maila sintagmatikoko egiturak lortzearren, erlature berezien azterketa empirikoa (definizio-formula estereotipatuak), definizioen analisi sintaktiko-semantikoa eta egitura semantikoen eraikuntza.

Analizatzaile bat inplementatu —pattern-matching hierarkiko partzialean oinarritua— eta iturburu-hiztegitiko izen, adjetibo eta aditzen definizioak analizatu dira —sintaktiko eta semantikoki—. Guztira 22215 definizio —beste hainbat adierari dagozkionak— dira horrela automatikoki tratatutakoak. Analiaren emaitza zehatzak ematen hasita, lehen-lehenik esan behar da analisi partzialeko mekanismoa izan arren LPPLko izen-definizioen %57.76a, aditzen %79.8a eta adjetiboen %69.04a gertatu direla erabat analizatuak, hau da definizioko hitz guztiak parekatuak. Hutsegite maila aski txikia izan da: izen-definizioen %7.9a, aditzen %2a eta adjetiboen %16.35a. Analisi hori burutzeko 159 patroia definitu ditugu.

Analisi-patroi bakoitzari egitura semantikoa eratzeko erregela bat lotu zaio eta hori erabili da HEBaren hasierako eraikuntza automatikoki egiteko garaian.

- Sistema urgazlearen gunea den *hiztegi ezagutza-basearen diseinua*. Hiztegi-ezagutzaren errepresentaziorako herentzidun sare semantikoen erako eredu bat zehaztu eta ezagutza horren egituratze-modua asmatu dugu.
- Analizatutako hiztegi horren azpimultzo batekin HEBaren *prototipo bat eraiki dugu*. Prototipo hori automatikoki eraiki da eta berorren abiapuntutzat 2607 definizio analizatu hartu dira, 6130 kontzeptuko sarea eratuz (THESAURUS azpi-EBa). Sare horretako nodoen arteko elkarlotze maila —batez beste eta HEBaren hasierako eraikuntzaren ondoren— 3.21 arkukoa da kontzeptuko (erlazio lexiko-semantikoen adierazpide diren arkuak). DICTIONNAIRE azpi-EBak berriz, 2400 unitate ditu; horietako bakoitzak hiztegitiko sarrera-hitz bat errepresentatzen du eta THESAURUSean dagozkion kontzeptuekin zuzenean loturik dago.

- *Dedukzio-mekanismo sail* baten ezarrera. Mekanismo horien bitartez ondokoak egiteko aukera egon da (edo dago):
  - a) Iturburu-hiztegitik eskuratu ezagutzen *aberasketa*, hiztegian —definizioetan batik bat— implizitua zena esplizitu bihurtuz.

Aberasketa horri esker, zeinetan sinonimia eta erlazio taxonomikoak ustiatu baitira gehienbat, prototipoko arkuen kopurua %10.7 gehitzea lortu da, eta honela, azkenean lortutako sarean, kontzeptuarteko erlazioen kopurua 21800era iritsi da.
  - b) Erabiltzailearen galderapeko arrazoiketa: HEBaren *aberasketa dinamikorako* aukera ematen duten mekanismoak ezarri eta oinarritzko *eskari-funtzionalitate sail* baten zehaztapen eta inplementazio informatikoa egin ditugu.
- HIZTSUAren *software-arkitekturaren diseinua*. Hiztegi-Sistema Urgazle Adimendunaren arkitektura diseinatu dugu. Inplementazio informatikoan alabaina, esan bezala, sistemaren gunea den "ezagutza-basea" da gehien landu dena. Sortutako sistemaren osagaiak HEBa, Dedukzio-Modulua, Komunikazio-Modulua eta Interfazea dira.

## **I.6 Txostenaren eskema.**

Ondokoan, txosten hau hiru partetan dago banatua.

Lehenbizikoan, II. kapitulan —lexikografia konputazionalaren nondik norakoak eman asmoz idatzitako sarrera laburraren ondoren— hiztegi ezagutza-basearen eraikibidea, hiztegitxo arrunt batetik abiatuz, nolakoa izan den azalduko dugu. Kapitulu horretan deskribatuko dugu lan mota horretarako ahalik orokorren nahi genukeen hiztegi bateko definizioen analisirako metodoa. HEBaren sorbidea bera zertan izan den aditzera eman ondoren, eta erabilitako prozeduren mugen ebaluazio bat eginda, III. kapitulan HEB horrentzat hautaturiko ezagutza-errepresentazioaren eredia deskribatuko dugu, bertan ezarritako dedukzio-mekanismoak eta hasierako eraikuntzaren ondotik egindako aberasketa dinamikoa zertan diren azaltzearekin batera.

Bigarren partean proiektuan sortu dugun hiztegi-sistema urgazlearen funtzionalitateaz arituko gara. Bi kapitulutan —IV.a eta V.a— definitu eta inplementatu ditugun oinarritzko funtzioez eta erabiltzaileak sistemarekin komunikatzeko izango dituen aukerez —erabiltzaile

## *I. kapitulua*

eta sistemaren arteko interfazeaz— ihardungo dugu. HIZTSUAren erabilera erakusteko adibide komentatuak txosten honen F eranskinean datoz<sup>1</sup>.

Txostenaren azken partean berriz, egindako lanaren etorkizuna eta ondorioak —VI. eta VII. kapituluak— izango ditugu mintzagai. Bertan azalduko ditugu, HIZTSUA beste ingurune batzuetan aplikatzeko posibilitateekin batera, lan honen ondoren ikergai geratutakoak eta, gure ustez, aurrera egiteko bide interesgarrienak diratekeenak.

---

<sup>1</sup> Txosten honen eranskinak aparteko liburuki batean bildu dira.

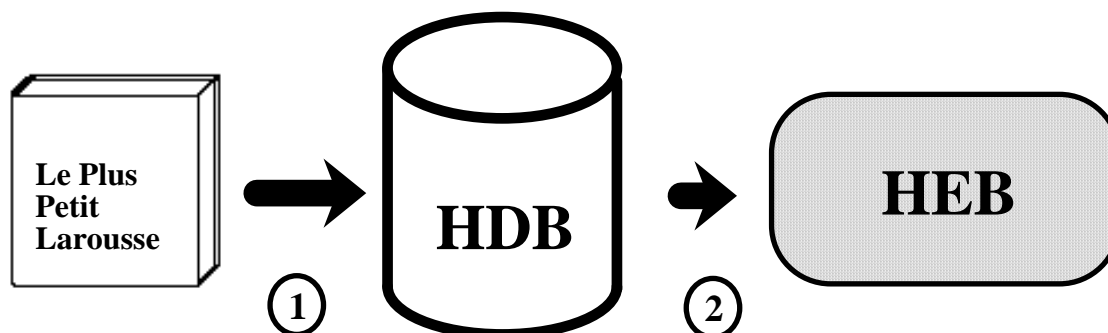
# LEHEN PARTEA: HIZTSUA ERAIKITZEKO PROZESUA

## **II. Hiztegi arruntetik Hiztegi-Ezagutza Baserantz.**

Bigarren kapitulu honetan deskribatuko dugu HIZTSUAren gune den Hiztegi-Ezagutza Basearen (HEB) eraikuntzan jarraitu dugun behetik goranzko bidea. Lexikografia konputazionala eta, bereziki, MRDetatiko ezagutza-eskuratzea zertan diren labur azaldu ondoren —kapituluko lehen bi ataletan—, iturburu gisa erabilitako hiztegia, hiztegi datu-basearen eraikuntza eta beronen gainean egindako azterketak azalduko ditugu.

Eraikuntza-prozesu horren ikuspegi orokorra ematearren, labur-labur azalduko ditugu orain frantsesezko hiztegitxo batetik abiatu HEBaren eraikuntzan jarraituriko kronologia eta metodoak.

II.1 irudian ikus daitezke eskematikoki iturburu-hiztegiaren gainean egindako eraldaketak, paperezko bertsioetik hasita HEBraino. Lehenengo etapan —1, irudian—, hiztegiko datuak datu-base erlazional batean grabatu egin dira: datu-base horri HDBa esango diogu (ikus § II.3 eta II.3.1). Datu-base honetako hiztegi-informazioen gainean egindako zenbait azterketa eta eraldatze —definizioetan, batez ere— direla medio, HDBa bera aberastua irtengo da (ikus § II.4etik II.5.3ra). Azterketa hauen ondorio gisa hiztegi-ezagutza errepresentatzeko proposamen bat egiten da: errepresentazio hori emaitzat duen definizioen analisi sintaktiko-semantikoa (§ II.5) izango da HEBaren prototipoa eraikitzena —2 etapa, irudian— eramango gaituena (ikus § II.6tik II.6.2ra).



II.1 irudia.- HEBaren eraikuntzako urratsen eskema.

## II.1 Lexikografia konputazionala: sarrera gisakoa.

Txosten honen lehen kapituluan esan bezala, lexikografia konputazionalaren arloan hainbat alderdi desberdin bildu ohi da.

Egungo lexikografian ordenadorea erabiltzea ohizko bihurtu da (Landau, 84; Sinclair, 87). Hirurogeigarreneko hamarkadatik hona metodo informatikoak sartu dira hiztegi gintzan eta gaur egun guztiz naturala iruditzen zaigu (Lenders, 90). Aitzitik, hiztegi baten eraikuntzan alderdi nagusia den hitzen esanahiaren deskribapena hiztegi gileei dagokie erabat, eta gainera hiztegi gile hauek lan horretarako ordenadorea erabiltzea ez dute hain argi ikusten. Ordenadorearen benetako erabilera beraz, lexikografi lanetarako euskarri izatera mugatua da: datuen grabazioa, corpusen lanketarako konkordantzi programak, testuetako informazio-bilaketarako sistemak, hiztegi gintzarako laguntza-inguruneak, etab. Ordenadorez lagunduriko lexikografia (CAL, Computer Aided Lexicography) da beraz.

Dena dela, gure proiektuaren motibazio nagusia ez da izan ordenadoreen erabilera lexikografian. Hori baino, W. Lenders-ek (1990) bere artikularen laburpenean adierazitakoa litzateke, funtsezko galdera hau egiten duenean: hiztegi gietako material lexikografikoaren bestelako erabilera batera eramango gintuzkeen *lexikoaren egituratze- eta eskuratze-eredurik* ez ote dago?

Hiztegi en edukinak ez duela balio handirik ez hizkuntzalaritzarako eta ez adimen artifizialerako diotenei erantzunez, K.C. Litkowsky-k (1980) duela hamabi urte azpimarratu zuen "lexikografoek egindako lanaz probetxatu eta ez berregitearen" premia, hiztegi gietan dagoen ezagutzaren araketak, LNPko sistemetako jakintza lexikalaren errepresentazio eta erabilerarako eredu baten garapenean izan zezakeen garrantziaz ziharduela.



Beranduago, Italiako Grosseton izandako mintegi batean —"Automating the Lexicon", 1986— ateratako ondorioetan<sup>1</sup>, iritziz hauek berrindartu egin ziren. A. Zampolli-k (1987) lexikologia eta lexikografia konputazionalerako hain interesgarri diren ondorio horien laburpena jasotzen du bere lanean.

Ordenadorez lagunduriko lexikografia alde batera utzirik, LNPrako lexiko-sistemen eraikuntzarako laguntzen arloa hurbilago dugu. Egile askoren lanetan azaltzen dira datu-base lexikal handien (LDB, Lexical Databases) eraketari begira egindako saioak, datu-base horien helburu nagusia LNPrako sistemen lexiko-iturri izatea delarik, bestelako erabilerarik kontuan hartzen baldin bada ere. Horien artean aipatu behar Italiaren Datu-Base Lexikal Multifuntzionala (IMLDB, Italian Multifunctional Lexical Database, (Zampolli, 87)). Elkarrekintza, malgutasuna, informazio lexikalaren berrerabilgarritasuna, dimentsioaniztasuna, eskola linguistiko desberdinekiko neutraltasuna, etab. dira beti aipatzen diren tasunak gaur egungo proiektu handi hauez hitz egitean. (Calzolari, 84a; 86; 90a; Calzolari & Picchi, 88) dirateke lan horien aurrekariak.

Ikuspuntu metodologikotik, aipa dezagun labur (R.J. Byrd *et al.*, 87) lana. Artikulu horretan hiztegien edukina automatikoki analizatu eta arakatzeko metodologia eta zenbait tresna ematen dira<sup>2</sup>. Hiztegiak euskarri magnetikora "botatzerakoan" aintzat hartu beharreko printzipio nagusia den "*do not throw away anything*" delakoa eman ondoren, egileek beren hiztegi-biltegitze eta -atzipenerako metodoa deskribatzen dute (DAM, Dictionary Access Method), zein, beren iritziz, datu-base arruntak baino egokiago bait da. DAMpeko fitxategiak —paperezko hiztegi baten biltegitze-prozesuaren ondorio— LDB formatura bihurtzen dira berehala, hiztegi-informazioaren atzipen azkar, malgu eta eroso eskaintzen duelarik. LDB horri egin beharreko eskari edota aldaketak bereziki diseinaturiko eskari-lengoiaren (LQL, Lexical Query Language) bitartez egin daitezke. Era horretan egituratutako hiztegiak analizatu eta ikertzeko zenbait teknika azaltzen dute egileek ondoren: iragazketa, definizioen gune sintaktikoaren bilaketa —normalean "genus" edo termino generikoa— matrize-eraketa (sinonimiari buruzko azterketetarako, adibidez), "sprouting" (erlazio baten gaineko itxitura trantsitiboaren kalkulua, "hazi-hitz" batetik hasita), etab. Artikuluaren azken aldera berriz, informazio sintaktiko eta semantikoaren erauzketaz, sinonimiaz eta erlazio honek planteatutako problemez, eta desanbiguazio lexikalaz dihardute. (Neff *et al.*, 88) eta (Neff & Boguraev, 89) aurreko lanarekin zerikusi estua duten lanak ditugu, non MRDtik LDBrako informazio-intsuldaketa automatizatzeaz ari bait dira egileak, LDB eredu berbera bere LQL lengoiarekin proposatuz.

---

<sup>1</sup> Ondorio horiek geroago onartuak izan ziren, New York-en egindako mintegian ("The Lexical Entry", 1986).

<sup>2</sup> New York-eko IBM T.J. Watson Research Center-en garatutako lanak dira, batik bat.

## II. kapitulua

LNP sistemetarako lexikoen eraikuntza automatikoa artikulu askotan erabilitako gaia dugu, eta horien artean —ikus baita txosten honen hurrengo atala ere— (Chodorow *et al.*, 88) eta (Wilks *et al.*, 89; 90) aipatuko ditugu. Hurbilekoak dira berriz (Grimes, 84; Chodorow *et al.*, 85), zeinetan ere hiztegi-ezagutzen lanketarako teknika eta metodologiak deskribatzen bait dira.

Europako Elkarteko ACQUILEX proiektuaren (Esprit BRA-3030: Acquisition of Lexical Knowledge for Natural Language Processing Systems, (Calzolari, 90b)) helburu nagusia MRDak erabiliz LNPko sistemetarako osagai lexikalak eraikitzekeo teknikak eta metodologiak garatzea izan da. Informazio lexikalaren erauzketan —sintaktiko-semantikoa, hainbat MRDtatik eta ingurune eleanitzean— datza proiektuaren funtsa, ezagutza lexikal eleanitzeko base (LKB, Lexical Knowledge Base; ikus § II.2) bakarria sortzea izaki helburu orokorra. LKB horrek kontzeptuen artean "nabigatzeko" aukera emango lioke erabiltzaileari, kontzeptuen eta kontzeptuarterko erlazio semantikoaren bitarteko atzipen-bideak eskainiz. ACQUILEXen ikergai nagusiak direla eta, hona hemen behinenak:

- Hiztegi-sarreretarako errepresentazio-lengoaia standard eta esplizituaren diseinua.
- Lengoaia naturalezko hiztegi-definizioetan inplizituki dagoen informazio semantiko guztiaren erauzketa.
- Kontzeptu eta kontzeptuarterko erlazioen moduan erauzitako informazio semantiko guztien errepresentazio-ereduen azterketa.
- LDB eta LKBen sorkuntza, atzipen eta tratamendurako oinarritzko programeriaren diseinua eta inplementazioa.

Literaturan hainbat aldiz aipaturiko "lengoaia naturalaren arloko zintzurgune estua" den problema aztertzea dute helburu ACQUILEXen, hots, neurri errealeko lexiko konputazional "handiak" erabilera desberdinetarako prestatzea; lexiko horietan informazio semantikoak esplizitu eta atzigarri izan behar duela azpimarratzen dute.

Aipatu berri diren lan guztietan, MRDetatik jakintza lexiko-semantikoa eskuratzearen alderdia hor ageri da nola edo hala. Gai horretaz arituko gara geu ere ondoko lerrootan.

Ikuspuntu teorikoago batetik, euskarri magnetikoan dauden hiztegien azterketaren helburutzat hizkuntzaren egitura semantikoari buruzko ikerketa egitea dukete beste egile batzuk (Wilks *et al.*, 90).

*Computational Lexicography for Natural Language Processing* (Boguraev & Briscoe arg., 89a) liburuan, LNParen aldetik egindako lexikografia konputazionalaren ikerrarloaren

ikuspegi orokorra ematen da<sup>1</sup>. Gure proiektu honetan egindako ikerketak badu harremanik liburu horretako kapituluetan zehar azaldutako zenbait lanekin, eta hortaz, txosten honetako hainbat lekutan beharrezko aipuak egingo zaizkio.

## II.2 MRDetako ezagutzaren eskuratzea.

LNPrako osagai lexikalen eratze lanetan bi modutara egin izan da lan: bi modu hauek G.A. Miller-ek (1985) "demo" eta "livre" deitutako jokabideak dira. "Demo" izeneko jokabidea informatikoki tratagarria baina sarrera kopuruaren aldetik oso mugatua den lexiko aberatsa eratzean datza. Jokabide horixe izan da nagusi LNPan —eta Adimean Artifizialean, oro har— azkeneko bi hamarraldiotan (Wilks *et al.*, 90). "Livre" delakoan berriz, oso bestelako bide bat egin ohi da, hiztegi arrunt baten makina-bertsioa hartu eta hizkuntz tratamenduko sistemaren beharretara egokitzen saiatuz. "Demo" jokabidearen arazoa sarrera bakoitza egiteko arrazoia espezifikoezia izatetik dator; hori dela medio, aurrakusi gabeko hizkuntz fenomenoei aurre egiterakoan sistemak porrot egiten du, beharrezko den ezagutza falta du eta. Beste jokabidean berriz, ezagutzaren iturritzat harturikoak, hau da hiztegiak eta entziklopediak, giza erabilerarako pentsatutako tresnak izatetik datoz problema gehienak.

"Livre" ihardupideak piztutako interesa gero eta handiago bihurtu da, 80.eko hamarkadan batez ere<sup>2</sup>. R. A. Amsler-ek eta (Amsler & White, 79; Amsler, 80, 81) ereindako haziaren ondoren, *The Merriam-Webster New Pocket Dictionary* (W7) (Chodorow *et al.*, 85; Markowitz *et al.*, 86; Binot & Jensen, 87; Jensen & Binot, 87) edo *LONGMAN Dictionary of Contemporary English* (LDOCE) (Michiels & Noël, 82a; Michiels *et al.*, 82b; Boguraev & Briscoe, 87; Jensen & Binot, 88; Boguraev & Briscoe eds., 89a; Wilks *et al.*, 90) bezalako hiztegi inportanteagoen gainean egindako lanak etorri ziren.

Lan horietan, hiztegia oso izaera bereziko *testu-corpus*zart hartzen da (van den Hurk & Meijs, 86; Boguraev *et al.*, 89b)<sup>3</sup>. Berezitasun hori corpus horretatik atera daitekeen informazio lexikalaren nolakotasunari zor zaio. Corpus orokor batetik hizkuntz elementuei buruzko informazio asko eta asko ateratzea baldin badago ere —eta testu-corpusen azterketa

---

<sup>1</sup> LDOCEren gainean egindako lanak besterik aurkitzen ez badira ere, deskribatutako teknikak nahiz lan desberdinetan ateratako ondorioak interes orokorreko zat har genitzake.

<sup>2</sup> Aurretiko zenbait lan aipatzeke ere ezin utzi, hala nola liburu gisa (Sparck-Jones, 86) argitaratutako K. Sparck-Jones-en tesia (1964) edo K.C. Litkowsky-ren lanak (1978; 80), garai hartako teknologia informatikoaren egoera eta hiztegiak euskari magnetikoan edukitzeko aukera urria zirela eta, aurrerago egiteko hamaika oztopo aurkitu zutenak.

<sup>3</sup> Lehen erreferentziaren izenburua (van den Hurk & Meijs, 86) aski adierazgarria dugu: "The Dictionary as a Corpus: Analyzing LDOCE's Definition Language". (Boguraev *et al.*, 89)n lexikoa eskuratzeko bi iturburu aipatzen dira, testu-korpus orokorrak eta MRDak; artikulu horretan, egileek metodologia bat finkatzen dute hiztegi-iturriok arakatzeko, hitzen ezaugarri semantikoak eskuratzearren, azken helburua lexiko konputazional baten alderdi semantikoa osatzea izaki.

## II. kapitulua

dugu, hain zuzen, hizkuntzalaritza informatikoaren barruan puri-purian dagoen arloetariko bat—, hiztegi batetik, eta bereziki hiztegi bateko definizioen analisitik, zuzenean atera daitekeen informazioa<sup>1</sup> nekez aterako genuke corpus orokorrago batetik.

Corpusetik zein MRDetatik jakintza lexikala ahal den neurrian automatikoki eskuratu nahiak dituen arazoei buruz dihardute, besteak beste, (Boguraev, 87; Boguraev & Briscoe eds., 89a; Calzolari, 90b) lanetan.

Baina hiztegia testu-corpus hutsa baino askoz ere gehiago da, hitzen esanahiari dagokionean. Hiztegia *oso entitate egituratua* dugu, jakintza orokor nahiz lexikalaren iturri estimagaitza (Calzolari, 86; 90a).

Gorago aipatu lanen artean, bi sail handi bereizi behar ditugu:

- MRDetatik bereziki —eta eskala handian— sintaxiari dagokion informazioa<sup>2</sup> eskuratzeko metodoen garapenera zuzendutako lanak (Boguraev & Briscoe, 87, esate baterako).
- Hiztegietan gorderik dagoen egitura semantikoari buruzko lanak (Amsler & White, 79; Amsler, 80, 81; Chodorow *et al.*, 85; Markowitz *et al.*, 86; Binot & Jensen, 87; Jensen & Binot, 87; Alshawi, 87, 89; Wilks *et al.*, 89, 90).

Lan horien emaitzatzat har genezake Hiztegi Ezagutza-Basearen kontzeptua. Hona Hiztegi Ezagutza-Basearen (LKB, Lexical Knowledge Base) definizioa: “. . . a repository of computational information about concepts intended to be generally useful in many application areas including computational linguistics, artificial intelligence, and information science.” (Amsler, 84:458). Amsler-entzat, Hiztegi Ezagutza-Basea "biltegi" orokorra litzateke —LNPko programa jakin batzuentzat baliagarri litzatekeen tresna berezia baino areago—, lexiko-kontzeptuei eta berorien erlazioei buruzko informazioa jasoko lukeena. Bere hitzaldiaren laburpenean aipatzen du definizio bildumak analizatu eta aztertzetik erator daitekeen ezagutza inplizituaren kontua. Horrela lortutako ezagutzaren gainean ezar daitezkeen inferentzi mekanismoak ere aipatzen ditu laburpen horren bukaera aldean, baserako ezagutza berrien eskuratzeari diharduenean: ezagutza berri hori eskuratzeari beharrezkoa izango da ezagutza-basean dagoen materialetik eratorri ezina dela frogatzen baldin bada.

---

<sup>1</sup> Nahiz eta, Calzolari-k (1990b) dioen bezala, informazio-eskuratze hori ez den behin ere osoa izango, eta ezta ere hutsik gabekoa edo erabat automatikoa.

<sup>2</sup> Lan horien lehenbiziko helburua MRDetako —argitaletxeetan grabatutako banda magnetikoak, paperezko hiztegiaren iturburu izateko jarioak— sarrerak analizatu eta bertako informazioa egituratuz LDBak (Lexical Data Base) (Neff *et al.*, 88) eratzea izan ohi da normalean.

Guri interesatu zaiguna —hizkuntzalaritza informatikoan zein adimen artifizialean premiazkoa denaren osagarri, bestalde (Wilks *et al.*, 90)— MRDan dagoen ezagutza esplizitu eta "gordearen" azterketa izan da. Beraz, gure xedea MRDtik informazio hori —sintaktiko, semantiko edo bestelakoa—*erauzi* eta erabilera desberdinetarako egokitu eta *egituratzea* izan da. Egituratze horren emaitzari esaten zaio hainbat lanetan Hiztegi Ezagutza-Basea (LKB) —(Amsler, 84; Calzolari, 86, 90a), esaterako— edo Machine Tractable Dictionary (MTD) (Wilks *et al.*, 89; 90). MRDetatik LKBetaranzko bidean —sarri LDB<sup>1</sup> erako bitarteko formatatik paseaz (Zampolli, 87; Calzolari, 90b)— bi alderdi eduki behar izaten dira kontuan (Wilks *et al.*, 90):

- 1) MRDan dagoen jakintza edo ezagutza zer-nolakoa den aztertu.
- 2) Ezagutza horren errepresentaziorik egokiena aukeratu.

Lehenari dagokionez, proiektu honetan hiztegi bat aztertu eta analizatu egin dugu helburu horrekin; bigarrena dela eta berriz, hiztegi-ezagutzaren errepresentaziorako proposamen bat egiten dugu.

Ondokoan —ondorengo bi ataletan bereziki— hiztegiko definizioez arituko gara, definiziook kontzeptuen eduki semantikoaren gordailutzat (Litkowsky, 78; 80) eta, beraz, eduki horiek eskuratzeko iturburu potentzialtzat hartuz. Definizioen egiturari buruz zenbait autoreren iritzia errepasatu, eta LPPLko definizioen gainean egindako ikerketa eta lortutako emaitzak azalduko ditugu.

## II.2.1 Definizio-esaldien egitura.

Hiztegiko definizio-testuen egitura aztertzen hasi bezain laster gauza batez ohartuko gara: badira —definizio batzuetan bederen— definizio-esaldi horiek idazteko modu finko samarrak. Modu horien artean, adibide bat ematearren, denok ezagutzen dugun *relatif à* hitzekin hasitako adjetibo-definizioak, adjetiboa dagokion izenarekin harreman zuzenean jartzen dutenak. Aristotelerengandik ikasia dugu bestalde, zein den kontzeptu bat definitzeko era "egokia", hau da, bere "genus" eman —honela, bere mota edo klasean sailkatuz— eta bere motako diren beste kontzeptuengandik bereizteko "differentia specifica" erantsiz.

---

<sup>1</sup> Calzolarien arabera (1990b), MRDa normalean fitxategi sekuentziala izaten da —fotokonposizio-bandak, esaterako—; LDBak aldiz, informazio-atzipena erraztuko duen halako egituraketa bat badu, horretarako eskaintzen dituen prozedurak gutxi edo asko interaktiboak direlarik. LKBan berriz, kontzeptu-kategoriak eta kontzeptuarterko erlazio lexiko-semantiko mota desberdinak daude errepresentaturik, are atzibide aberatsagoa eskainiz. Inferentzia eta dedukzioaren nozioak ere ageri dira LKB deitutako horretan, eta, egilearen arabera, gehienetan kontzeptu-sare gisa antolatuta egon ohi dira.

## II. kapitulua

Lexikografoen lanetan ageri da gure lanaren lehen parte honetarako funtsezkoak diren *formula* edo *definizio-eskema* horien berri. Esaterako, M. Moliner-ek (1983), bere *Diccionario de uso del español* hiztegiaren hitzaurrean, sarrerak definitzeko definizio-formula eta -arau multzo oso bat darabilela dio. Autore horren hitzetan, bere definizio-teoriari jarraiki —"genus et differentia" motako formula klasikoan oinarritua batik bat— kontzeptuen hierarkia (etengabe goranzkoa) eratzea posible da, banako edo indibiduoetatik hasi eta mota goreneraino. Hitzak definitzeko jokabide zorrotz bera duten beste lexikografoen artean Sinclair (1987) aipa daiteke; definizioak bete behar lituzkeen baldintzez ari dela, Sinclair-ek, argitasunaren eta zehaztasunaren izenean, definizio batek "genus" bat eta definituriko hitza —edo "definiendum" a— bere kuasi-sinonimo guztietarik bereizteko adina "differentia" behar dituela dio. Gaurregungo eskola lexikografiko hauen beste arau bat sinonimia definizio-formula gisa erabat baztertzea litzateke, definizioetako gurpil zoro eta tautologiekin batera (Moliner, 83).

S.I. Landau-ren arabera (1984), lexikografoak kontzeptu bat definitzerakoan duen erroka nagusia zehaztasun espezifikoarene eta erreferente-domeinuaren arteko oreka bilatzean datza. Definizio zehatzegiak hainbat eta hainbat erreferente bere domeinutik at uzteko arriskua du. Aitzitik, definizio orokorregiak —hitzaren ia erreferente posible guztiak biltzen dituenak— "zehaztasun falta" izango du beti.

Lexikografia konputazionalaren arloan, ikerlari askok aztertu ditu definizioak egitura aldetik. Hiztegian hitzak beste hitz batzuk usatuz —horiek ere hiztegikoak, jeneralean— definitzen direlarik, definizioen *zirkulartasuna* hortxe dagoen zerbait da (Smith & Maxwell, 80). J. Olney-ren esanetan —(Smith & Maxwell, 80)n aipatua— hiztegiko hitzen %50a ageri da hiztegi bereko definizio-testuetan. Definizioen zirkulartasunaren kontu hori hainbat lanetan aipatzen da, eta horien artean, ikerlanaren jomuga, hiztegietako definizio klasikoak ustiatuz, egitura hierarkikoak —hiperonimia eta hiponimia erlazioak— eraiki eta aztertzea denetan (Vossen, 89).

Definizioen beste alde interesgarri bat definitutako hitzen esanahia *esanahi bakunagoen osaketa* bezala ikustean datza. Horri buruz, P.N. Johnson Laird-ek (1981:107) ondokoa dio:

A dictionary definition is made up of a **structured set of semantic elements** which decompose the meaning of the word into its semantic constituents.

Hitzen *hedadura* eta *besarkadurari* dagokionean berriz, *Dictionnaire de Linguistique* (Dubois *et al.*, 73:427) hiztegian argi eta garbi bereizten dira semantikaren eremua eta erreferenteena:

En termes de logique, on peut dire que la définition du mot "chaise" en compréhension intéresse la sémantique, tandis que la définition du mot "chaise" en extension ("A, B, C, N sont des chaises") intéresse une théorie de la référence.

(Wilks *et al.*, 89)n, beren MTDko hitz-adiera bakoitza errepresentatzeko sortutako frame-egiturez ari direla, frame horiek errepresentatutako kontzeptuaren alderdi intentsionalari dagozkiola diote.

Amsler-entzat (1984:458), hiztegia giza jakintzaren indize moduko bat litzateke. Definizio-esaldiak berriz —LKB handiak eraikitze lanaz ari delarik, LNPko sistemetan erabiliak izateko batik bat—, kontzeptuen esanahiaren errepresentazio gisa erabiltzeko ez-nahikotzat jotzen ditu:

A dictionary definition provides the minimal information necessary to evoke the concept it defines in the mind of a human reader who already knows to what this concept refers. It is neither intended nor capable of serving as the actual "meaning" of that concept.

(Vossen, 89)n, hiztegi arrunt bateko definizio edo "esanahi-deskribapenei" buruz diharduela, definizio-sistemak dituen hiru ezaugarri aipatzen ditu egileak:

- *Erredundantzia*: definizio-sisteman oso era naturalean adierazten da zenbait informazioen erredundantzia —hiponimiaren trantsitibotasunari esker erator daitekeen informazioa, esaterako—; beste zenbait ordea, sarrera konkretu bati dagokion informazio espezifiko da.
- *Atzigarritasuna*, definizioetako hitzak beren definizioen erreferentzia baitira.
- *Erlazioak*, sarrera-hitzen eta beren definizioetako hitzen artean baino, sarrera-hitzen eta beren definizioetako hitzekin eratutako hizkuntz espresioen artean. Horrek esan nahi du definizioetako hitzak ez ezik hitz horiek definizio-espresioan betetzen duten funtzioa ere kontuan hartu beharrekoa dela, inferentzia sistematikoak egingo baldin badira behintzat.

Azkenik, hiztegi bateko definizioen hizkuntzaren ezaugarririk behinenetakoa bere *metalengoaia* izaeran<sup>1</sup> letzake. Lan desberdinetan aipatutako ezaugarri horrek, definizioak idazteko orduan lexikografoek erabilitako formulen kontuarekin erabat loturik dagoena, aparteko interesa izan du guretzat hiztegi-egituretan dauden erlazio inplizituen bilaketarako. Hara zer dioen G. Ferrari (1980:259) batek, adibidez:

La définition que le dictionnaire attache à chaque entrée, d'un côté a la fonction de nous faire comprendre le signifié d'un mot, de l'autre représente effectivement les traits contextuels que nous cherchons. Elle appartient, donc, au **langage**, puisqu'elle nous renseigne sur une certaine

---

<sup>1</sup> “La **métalangue** est une langue artificielle servant à décrire une langue naturelle (1) dont les termes sont ceux de la langue objet d'analyse, mais qui ont une seule acception et (2) dont les règles de syntaxe sont aussi celles de la langue analysée. La métalangue est, par exemple, le **langage grammatical**, dont le linguiste se sert pour décrire le fonctionnement de la langue; c'est le **langage lexicographique**, dont l'auteur de dictionnaire se sert pour les définitions des mots.” (Dubois et al., 73:317).

## II. kapitulua

réalité, parfois concrète, parfois linguistique, et à un **métalangage**, puisque chaque élément, ou certains éléments des plus significatifs, peut être considéré comme le symbole de certain trait.

Frantsesezko *beauté* hitzaren definizioa —*qualité de ce qui est beau*— komentagai hartuz, Ferrarik dio definizio horrek edertasuna zer den ez ezik edertasuna kalitateen klaseko zerbait dela ere adierazten digula, eta beraz, klase horrentzat definitutako zenbait konpatibilitate izango dituela.

Aurrerago, hiztegietaiko definizioen sailkapen bat ematen digu. Bere iritziz, sarrera bat definitzeko hiru modu daude oro har (Smith eta Maxwell-ek (1980) emandako sailkapenaren antzeko):

- a) sinonimo baten bitartez;
- b) deskribapen bat erabiliz (definizioa);
- c) oinarri-sarrera batekiko erlazio sintaktiko baten bitartez.

Definizio-lengoiaren metalengoitutasuna dela eta berriz, Calzolarik (1986; 90a) hiztegioko definizio-esaldiak nolabait kontrolaturik daudela azpimarratzen du, kontrol hori konszienteki egiten ez bada ere: tradizioa eta konbentzio lexikografikoak lirateke definizioak erabat libre ez idazteko dagoen arrazoirik nagusienetako bat. Hori horrela, hortxe dugu kontzeptuarterko erlazio semantiko desberdinak erauzteko orduan hain inportantea datekeen definizio-formulen multzoa (Zampolli, 87).

Derragun, atal honekin bukatzeko, lexikografia konputazionalen definizioei buruz egindako lanen artean mota desberdinak bereiz daitezkeela: definizioen azterketa maila grafemiko batean geratzen den lanetatik hasi eta jomugatzat definizio-esaldien analisi lexiko-semantikoa burutzea dutenataraino.

Lehenengo multzoan, hau da, MRD bateko sarreraren *analisi grafemiko*<sup>1</sup> baten bitartez LDBa eratzeko helburua duten proiektuen artean —proiektu askotan lehen fasea izan ohi dena, bestalde— (Neff *et al.*, 88; Neff & Boguraev, 89; Martí & Castellón, 90) aipa litezke.

Beste multzo handi bat sarreraren *genusaren erauzketa automatiko edo erdiautomatiko* helburu duten lanek osatzen dute; helburu horixe da nagusia zenbait proiektutan (Yoshida *et al.*, 82; Calzolari, 84b; Chodorow *et al.*, 85; Tsurumaru *et al.*, 86; Chodorow *et al.*, 88), eta autore gehienek iritziz, definiendumaren genusa erauzteko jeneralean diferentzi partearen

---

<sup>1</sup> Maila grafemikoko analisi hori burutzeko sarreraren gramatika bat zehaztu eta gramatika horrentzako analizatzaile bat egin behar izaten da; analizatzaile horrek sarreraren atal desberdinak bereizteko gauza izan behar du: definitutako hitza, kategoria sintaktikoa, errepresentazio fonetikoa, erabilere-markak, definizioa, adibideak, etab.



analisi eta ustiaketa<sup>1</sup> baino errazago gertatu ohi da. Lan horietan, sarrera baten definizio-testuan elementu generikoa dena —genus, hiperonimo, gainklase, etab.— ediretea da helburua, berorren eta definitutako hitzaren arteko lotura hierarkiko bat ezartzearen.

Baina hiperonimo-erazketa baino harantzago joatearen premia ere, hots, definizioen benetako analisi osoa egitearena, aitortzen da literaturan (Michiels & Noël, 82; Wilks *et al.*, 89, 90). Calzolarik (1983), "differentia" eta alde "erlazonalen" —hau da, definizio-esaldietan gai generikoa edo genusa eta espezifikoa edo "differentia" harremanetan jartzen dituzten eraikuntzen— lanketa automatikoak eskaintzen duen interesa eta zailtasuna azpimarratzen ditu. Orduan, *definizio-esaldien analisi lexiko-semantikoa* bihurtzen da xede.

Definizioen analisiaren helburuen artean, Litkowsky-k (1978) hitz baten esanahiaren errepresentazio eraginkorrago bat aurkitzea aipatzen du, halako analisiak edozein hiztegitan aurki daitezkeen informazio erredundante kopuru eskergak eta gurpil zoroak deuseztatzeko duen interesaz aparte.

Definizioak analizatu eta bertatik informazio semantikoa ateratzeko helburuarekin egindako lanak aipatzen hasita, hor ditugu S. Yoshida-k eta (Yoshida *et al.*, 82) edo G. Arangok (1983) egindakoak, zeintzuk, aldeak alde, definizio-esaldiak analizatu eta, horrela, esaldi horiei —eta beraz, definituriko kontzeptuei— lotutako errepresentazio semantikoa lortzea baitute xede. Horrek aukera ematen du hiztegiko kontzeptuen artean lotura-sistema bat ezartzeko, hiztegian bertan dauden erlazio taxonomikoei —baina ez bakarrik horiei— esker batik bat.

Ahlswede-ren taldearen lanetan —kontzeptuen eta kontzeptuarteko erlazioen errepresentazio batera iristeko asmoz, horiek ere—, bilakabide interesgarria antzeman daiteke W7 hiztegiko definizioetako erlazio lexiko-semantikoaren erazketarako erabilitako prozeduretan: definizio-esaldiak analizatzeko ingeleserako analizatzaile sintaktiko orokor baten<sup>2</sup> erabileratik hasi (Ahlswede, 85; Ahlswede *et al.*, 86), eta analizatzaile txikiz osatutako multzo baten bitartezko analisi erdiautomatikoraino ("dumb parsing", (Ahlswede, 89)), azken kasu horretan informazio interesgarri baina partzialen erazketari begira egindako teknikak garatu direlarik.

---

<sup>1</sup> (Wilks *et al.*, 89)n diotenez, genusaren erazketa aditzen definizioetan errazago da izenenetan baino; alabaina, izenen definizioak egokiago omen dira "differentia"k ateratzeko orduan.

<sup>2</sup> Linguistic String Parser (LSP), (Sager, 81).

## II.2.2 Definizioetan gordetako kontzeptuarterko erlazioak.

Hiztegi arruntetako definizioetatik atera daitezkeen erlazio lexiko-semanticoen desberdinen ikuspegi orokorra eman nahi genuke azpiatal honetan, horretarako gorago aipaturiko lanetan oinarrituz.

Hiztegiaren eduki semanticoa, definizioetan dagoena (Litkowsky, 80), elkarrekin erlazonaturik dauden kontzeptuen —motak, kategoriak (Calzolari & Picchi, 88)— multzo bezala hartu ohi da. Kontzeptuak hiztegian definiturik dauden hitz-adiera desberdinak lirakeke eta kontzeptuarterko erlazioak definizio-esaldien bitartez ezarriko lirakeke. Erlazio horiek era askotakoak dira.

Erlazio horien sailkapen desberdinak egin ohi dira, autoreen arabera. Tipikoenetako bat erlazioak bi sailetan, *paradigmatikoak* eta *sintagmatikoak* (Calzolari, 84b; van den Hurk & Meijs, 86), banatzen dituena litzateke. Erlazio paradigmatikoak kategoriatik bereko hitzen artean ezarritakoak lirakeke, eta sintagmatikoak kategoriatik bereko hitzen artean daudenak (van den Hurk & Meijs, 86). Azken erreferentzia horretan, definizioen bitartez ezarritako erlazioak sailkatzeko erlazio *bertikal* eta *horizontalen* kontzeptuak ere badarabilzate. Erlazio bertikala definiendumaren eta bere genusaren artekoa litzateke —hiperonimia erlazioa—, eta horizontala berriz, genusaren eta differentia osatzen duten aurre- edo post-modifikatzaileen artean ezarritakoa.

*Hiperonimia* eta bere alderantzikoa den *hiponimia* dira, dudarik gabe, definizioetan aurki daitezkeen erlazioen artean hobekien landu direnak. R.A. Amsler (1980; 81) oso aipatua da hiztegian erlazio taxonomikoei esker ezarritako *egitura hierarkikoaz* —definizio-modu tradizional aristotelikoaz, batez ere— iharduten denean. Berari zor zaio egitura hori deskribatzeko sortutako hierarkia mihiztatuaren kontzeptua ("tangled hierarchy"). Kontzeptu horren inguruan, eta bestelako erlazio lexikalei esker ezarritako egitura hierarkikoetara hedatuz, (Calzolari, 83:48)n ondokoa irakur daiteke, hizkuntzaren lexiko-sistemari buruz:

. . . not organized as a unique hierarchical structure, and not even as several independent hierarchies . . . as the composition and overlapping of many tangled hierarchies organized on many different types of relations (at a paradigmatic level of substitutability).

Autore asko dator bat hierarkia hiponimikoak hierarkia dexente zabal baina sakonera gutxikoak direla esaterakoan (Chodorow *et al.*, 88). Wilks-en taldearen lanetan (Wilks *et al.*, 89; 90), hierarkien mihiztatze hori hierarkiak hitzekin eraikitzeagatik datorrela diote: autore horien arabera, mihiztadura ia erabat desagertuko litzateke hierarkiok kontzeptuekin —hitz-adierak edo zentzuak— sortuz gero, horretarako definizioetan ageri diren hitzak aldeztatik lexikalki desanbiguatzea derrigorrezko izanik.

Calzolariren (1984b) arabera, hiponimia erlazioak sortzen dituen arbola hierarkikoek semantikoki koherente diren azpimultzo zenbaitetan banatzen du lexikoa.

Kontzeptu orokorragoetatik espezifikagoetara legokeen *atributu-herentzia* ere hainbat lanetan jorratutako gaia dugu (Chodorow *et al.*, 85; Calzolari, 86,90a; Calzolari & Picchi, 88; Miller *et al.*, 88).

Zenbait autoreren arabera, adjetiboen arteko erlazio hiponimikoak ez dira oso argiak (Ahlswede, 85), izan ere, adjetiboen definizioa emateko, askotan jotzen baita beste kategoria sintaktiko bateko hitz bat erabiltzera, definitutako adjetiboa harekin lotuz. (Miller *et al.*, 88)n dudatan jartzen dute erlazio horien ezarrera aditzen artean ere, bertan dauden herentzi arazoak direla medio kasu honetan.

Teubert-ek (1984) kontzeptu multzoaren eta erlazioen "sare" ikuspegia ematen digu, partez hierarkiko eta partez *asoziatibo* den erlazio kontzeptualen sare gisa aurkeztuz.

Kontzeptuen artean hierarkiarik ezartzen ez duten erlazioen artean berriz, bistakoa da *sinonimia* dela aipatu beharrekoa, hiztegietan betidanik definizio-formula gisa usatutakoa eta, oraindik orain, (Miller *et al.*, 88)<sup>1</sup> bezalako lexikografia konputazionalak lanetan ere erabilia. Sinonimia eta bere arazoak —erlazio honen trantsitibotasunak aurkeztutakoak bereziki— landu dira (Warnesson & Marcotorchino, 82; Calzolari, 83; Warnesson, 84; Sparck-Jones, 86) bezalako lanetan. Sinonimiarekin batera *antonimia* ere aurkitu ohi da, batez ere adjetiboak definitzeko erabilia. Kasu honetan, arazoa antonimia kontzeptuarteko edo hitzen arteko erlazioa (Miller *et al.*, 88) den jakitean letzake.

Erlazio *meronimikoek* —edo ezagunago ditugunez, "partea-oso" erlazioek— osatzen dute hiztegian hierarki eratzaile diren beste erlazio multzo handia. Izan ere, meronimiaren kontzeptua ez baita "partea-oso" erlazioa mugatzen, baizik eta "osagaia-objetua", "elementua-multzoa", "gaia-objetua", "lekua-ingurua", etab. ere hartzen ditu bere baitan (Winston *et al.*, 87). Calzolariren artikulua batzuetan (1983; 88), erlazio horiek hiztegiko definizioetatik lortzeari begira egindako lanak aipatzen dira. (Miller *et al.*, 88)n berriz, hierarkia meronimikoetako herentziaren arazoa tratatzen da. Horri buruz oso interesgarria da (Winston *et al.*, 87) artikulua, non erlazio horien trantsitibotasunaren arazoa aztertzen baita, semantika lexikalaren ikuspuntutik. Aipatu artikuluan, erlazio meronimiko desberdinak zertan diren azaltzeaz gain, beren esistentzia egiaztatzeko (ingelesez) zenbait test aurkezten dira, erlazio horiek erlazio lexiko-semantikoaren munduan kokatuz. Aipatu berri ditugun

---

<sup>1</sup> (Miller *et al.*, 88)n WordNet izeneko proiektua deskribatzen da, teoria psikolinguistikoen arabera sortutako ingelesaren erreferentzia lexikaleko sistema gisa. Kontzeptu bakoitza adierazteko sinonimo multzo bat erabiltzen da.

## II. kapitulua

trantsitibotasun arazoentzako soluziobide bat eskain lezakete hierarkia mistoek, hau da, erlazio meronimiko eta bestelakoen bitartez eratutako hierarkiek.

*Eratorpena* ere analizagarria da hiztegi-tako definizioetan. Zerikusi morfologikoa duten hitzen arteko erlazioak —aditzetiko izen eta adjetiboak, izenetiko adjetiboak, etab. (Calzolari, 83)— erauz daitezke zenbait definiziotatik, horretara beren arteko erlazio semantikoa esplizitu bihurtuz (Calzolari, 86; 90a).

Aditzen definiziotatik bereziki atera daitezke *kasu-* edo *tema-erlazioak*<sup>1</sup>. Aditz batek denotatutako egintzaren agente "tipiko" edo "prototipikoa" (Ahlsvede *et al.*, 86) —"generic agent" (Markowitz *et al.*, 86)n— aise erauz daiteke "*ce*" "*qui*" *forme-verbale* tankerako eskema duten definiziotatik. Halaber, zenbait definizioak ematen digu aukera egintza baten objektu tipikoa lortzeko edo bestelako erlazio tematikoak erauzteko. Definizioetan bila ditzakegu autore desberdinek argumentu lexikalizatuak (Calzolari, 86, 90a) edo "preferred case fillers" (Wilks *et al.*, 89) deritzetenak, askotan, baina ez derrigor, azaleko mailako eratorpen morfologikoko fenomeno baten pean gordeak.

*Hautapen-murriztapenei* ("selectional restrictions") buruzko informazioa ere lor daiteke definizio-esaldietatik, (Litkowsky, 80)n edo (Markowitz *et al.*, 86; Calzolari, 86, 90)en ikus daitekeenez, azken kasu horietan adjetiboaren definizioez ari direlarik. Halaber, *kokamendu-erlazioak* eta *agerkidetzari* buruzko murriztapenak ere aurki daitezke hiztegi-tako definizioetan (Calzolari, 83).

Hiztegi espezifikotatik erauzgarri diren erlazioez diharduela, Ahlsvede-k (1985) *azpilengoia jakin bateko erlazio espezifikoa*<sup>2</sup> aipatzen ditu, hizkuntzaren lexiko orokorrari ez dagozkion erlazioak alegia. (Ahlsvede *et al.*, 86)n berriz, oraingoan hizkuntzaren hiztegi orokorraz, CHILDBEZALAKO erlazio bereziak—*lamb* eta *sheepen* arteko erlazioa, esate baterako—, taxonomia biologikoari dagozkionak, etab. erauz daitezkeela erakusten dute. (Wilks *et al.*, 90)en, lehen aipatutako erlazio taxonomiko eta meronimikoak alde batera utziz, definizio-tako diferentzi partetik lortutako erlazio berezi zenbait azaltzen dute, "*means of*", "*measure of*", eta abarren tankerako eraikuntzetatik erauziak. Azkenik, (Calzolari, 83)n lengoia berezituaren azpisistema terminologikoa erauzteko aukerak seinatzen dira.

---

<sup>1</sup> Azpimarratzekoa da hiztegi-eguztuen iturririk aberatsenetako bat adibideetan datzala. Bertan aurki daitezkeen informazioen artean, kasu-erlazioen argumentuena litzateke ohargarrienetakoa (Wilks *et al.*, 89). Gurean definizioen azterketara mugatu izana da, hemen, ezagutza-iturri oparo horretaz gehiago ez aritzeko arrazoia.

<sup>2</sup> Apoplejiaren inguruko kasuen medikuntza-azpilengoian aurki daitezkeen erlazio espezifikoei dagokienez, Ahlsvede-k (1985) NNABLE aipatzen du —funtzio fisiko edo mental baten akatsei buruz ihardutean, baldintzaren eta funtzioaren arteko erlazioak, hala nola *aphasia* eta *speech*, edo *amnesia* eta *memory*ren artekoak— eta FUNC —giza-gorputzaren atal baten funtzio tipikoaren deskribapena edo litzatekeena—.

Txosten honetako II.4.6 azpiatalean laburbilduko ditugu gure proiektuan ikergai izan dugun hiztegitik ateratako erlazioak.

### **II.3 Le Plus Petit Larousse (LPPL): hiztegiaren ezaugarriak. Hiztegi Datu-Basea.**

Proiektuaren oinarri gisa erabili dugun paperezko hiztegia aurkeztuko dugu atal honetan. Hiztegi horretaz baliatuz egin dugu zenbait azterketa enpiriko, definizioen gainean batez ere, eta berori hartu dugu sistemak eskuratu beharreko hiztegi-ezagutzaren iturburutzat.

Ondoren —§ II.3.1— LPPLko informazioa eta egindako azterketen emaitzak gordetzeko nahiz azterketa enpiriko horietarako abiapuntu gisa erabili dugun Hiztegi Datu-Basearen (HDB) sorkuntza emango dugu ezagutzera. § II.4en berriz, definizioen analisi sintaktiko-semanticoa (§ II.5) egitera eramán gaituen hiztegi-definizioen azterketarako metodoa zertan den azalduko dugu. Bukatzeko, § II.6n HEBaren eraikibidea —HDBeko informazio analizatu eta birmoldatutik abiatuz— nolakoa izan den deskribatu eta § II.7n analisi-prozedura horien mugak komentatuko ditugu.

*Le Plus Petit Larousse* (Paris: Librairie Larousse, 1980) frantsesezko hiztegitxo bat da. Hiztegi esplikatzaile labur honek ondoko informazio hau dakar sarrera edo artikulu bakoitzeko:

- *Sarreraren ortografia*, zenbaitetan forma flexionatu batzuk ere azalduz (adjetiboen femeninoa bereziki).
- *Ebakera* (ez beti).
- *Kategoria sintaktikoa* (edo hitzaren izaera morfologikoa).
- *Erabileremua*: Fig., Mus., etab.
- Adiera desberdinen *definizioa*. Definiziook puntuen bitartez bereizirik datoz, sarrera bakoitzean. Sarri, hitz sinonimoetara bidaltzen gaituzten erreferentzia hutsak dira. Batzuetan, informazio osagarri zenbait badute (aditz irregular batzuen konjugazioa, adibidez).
- *Adibideak* (ez beti).

Hiztegi honen neurriak zehaztearren, eman ditzagun orain zenbait datu berorren edukinaz.

LPPLk 22899 hitz-adiera dauzka, 15953 sarreratan (hitz baten homografoak sarrera desberdinetan datoz LPPLn); beraz, 1.44 adiera ditugu sarrera-unitateko, batez beste. Horrez

## II. kapitulua

gainera, 1980 forma flexionatu ere ematen ditu (adjetibo eta izenen femenino edo plurala, etab.). Sarrera kopuru handi bat (11182, %70.09a) hitz monosemikoak dira, hau da, adiera bakarrekoak. Adiera bat edo bi dituzten sarrera kopurua %91.38a da. Definizioak, jeneralean, oso laburrak dira: adiera guztien %74.57ak 4 hitz edo gutxiagoko definizioak dauzka, batezbesteko luzera 3.27 hitzekoa delarik.

LPPLn bi definizio mota desberdin aurki daitezke. Alde batetik, hiztegian gehien ageri diren guk A motakoak deiturikoak: definizio tipikoak, non esaldi bat erabiltzen baita sarrera-hitzaren esanahia definitzeko. Aitzitik, B motako definizioak "adibideen bitarteko definizioak" ditugu, esan nahi baita, kasu horietan sarrera-hitza baino haren erabilera konkretu bat azaltzen duela definizio-esaldiak, erabilera hori adibide baten bitartez emanik. Esan bezala, LPPLko definizio gehienak A motakoak ditugu. Hirurehun besterik ez dira B motakoak, hau da, 300 definizio horietan definizio-esaldia aurretik emandako adibidearen parafraze bat da. LPPLko definizioetarik %14ak soilik ematen ditu erabilera-adibideak (B motako adibideak barne).

Kategoriaka, LPPLko azpimultzo garrantzitsuenak honako hauek ditugu (A motako definiziodun adierak):

- a) Izenak        13744 adiera
- b) Aditzak       5247 adiera
- c) Adjetiboak   3731 adiera
- d) Adberbioak   173 adiera

Azpimultzo horiek ez dira disjuntuak: behin baino gehiagotan kontatu behar izan diren adierak bai baitira horien artean, hala nola 'adj. et n.' edo 'adj. num. et m. inv.'<sup>1</sup> tankerako kategoriak dituzten sarrerak, izan ere azpimultzo bat baino gehiagori baitagozkio horiek. LPPLa osatzen duten gainerako adierak kategoria desberdinetakoak ditugu: konjuntzioak, aurrizkiak, preposizioak, izenordainak, etab.

### II.3.1 Hiztegi Datu-Basearen eraketa.

LPPL hiztegia euskarri magnetikoan ez geneukanez grabaketari ekin genion, biltegitzat datu-base erlazionaleko sistema arrunt bat hautatuz. Grabaketa horren emaitza dugu guk Hiztegi Datu-Basea (HDB) deiturikoa.

---

<sup>1</sup> Hiztegietan ohizko diren laburdurak erabiliz. Adibidez, 'adjectif et nom' adierazteko erabilitako 'adj. et n.', edo 'adjectif numéral et mot invariable' adierazten duen 'adj. num. et m. inv.'.

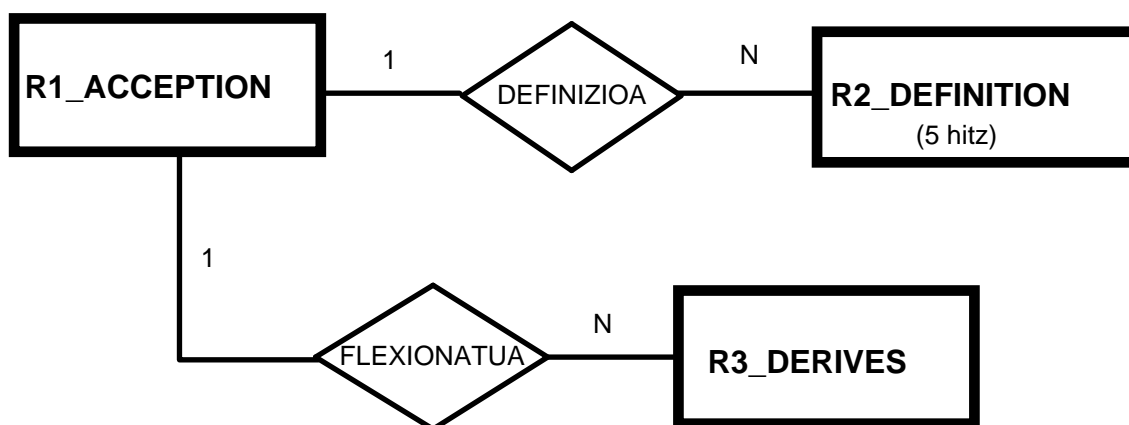
VAX ordenadorean dabilen RDB/VMS Datu-Baseen Gestiorako Sistema aukeratu genuen, horixe baitzen eskueran genuena proiektu honen hastapenetan. Zein informazio gorde eta zein ez zehaztu eta gero, eta jakina, datu-basearen egitura definiturik, grabaketari ekin zitzaion horretarako bereziki idatzitako programa baten laguntzarekin.

HDBaren egitura definitzeko aukera horrek baditu bere abantailak MRDeKiko, izan ere MRDa ez baita gehienetan izaten argitaratzailearenean hiztegiaren paperezko bertsioaren iturri izateko grabatutako banda magnetiko soila baizik. Horrek dakarren maila grafemikoko azterketa —karaktere-mota desberdinen bitartez hiztegi artikuluaren atalak ezagutu beharra, ikur berezien ezagutza, etab.— lan neketsua eta aspergarria izan ohi da maiz.

HDBrako sortutako lehen eskema kontzeptualean hiru entitate nagusi (taulak, datu-basean) pentsatu genituen:

- R1\_ACCEPTION. Bertan adiera bati dagozkion informazio eta ezaugarriak gordeko ziren: adiera bakoitza unibokoki bereizteko gakoa (sarrera, homografo-adierazlea, adiera-zenbakia), kategoria sintaktikoa, erabileremua, definizio mota eta adibidea.
- R2\_DEFINITION. Adieraren definizioa osatzen duten hitzak gordetzeko (5 hitz lerroko —eremu bakoitzean hitz bat—, LPPLko lagin baten gainean eginiko azterketatxo estatistiko batek LPPLko definizio-esaldien %93ak 4 hitz edo gutxiago daukala erakutsi baitzuen): adiera-gakoa eta definizio-esaldiko bost hitz.
- R3\_DERIVES. Taula lagungarri hau sarreren forma flexionatuei buruzko informazioa —adjetiboen femeninoa, batik bat— gordetzeko erabiliko zen: forma flexionatua, berorren forma kanonikoaren sarrera (gakoa) eta izaera morfologikoa.

Hona, irudi bidez, HDBaren hasierako eskema kontzeptuala:



II.2 irudia.- HDBaren eskema kontzeptuala.

## II. kapitulua

LPPLn egonik HDBan jaso ez dugun informazioa dela eta berriz, esan beharra dugu LPPLren gure makina-bertsio honi LPPLko sarrera bakan batzuetan dagoen informazio berezia besterik ez zaiola falta. Honela, HDBan ez dira jaso ez informazio fonetikoak (ebakera) eta ezta aditz irregularren konjugazioari dagokiona ere.

Grabaketan hartutako arauak buruz ondokoak azpimarratu nahi genituzke:

- Adiera desberdinen definizioak, LPPLn puntuen bitartez bereziak, zenbakitzeko erabakia hartu zen. Horrez gain, homografo-adierazleak ere sartu ziren, horretara datu-baseko taulen atzigako nagusia eratuz: sarrera, homografo-adierazlea, adiera-zenbakia. Eragiketa horien ondoren, LPPLn —bestelako bereizkuntzarik gabe— emandako *avocat* sarrera bien ordeztu, ondoko hiru lerroak dauzkagu HDBan:

*avocat I 1: qui plaide en justice*

*avocat I 2: intercesseur*

*avocat II 1: fruit à noyau en forme de poire*

- Adiera bakoitzaren kategoria sintaktikoa HDBan esplizituki adieraztea erabaki zen, nahiz eta LPPLn horrela egon ez adiera guztien kasuan, aurreko adierarekiko desberdinak direnetan soilik baizik.
- R1\_ACCEPTION taulan definizio motari dagokion eremuko informazioa eskuz erregistratu zen, LPPLko bi definizio motak ongi bereiztearren (ikus § II.3).
- Definizio-testuko hitz bakoitzak aparteko eremu bat hartu zuen datu-basean, definizioen gainean egin beharreko lanak horrela errazteko asmotan (eta kontuan izanik gainera, horrela jokatzek ez zekarrela askoz lan handiagorik grabatzerakoan). Puntuazio-ikurrak eta ikur bereziak (komak, parentesiak, etab.) aparteko "hitz" bezala tratatu ziren grabaketan.

## II.4 Definizioen azterketarako metodoa.

Atal honetan iturburu-hiztegiko definizioak analizatzeko sortu eta jarraitutako metodoa zertan datzan azalduko dugu.

Metodo honen azken urratsean burutu dugun definizioen analisi edo "parsing"erako Alshawi-k deskribatutakoaren (1987; 89) antzeko analizatzailea erabili dugu. Analizatzaile hori —§ II.5 atalean labur-labur azalduko— eskema-hierarkiatan oinarritzen da. Alabaina, Alshawiren lanetan ez da zehazten zer eta nola egin, analizatzaileak, definizioak analizatzeko, behar dituen hierarkiok eratzeke.



Horregatik, gure lan honetan ongi finkatutako hierarkiak eratzeari begira pentsatutako metodoa diseinatzen ahalegindu gara. Metodologia horren diseinuari lotu bagatzaizkio, bi arrazoiongatik izan da batez ere:

- a) Hierarkien eraikuntza oso urrats garrantzitsutzat jotzen dugulako, analisi zuzena egin nahi bada kontu handiz burutu beharrekoa.
- b) Analisi-prozesutik ahalik eta gehien atera nahi bada, ongi oinarritutako hierarkia sendoak eratu behar direlako; eratze hori taxuz egingo bada, ez zaigu nahikoa iruditzen hierarki eratzailleak definizioetan ager daitezkeen perpaus-egiturei buruz izan dezakeen intuizioa.

Metodoaren xedea, definizioen analisiari berari ekin aurretik, hiztegi-definizioen nondik norakoak zehaztea da, horretarako zenbait prozedura estatistiko eta enpiriko erabiliz. Horrela bada, hiztegiko definizioak zer-nolakoak diren zehazte horretan, bertan agertzen diren egitura sintaktikoak zein diren, lexikografoak definizioak idazteko erabiltzen dituen formula bereziak —estereotipatuak edo ez hain estereotipatuak— nolakoak diren, etab. zedarritzen saiatu beharko dugu. Atal honetan —§ II.4— eta ondorengoetan metodoa osatzen duten urratsak eman eta urrats horietako bakoitzean lortutako emaitzak azalduko ditugu.

Metodoaren helburu nagusia beraz, hiztegiko definizioen artean ageri diren mota desberdinak *semantikoki* ezaugarritzea litzateke. Hona metodoa osatzen duten urratsak zein diren:

- 1) Definizio-esaldietan ageri diren hitzen *maiztasun-azterketa* (§ II.4.1).
- 2) Definizioetako hitz-agerpenen *etiketatze eta desanbiguatze lexikala* (§ II.4.2).
- 3) Definizio-esaldien egituraren errepresentazio diren *kategori eskemen zerrendak* lortu eta aztertu (§ II.4.3).
- 4) Aurreko urratsean lortutako *zerrenden behetik gorako analisia*, goragoko mailako egitura sintaktikoak lortzearren (§ II.4.4).
- 5) Definizio-formula estereotipatuen gaineko azterketa enpirikoa. *Erlatore berezitzat* hartuko direnak finkatu (§ II.4.5 et II.4.6).
- 6) Patroi-hierarkiak prestatu eta definizioen *analisi sintaktiko-semantikoa* burutu (§ II.5 et II.5.1).
- 7) Patroi bakoitzari dagokion *egitura semantikoa eraikitze*ko erregela esleitu (§ II.5.3).

### II.4.1 Maiztasun-azterketa.

Autore anitzek (Amsler, 80; Calzolari, 84b; Papegaaij *et al.*, 86) egiaztatu du hiztegia den corpusean zenbait hitzek hartzen duen garrantzi erlatiboa —maiztasunari begiratuta—. Y. Wilks-ek (1972) egindako oharpena, bere CSD sistemako "elementu definitzaileen" eta W7 hiztegiaren gaineko maiztasun-azterketa baten emaitzen arteko antz handiaz, esanguratsua da oso<sup>1</sup>.

Definizio-azpilengoaiaren mamia ez datza, jakina, bere hiztegiaren espezifikotasunean bakarrik. N. Calzolarik (1986, 90a), definizio-azpilengoaia horren esistentzia adierazten duelarik, bere inportantzia azpilengoaia horrek dituzkeen murriztapen lexiko-gramatikaletan legokeela dio, berorren ustiagarritasunari begira oso baliagarri bihurtuko luketenak. (Calzolari & Picchi, 88)n kategori eta erlazio-hiztegi oinarrikoa finkatu dute (ikus § II.5 aurrerago), horretarako alderdi "kuantitatibo eta intuitiboak" kontuan edukiz.

Gure proiektuan, HDBaren gainean egin ditugun lehendabiziko azterketak azterketa estatistikoak izan dira. Azterketon ikergai, definizioetako hitzen maiztasuna eta izaera gramatikala izan dira gehienbat. Horien ondorioz jakin dugu, esate baterako, zein diren definizio-esaldietan maizen ageri diren hitzak, definizio-metalengoiari dagozkion hitzak eta esaera bereziak, hots, kontzeptuak definitzeko lexikografoak erabilitako formulak nolakoak diren, edota sarreren definiziorako lexikografoak zer-nolako gramatika erabili duen.

Azterketa estatistiko horiek burutu eta gero, definizioetako hitzen aurreneko sailkapen bat eta definizio-eskemen bilduma finkatu ditugu. Sailkapen horiek gidatu gaituzte azkenean ezagutzaren errepresentaziorako egin dugun hautuan.

LPPLren gainean egindako lehen azterketa estatistikoak erakutsi du, besteak beste, definizioetan ageri diren hitzen multzoan (75951 hitz-forma) 12350 hitz-forma desberdin daudela. Azpimarratzekoa da azken kopuru hori definizioetako forma desberdinen agerpenei dagokiela, izan ere kontaketa horretarako ez baitzen burutzen agerpenen inongo analisi morfologikorik.

Azterketa horretan ondoko arazoak izan ditugu:

- Hitz-multzoak, lokuzioak, etab. ez ditugu tratatu.
- Arazo morfologikoak: *petit*, *petite*, *petits* hitz-forma diferentetzat hartu dira hasiera batean (elkarren arteko erlazioerik gabekotzat).

---

<sup>1</sup> Zehatzago esanik, Wilks-ek azaltzen duen kointzidentzia azpimarragarria W7an usuen ageri diren 50 hitzen eta bere sistemako "elementu definitzaileen" artekoa da.

- Ortografi arazoak: *de*, *d'* hitz-forma diferentetzat jo dira.
- Hiztegiaren grabaketan egindako sakatze-hutsak.

Azterketa estatistiko horiek garbi asko azaltzen dute delako hitz gramatikalak (preposizio, konjuntzio, etab.) direla usuen ageri direnak. Horien artean baina, alde nabarmenak daude segun eta zer kategoriari dagozkion: adibidez, *de* da forma kanoniko usuena (%10.57) izen-definizioetako hitzen artean, baina adjetiboenetakoan artean *qui* da maizen ageri dena (%8.80). Honek argi adierazten du badirela, izen-definizioen artean, *nom "de" complément*<sup>1</sup> eskema betetzen dutenak, hau da "genus et differentia" erako definizio klasikoak alegia. Adjetiboen kasuan berriz, "*qui*" *groupe-verbal* moduko eskema legoke datu horiek idarokitakoan artean, eskemako aditz-sintagma horrek adieraziko duelarik definitutako adjetiboa hartzen duenak egindako edota jasandako egintza.

Hitz gramatikalen maiztasun erlatiboki altu horiek alde batera utzita, sarrera-azpimultzo desberdinei dagozkien definizioetako hitzen kategoria gramatikaren aldetik ere alde nabarmenak ageri dira: esate baterako, izenen definizioetan izenak dira maiztasun handieneko hitzen artean gailentzen direnak: *action, partie, personne, plante, chose, corps, sorte, ensemble, état, instrument*, etab. bezalako hitzak, hau da kontzeptu-hierarkian toki gorenak hartuko dituztenak, hiponimo asko dituztenak alegia; aditzen definizioetan berriz, izenak gabe aditz-infinitiboak aurkituko ditugu maiztasun handienekoen artean, hala nola *faire, rendre, mettre, être, donner, avoir, prendre* (euskarri-aditzak, erlatoreak askotan), aditzen "genus" inportanteenak edo adieraziz. Adberbio, izenordain, artikulua eta konjuntzioen hiztegi-definizioen azterketak mota honetako sarrerak definitzeko erabili ohi den beste era bat erakusten du: kasu horietan lexikografoak, maiz, gramatikaren metalengoiara jotzen du hitzak definitzeko garaian, eta hori dela medio horra hor *temps, quantité, lieu, façon, négation, adv.* bezalako hitzak maiztasun handienekoen artean. Bukatzeko, adjetiboen definizioetan *a, peut, est, dit, fait* bezalako aditz-forma jokatuak maiz samar agertzeak ere arestian aipatu adjetibo-sintagma —"*qui*" *groupe-verbal* — egituraren presentzia salatuko luke.

Maiztasun-azterketa hauek eraman gaituzte hiztegi kozkor honetako definizioetan ageri diren hitzen estreinako sailkapen bat egitera, geroago ezagutzaren adierazpiderako formalismoan erabiltzeko helburuarekin. Ondoko zerrendetan ikus daitezke —ez dugu exhaustibotasunik bilatzen— definizioetako hitz erabilienak (agerpen-maiztasuna —forma

---

<sup>1</sup> Eskemon sintaxia adierazteko EBNF (Extended Backus-Naur Form) formalismoa erabiliko dugu.

## II. kapitulua

desberdinen kopururak kanonikoaren pean bildurik<sup>1</sup>, jeneralean— %1ekoa edo handiagoa denean), sail bakoitzean hitzak maiztasun handienetik txikienera ordenatuta ageri direlarik:

### IZENEN DEFINIZIOAK

**Determinatzaileak:** *le, un, des, du, au, deux, certain*

**Genus erabilienak:** *action, partie, personne, plante, chose, fait, sorte, ensemble, état, instrument, pièce, caractère, lieu, animal, oiseau, homme, art, forme, qualité, appareil, liquide, objet, fruit, maladie, fleurs, métal, mouvement, arbre, terre, femme, produit, enfant, poisson, substance, vêtement, couleur, jeu*

**Modifikatzaileak:** *petit, est, grand, fait, forme, gros, manque, a, produit<sup>2</sup>*

**Modifikatzaileen sartzaileak:** *de, qui, à, des, du, pour, en, dans, par, au, sur, que, sans, avec, entre*

**Erreferentziak:** *qui, ce, on, se, que, son, certain, autre, celui*

**Konektoreak:** *et, ou*

### ADITZEN DEFINIZIOAK

**Determinatzaileak:** *le, un, des, du, au*

**Genus erabilienak:** *faire, rendre, mettre, être, donner, avoir, prendre, devenir, ôter, tirer, changer, passer, produire, causer, couvrir, jeter, tomber, établir, enlever, fixer, porter, pousser, séparer, couper, exciter, orner, perdre, marquer, réduire, réunir, serrer, disposer, obtenir, ouvrir, agiter, exprimer, sortir, tourner, tromper, appliquer, dire, diriger, détruire, engager, examiner, former, frapper, joindre, nettoyer, parler, prononcer*

**Kasu-sartzaileak:** *de, en, à, des, avec, par, dans, du, sur, au, pour, sans*

**Erreferentziak:** *se, ce, qui*

**Konektoreak:** *et, ou*

---

<sup>1</sup> Emaizta hauetara iritsiko baldin baginen beharrezkoa zen, noski, forma kanoniko bakarrari zegozkion forma desberdinak eskuz, besterik ezean, biltzea (maiztasun handieneko kasuetan besterik ez dugu egin), arestian aipatutako arazoak nolabait ekiditekotan.

<sup>2</sup> Genusaren modifikatzaile bezala sailkatu ditugu nahiz adjetiboak —*petit, grand, etab.*—, nahiz izenordain erlatibo baten ondoren etorri ohi diren aditz-instantziak ('*qui fait*', '*qui forme*', '*qui manque*', etab.).

## ADJETIBOEN DEFINIZIOAK

**Determinatzaileak:** *le, un, du, au, des, deux*

**Genus erabilienak:** *grand, vif, bon, difficile, plein, clair, couvert, digne, mauvais, petit, agréable*

**Egintza- edo egoera-adierazleak:** *a, peut, est, dit, fait, forme, aime, sert, tient, dure, agit, contient, manque, produit*

**Sartzaileak:** *qui, relatif, que, propre, conforme, contraire*

**Modifikatzaileak:** *ne, non, pas, très, peu, plus, bien, beaucoup, après*

**Erreferentziak:** *qui, se, en, que, on, son, dont*

**Konektoreak:** *et, ou*

Benetan interesgarri litzateke emaitzok frantsesezko beste testu-corpusen gainean egindako hitzen maiztasun-azterketa orokorrekin erkatu ahal izatea. Izan ere kasu nabarmenak baitaude, hala nola *il* hitzaren agerpen kopuru eskasaren ondoan —21 bider (0.0280%) besterik ez LPPL osoan—, hor ageri dira *action* (0.8147%), *partie* (0.2213%) edo *ensemble* (0.1640%), gure ustez hiztegi-definizioen corpusean bestelako testuetan baino askoz ere usuago ageri direnak.

### **II.4.2 Desanbiguatze lexikala. Etiketatzeko edo "tagging" a.**

Hiztegiko definizioetako hitz-agerpenen desanbiguatzea zentzu-adierazgailu bat eranstean datza —homografo-adierazlea eta adiera-zenbakia edo—, agerpen bakoitza hitzaren beste esanahietarik bereizteko asmoz. Hori egiteko premia, gure kasuan, helburu dugun kontzeptu-sareak berak sortua da, sare horretako nodoak ez baitira hitzen errepresentazio izango baizik eta hitz-adiera edo kontzeptuena —ezagutza tesaurikoaren errepresentaziorako egin dugun hautuak dakarren premia, azken finean—.

Tagging edo etiketatze delako eragiketaren xedea definizio-testuetako hitz-agerpen bakoitzari dagokion kategoria gramatikala esleitzea da. Normala denez, bai etiketatze-lanetarako zein desanbiguaketarako LPPLn bertan dagoen informazioaz baliatu gara, hasiera batean behintzat.

Estreinako arazoa hiztegi-sarrera ez diren hitz-agerpenek aurkeztu ziguten (21373 hitz-agerpen zehatz-mehatz, agerpen kopuru osoaren %30a ia); multzo hori aztertu ondoren kasu desberdinak zeudela ikusi ahal izan genuen:

## II. kapitulua

- a) LPPLren berezko akatsak: definizioetan maiz samar ageri diren hitzak ez dira hiztegiko sarrera gisa agertzen. Horien artean ditugu *au*, *course*, *et*, *non*, *pas* (adberbioa), *voyager*, etab.
- b) Izenen, adjetiboen (femeninoa eta plurala) eta aditzen (forma jokatuak, gerundioak, partizipioak, etab.) aldakuntza morfolo­giko erregularrak: *grandes*, *peut*, *est*.
- c) Aldaketa ortografikoak, bereziki apostrofodunak eta laburdurak: *d'*, *adv*.
- d) Puntuazio-ikurrak (agerpentzat hartuak, LPPLa makinaratu zen momentuan).
- e) Transkribatze-erroreak eta beste.

Automatikoki etiketatu edo desanbiguezinak ziren hitz-agerpen horien kopurua ahalik eta gehien gutxitzearren, ondoko lanei ekin genien:

- 1) Beste hiztegi txiki batetik hartu (*Larousse de Poche*, Paris: Librairie Larousse, 1979) eta zenbait sarrera (31 hitz, 73 adiera) beren definizioekin datu-basean sartzea erabaki genuen, LPPLn falta ziren sarrera nabarmenen hutsunea betetzeko (ikus gorago).
- 2) LPPLko definizioetako hitz-agerpen bakoitzari bere forma kanonikoa esleitu zitzaion. Lan hori eskuz osatu zen agerpen usuenen (agerpen-maiztasuna  $\geq 10$ ) dudazko kasuetan, kasu horietan analizatzaile morfolo­giko batek emango ligukeen analisi-aniztasuna ebatziz: *fait* agerpenari adibidez, *fait* (izena) edota *faire* aditz-infinitiboa esleitu zaizkio kanoniko gisa, kasu bakoitzean betetzen zuen funtzioaren arabera. Gainerako kasuetan, oraingoz, agerpena bera hartu dugu forma kanonikotzat.

Forma kanonikoa esleitzearekin batera agerpen bakoitzari aldakuntza ortomorfologikoa esango diogun informazioa ere erantsi genion<sup>1</sup>, hau da, forma kanonikoa eta agerpena desberdinak direlarik bien arteko aldea azaltzen duen arrazoi morfolo­giko edota ortografikoa adierazten duena. Adibidez, *fait* agerpenak, *faire* kanonikoari dagokiolarik, 'PI3' (indikati­boko orainaldiko 3. pertsona adieraziz) edo 'PAR' (partizipio forma) izan ditzake aldakuntza-kode bezala; edo bestela, forma kanonikoa *fait* bera denean, aldakuntza-kode hutsa, kasu horretan izena izango baita. Beste adibide bat ematearren hona *d'* agerpenarena, kanoniko gisa *de* esleituko zaiona eta 'APS' (apostrofoduna) aldakuntza-kodetzat.

---

<sup>1</sup> Eransketa guzti hauek HDBko R2\_DEFINITION taulan egin ditugu, horretarako beharrezko ziren eremuak bertan definituz.

- 3) LPPLko zenbait sarreratan ageri den forma flexionatuei buruzko informazioaz — HDBan grabatua zegoena— ere baliatu ginen: batez ere adjetiboen femeninoak izan ziren, beren agerpenei dagokien forma kanonikoa (maskulinoa, hori baita LPPLk sarrera gisa ematen duena) esleitzeko erabilitakoak.

Guzti hori egin eta gero, automatikoki etiketaezinak ziren forma kaniko desberdinen kopurua 4770era gutxitua genuen, 8172 agerpen denetara (agerpen kopuru osoaren %10a pasatxo).

Desanbiguatze-programa egikaritu ondoren konturatu ginen agerpen-maiztasun handieneko hitzen artean bazirela desanbiguatzeke geratutakoak, eta baita kategoriarekin etiketatu gabekoak ere hiztegian kategoriatzeko sarrera polisemikoei zegozkielako: esaterako, *en* preposizio edo izenordain pertsonala ('*prép. et pron. pers.*'). Anbiguetate horrek hor irautea gogaikarria zitzaigun oso definizioen eskema sintaktikoak lortzeari begira —gure hurrengo helburua—, zeren eskema horietako asko eta asko geratuko baitzen anbiguo hitz horietakoren baten kategoriaren faltan. Horregatik, horietako batzuk eskuz desanbiguatzeke erabakia hartu genuen; beste batzuk berriz, definizioetan erabilitako lengoai mota edo kategoriatzeko desberdinen agerkidetzari buruzko datuak erabiliz desanbiguatu genituen, heuristikoki<sup>1</sup>.

Egindako prozesu horien ondoren<sup>2</sup>, etiketatzeari eta desanbiguatze lexikalari (zentzu-adierazgailuaren esleipena) dagozkion portzentaiak honako hauek ditugu:

- 67779 hitz-agerpen (%89.2a) beren kategoriatzeko gramatikarekin etiketatuak; horietarik 26032k, %34.11k, badute jada zein zentzuri dagozkion zehazten duen adierazgailua (hitz monosemikoak dira gehienbat, jakina).
- 15508 definizio, adieren %67.45a, erabat etiketatuak (kategoria hitz-agerpen guztietan esleitua).

Beharrezko genduz orain analizatzaile morfologiko bat kategoriarekin etiketatu gabe geratutako kasu gehienak konpondu ahal izateko (pluralean dauden formak gehienbat). Zentzuen desanbiguaketa lexikalak berriz, arazo gaitzagoak aurkezten ditu, erabat askatzeko prozedura erdiautomatikoak eskatuko dituztenak. Problema korapilatsua, eskuz egiteko ere benetan zaila: definizioetan erabilitako baina hiztegiko sarrera ez diren zentzuak, erabilitako

---

<sup>1</sup> Adibidez *qui* izenordaina, desanbiguatzeke geratua '*pron. inter.*' kategoriako adieragatik, '*pron. rel.*' bezala etiketatu genuen hiztegi osoan (izan ere, nekez agertuko baita galdezko izenordain bezala definizioetan).

<sup>2</sup> Datu horiek HDBaren egungo egoerari dagozkio, hau da, aipatu berri ditugun —edo hurrengo ataletan azalduko ditugun— etiketatze- eta desanbiguatze-prozesu automatiko eta erdiautomatiko horiek burutu ondorengo egoerari.

## II. kapitulua

zentzua definitutako zenbaiten nahasketa deneko kasuak, hiztegitik adituak ere hitzari dagokion zentzua bereizteko lanak izango lituzkeeneko kasuak, etc..

II.3 irudian ikus daiteke zer-nolako itxura duen HDBko informazioak, azaldu berri ditugun prozesuak burutu ondoren; horra hor R2\_DEFINITION taulako bi lerro, *spatule I I* eta *bolide I I* sarrerei dagokien informazioarekin.

<b><i>spatule I I:</i></b>	<i>sorte</i>	<i>de</i>	<i>cuiller</i>	<i>plate</i>	(definizio-esaldia)
	<i>sorte I</i>	<i>de I</i>	<i>cuiller II</i>	<i>plat I</i>	(forma kanonikoak)
	f.	prép.	f.	adj.	(kategoriak)
				F	(ald. orto-morfologikoak)
<b><i>bolide I I:</i></b>	<i>véhicule</i>	<i>qui</i>	<i>va</i>	<i>très vite</i>	(definizio-esaldia)
	<i>véhicule II</i>	<i>qui I</i>	<i>aller I</i>	<i>très II vite II</i>	(forma kanonikoak)
	m.	pron. rel.	vi.	adv. adv.	(kategoriak)
			PI3		(ald. orto-morfologikoak)

**II.3 irudia.-** HDBko bi sarrera, etiketatze- eta desanbiguaketa-prozesuak burutu eta gero.

### II.4.3 Definizio-eskemak.

Hiztegiko definizioetan erabilitako egiturak zer-nolakoak diren jakitearren, *definizio-eskemen* zerrendak ateratzea izan zen gure hurrengo pausoa, horretarako datu-basean zegoen informazioaren egituraketaz baliatuz. Horrek berehala eraman gintuen definizioen behin-behineko gramatikatxo bat zehaztera.

Definizio-eskema hauek "kategoria morfosintaktikoz osatutako sekuentzia" bezala hartzen ditugu. Aurreneko fase batean, definizio-eskema horien lorpena etiketatze-prozesuan "erabat" etiketaturiko definizioetarako egin genuen (izatez, beren lehen bost hitzetan etiketatu etarako). Dударik gabe, aukera horrek definiziorik laburrenak hartzen ditu, etiketatze maila handiagoa dutenak alegia.

Bestalde, hemen ez ditugu aipatuko definizio-metalengoiari dagozkion hitz berezien funtzioan adieraz daitezkeen eskemak, eskema sintaktiko hutsak baizik, beti ere kontuan hartuz eskema horiei egitura semantiko egokiak esleitu beharko dizkiegula ondoren.

Definizioen analisi sintaktiko-semantikorako egindako hautuak kategoria desberdinetako sarreren definizioetarako hierarkia bereziak eskatuko zizkigula kontuan hartuz, izen, adjetibo, aditz eta adberbioen definizioetatik lista desberdinak atera genituen. LPPLn dauden



adberbio-definizioen kopuru murrizak —173, %0.76a—, eta definizio horien eskemak nahiko "sakabanatuak" ageri zirela ikusteak adberbioen azterketa alde batera uztera bultzatu gintuen. Beraz hemendik aurrera izen, adjetiboen eta aditzen definizioez soilik arituko gara.

Definizio-eskema sintaktikoen aurreneko zerrendak hitz-agerpen gehiago etiketatzeko —eskemen analogian oinarrituta— erabili genituen. Adibidez, *mot-quelconque préposition nom* zuten definizioetako lehenengo hitzari 'NOM' etiketa esleitzeko ez zegoen arazorik, izan ere izen-definizio homologagarrien gehiengoa baitzen *nom préposition nom* erakoa (701 vs. 7 besterik ez bestelako eskemekikoak). Beste adibide bat ematearren, ikus dezagun *nom préposition article mot-quelconque* eskema duten definizioen kasua. Horrelako 36 izen-definizio zeuden, non kategori etiketa laugarren hitzari falta baitzitzaion. *nom préposition article nom* bezalako eskema 146 izen-definizioek zeukaten, eta horixe zen *nom préposition article* (ordena horretan) hasiera zuen eskema bakarra. Era horretako azterketak *mot-quelconque* elementua zuten eskema erabilienekin eginez, eta eskemalortze/etiketatzeko prozedura hainbat bider errepikatuta, iritsi ginen atal honen bukaeran emango ditugun emaitzetara.

Jabetzen gara, noski, ihardupide honek sortuko bide zituen errakuntzez, zenbait kasutan estrategia heuristiko hutsetan oinarritzen den neurrian. Hala eta guztiz ere, errakuntza horiek gutxi direlakoan gaude, eta batez ere, gure helburua ziren egitura sintaktikoen maiztasunari buruzko emaitzak ez dituztela funtsean aldatu pentsatzen dugu.

Definizio-eskemei buruzko datu estatistikoak, eta § II.4.5 azpiatalean azalduko dugun definizio-formula estereotipatuen gaineko ikerketa enpirikoa, abiapuntu egokiak izan ditugu analisi sintaktiko-semantikorako behar ziren patroihierarkiak zehaztu eta eratzeko.

Ondoren izen, aditz eta adjetiboen definizioen gramatikatxo bat azalduko dugu EBNF formalismoan (gramatika hori betetzen bide duten definizio kopuruen estimazioak emanez). Arestian azalduko prozesu horien emaitza dugu gramatika hau. Kopuru horiek berriz, estimazio hutsak direla esan behar dugu, proiektuaren etapa horretan definizio-egiturei buruz egindako hipotesietatik —aurreko azterketetako emaitzetan oinarrituz— eratorriak, ondoren analisi sintaktiko-semantikoaren urrats desberdinetan erabilitako gramatika edo hierarkiak zehazteko lagungarri izan ditugunak.

IZENEN DEFINIZIOAK: izen-sintagmaren egitura.

*définition\_de\_nom* = [adj\_qualificatif] nom {modificateur} {"", "définition\_de\_nom"} | proposition\_relative.

*modificateur* = adj\_qualificatif | préposition [quantificateur] [adj\_qualificatif] nom [modificateur] | préposition\_infinif\_verbal | [préposition\_pron\_dém] proposition\_relative.

## II. kapitulua

Estimazioa: LPPLko izen-definizioetarik %60-85ek betetzen du gramatika hori.

ADITZEN DEFINIZIOAK: aditza infinitiboan duen aditz-sintagmaren egitura.

*définition\_de\_verbe* = [*pron\_pers*] *infinitif\_verbal* {*complément*} {"", "*définition\_de\_verbe*}.

*complément* = [*préposition*] [[*quantificateur*] *nom*] | *préposition infinitif\_verbal* | *adj\_qualificatif* | *adverbe*.

Estimazioa: LPPLko aditz-definizioetarik %70-90ek betetzen du gramatika hori.

ADJETIBOEN DEFINIZIOAK

*définition\_de\_adjectif* = ([*modificateur*] *adj\_qualificatif* | [*adj\_qualificatif*] *préposition* ([*quantificateur*] *nom* | *infinitif\_verbal*) | *proposition\_relative*) {"", "*définition\_de\_adjectif*}.

*modificateur* = *adverbe*.

Estimazioa: LPPLko adjetibo-definizioetarik %60-90ek betetzen du gramatika hori.

### II.4.4 Behetik gorako analisi sintaktikoa: kategoria sintagmatikodun eskemen lorpena.

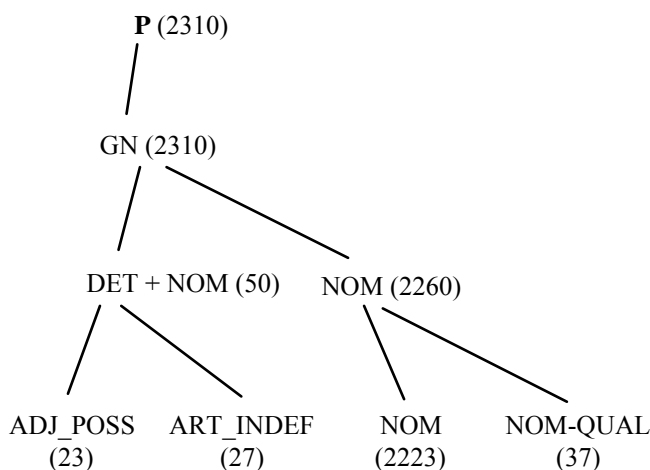
Kategori eskemen zerrendak lortu ondoren, eskemon behetik gorako edo bottom-up erako analisiari ekin genion, kategoria sintagmatikozko sekuentzien maiztasunak lortzeko asmotan. Kontua da orain definizio-esaldietan erabilitako sintagma-sekuentzien zerrendak lortzea — hitz-agerpen bakar bati dagozkion kategoriak baino maila altuagoko egiturez osatuak—, maiztasun absolutuz ordenaturik.

<p>ADJ_POSS+NOM (23 bider)</p> <pre>       GN      /  \     DET  NOM         ADJ_POSS           </pre>	<p>ART_INDEF+NOM (27 bider)</p> <pre>       GN      /  \     DET  NOM         ART_INDEF           </pre>
<p>NOM (2223 bider)</p> <pre>       GN               NOM               NOM           </pre>	<p>NOM-QUAL (37 bider)</p> <pre>       GN               NOM               NOM-QUAL           </pre>

**II.4 irudia.-** Lau eskema sintagmatiko, dagozkien maiztasunekin.

Analisi hori burutu ahal izateko gramatika bat definitu behar izan dugu: aurreko azpiatalean aipatu kategori eskemetako bakoitzetik bertan ageri diren sintagmen funtzioan adierazitako eskema berri bat lortu dugu. Eskema horiek bildu egin ditugu, adabegi bakoitzeko maiztasunak metatuz (ikus II.4 irudia).

II.4 irudi horretako eskemak bilduz gero, 2310 perpaus-eskema (P) lortuko ditugu (ikus II.5 irudia), denak izen-sintagma egitura (GN) desberdinak dituztenak gainera (50 determinatzaile bat eta izen batez osaturikoak, eta 2260 izena besterik ez dutenak).



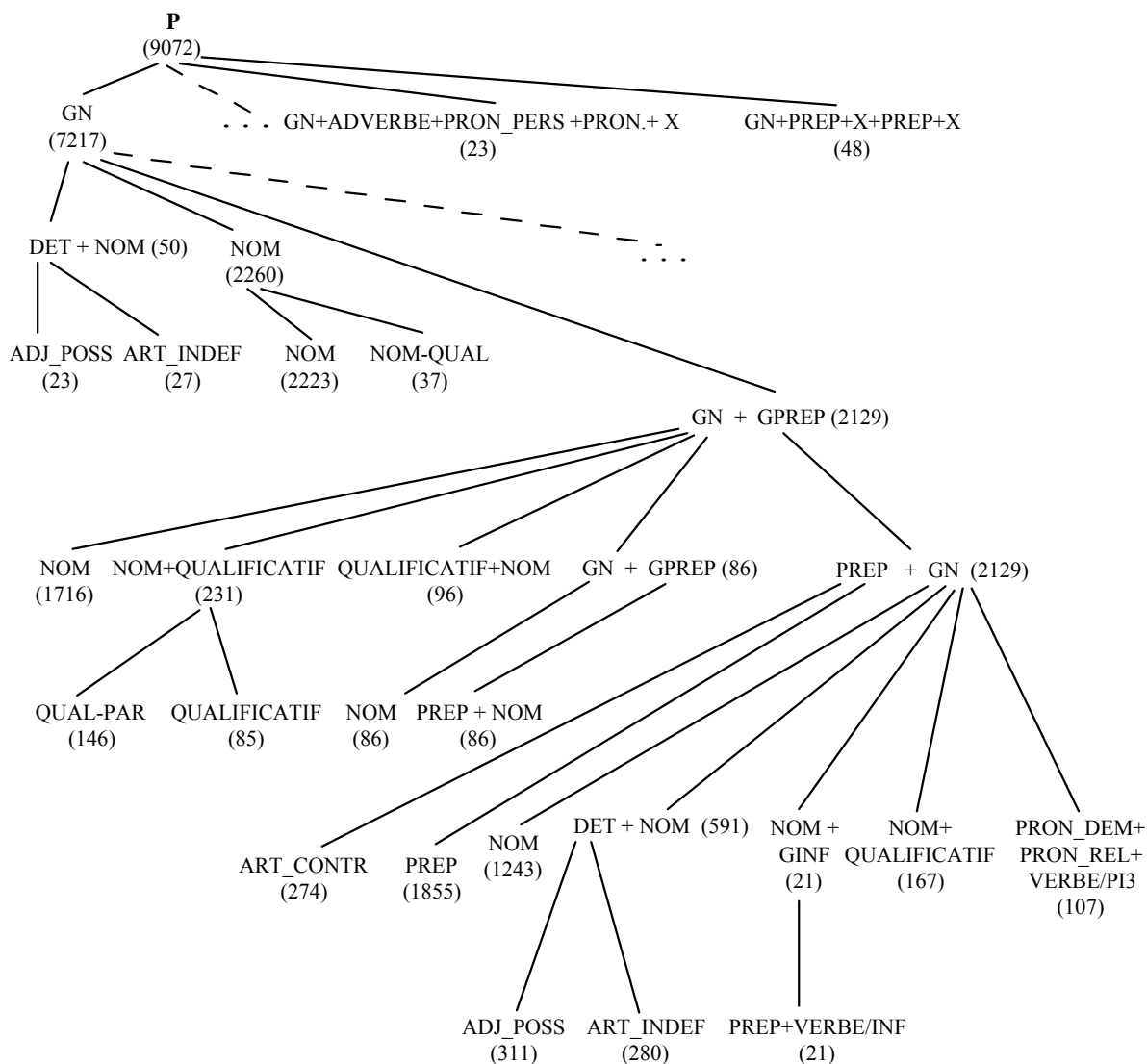
**II.5 irudia.-** Aurreko irudiko eskema sintagmatikoen bilketa<sup>1</sup>.

II.6 irudian ikus daiteke eskema-bilketa prozesu horiek izenen definizioen kasuan emandako emaitzaren parte bat: kontuan hartutako 9072 analisien artean, 7217k izen-sintagmak errepresentatzen dituzte, eta gainerakoak kategoria ezezagunen bat duten eskema ez-osoer dagozkie (irudian X ikurra duten eskemak).

Behetik gorako analizatzailea aplikazio honetarako bereziki diseinatu eta inplementatu zen (LISPez). Kategori eskemetako kategori sekuentzia bakoitzeko sintagma-sekuentzia bakarra ematen du analizatzaile honek. Analizatu beharreko kategori eskemako osagaiekin diagrama bat ("chart") eratu ondoren, analizatzaileak osagai posibleen sekuentzia bat aukeratuko du; aukera hori beti osagairik luzeenari (hitz kopuruari begiraturaz) egokituko zaio. Hitz kopuru berdineko bi osagai edo gehiago baldin badaude, eta horiek baldin badira luzeenak, analizatzaileak emaitza gisa kasu guztiak kontatuko ditu. Horretara, eskema kopuru totala analizatutako eskemen kopurua baino handiago izan badaiteke ere, analisi-aukera interesgarriak galtzeko arriskurik ez da izango.

---

<sup>1</sup> Ohar bedi irakurlea, kasu honetan zuhaitzak ez duela —zuhaitz sintaktiko klasiko baten gisan— adabegi bakoitzari dagokion egitura errepresentatu nahi, adabegi bakoitzean bildutako azpiegituren multzoa baizik.



**II.6 irudia.-** Izen-definizioen eskema sintagmatikoak (partez).

**II.4.5 Azterketa estatistikoa eta definizio-formulei buruzko ikerketa empirikoa. Erlatore bereziak.**

Definizioetako hiztegiaren mailako maiztasun-azterketen ondoren —alde batera utzita orain definizioen egitura sintaktiko orokorrari dagozkionak—, definizio-metalengoiako eraikuntza bereziak aztertzen hasi ginen.

*Erlatore berezien* —metalengoiatzko egitura horiei horrela deritzegu— bilaketan abiapuntu gisa izenen, aditzen, adjetiboen eta adverbioen definizioetako hitz-agerpenen maiztasun kalkuluak hartu ditugu. Kalkulu horien emaitzak eta gai honetaz hainbat autorek egindako azterketak (Calzolari, 84b; Markowitz *et al.*, 86; Vossen *et al.*, 89) izan ditugu gidari LPPLren gainean erlature bereziori buruz egin dugun ikerketa empirikoan.

## II. kapitulua

Aurreko ataletan aipatu dugu jadanik metalengoaia lexikografikoaren —edo, hobe esanda, definitorioaren— kontzeptua (Ferrari, 80; Calzolari, 86, 90a; Calzolari & Picchi, 88). (Guckler, 78)n, definizioetan maiz ageri diren hitzak direla eta, hitzok hizkuntzaren egitura semantikoren "gunea" edo osatuko luketela esaten da. Bere proiektuaren helburua "lengoaia naturalaren deskribapen semantikorako metalengoaia" definitzea litzateke hain zuzen ere, horretarako lexikografoek beren lanean erabilitako terminoetan oinarrituz. Gune hori semantikoki desanbiguatutako hitzek osatuko lukete, izatez kontzeptuarteko erlazioen deskribapenerako "primitibo semantikoak" edo lirakeenak.

Erlatore bereziak, hiztegi-definizioetan, ondoko kasuetan agertu ohi dira:

- Maiz samar erabilitako eraikuntzetan.
- Semantikoki "hutsak" edo "zehaztasun gutxikoak" diren hitzetan<sup>1</sup>.
- Definizioa osatzen duen sintagmaren —izen-, aditz-sintagmaren— gune sintaktikoa eta semantikoa (genusa) bat ez datozenean. Ideia hau zenbait lanetan aipatua da (van den Hurk & Meijs, 86; Vossen, 89; Vossen *et al.*, 89), eta beti azpimarratzen da hori gertatzen dela kategori aldaketa bat dagoenean, hau da definiendumaren kategoria eta "definiens"aren karga semantikoa daraman osagaiarena desberdinak direnean.
- Teorikoki ez-anbiguo diren eraikuntzetan (hau da, '*ensemble de*' formulaz egindako eraikuntza bat daukagunean ongi dakigu *ensemblen* zein adierari, eta batez ere *deren* zein adierari, dagokien hitz horien erabilera testuinguru konkretu horretan).

Lengoaia lexikografikoaren eraikuntza espezifiko hauen gainean egin diren lanen artean azpimarratzekoak dira, gure iritziz, P. Vossen, W. Meijs eta berorien taldeak egindakoak. Lan horiek LINKS proiektuaren barruan garatu dira batik bat; proiektu horren helburu nagusia LDOCE hiztegiko definizioak sistematikoki errepresentatzeko datu-base semantiko bat eratzea litzateke (Vossen *et al.*, 89).

Autore horiek hiztegiko definizio edo "esanahi-deskribapenak" sailkatu egin dituzte — ikuspuntu erlazional batez, eta bereziki "definizio-gunearen karga semantikoari" begiratuz— lau eraikuntza motatan: gune ez-komplexuak ("non-complex kernels") alde batetik, hau da, "link", "synonym" eta "shunter" direlakoak, eta gune konplexu edo "linker"ak bestetik (Vossen, 89; Vossen *et al.*, 89; Vossen & Serail, 90).

---

<sup>1</sup> Semantikoki "zehaztasun gutxi" duten hitzak, hala nola *member*, *part*, *etab.*, edo "hutsak" *thing*, *everything*, edo *something* bezala (van den Hurk & Meijs, 86).

Autore horien arabera, definizio batean "*link*" motako eraikuntza bat dago "genus et differentia" erako "benetako" definizio bat denean, hau da, definiensaren gune sintaktikoa eta definiendumaren genusa —semantikoa— bat datozenean, eta genus horren aldamenean bere eremu semantikoa murrizten duten aurre- edo post-modifikatzaileak daudenean (differentia). Horrelakoetatik erauz daitezke, erlazio hiper-hiponimikoez gainera, definiendum-differentia edo genus-differentia erlazioak.

Eraikuntza mota horren muturreko kasua litzateke, autore horien iritziz, "*synonym*" delakoa, non eduki semantikoa jeneralean hitz bakar batean egon ohi baita.

Definizioaren karga semantikoa haren gune sintaktikoa ez den —eta normalean beste kategoria sintaktiko bateko— hitz batera "lerratzen" denean berriz, "*shunter*" motako eraikuntza baten aurrean geundeke. Eraikuntza mota honen adibide bezala '*act of*', '*quality of*', etab. bezalako eraikuntzak ematen dituzte, zeintzuk izen-kontzeptu baten esanahia —definienduma— aditz edota adjetibo baterantz lerrarazten baitute. Definizio horien gune sintaktikoa informazio handirik ematen ez duen hiperonimo "orokorregi" bat litzateke, autore hauen arabera.

Azkenik, "*linker*" delakoak "shunter"en antzeko eraikuntzak lirateke, "linker" a bera of preposizioa eta definitutako hitzaren kategoria bereko izan ohi den osagai batez osatuak. "Shunter"en kasuan bezalaxe, definizioaren karga semantikoa ez dago haren gune sintaktikoaren baitan, beste elementuren batean baizik. Hemen ere hiperonimo orokorregiak aurkituko dira, eta definizioak ezarritako erlazioerik inportanteena definienduma eta of preposizioaren ondotik datorren osagaiaren artekoa izango da. Hauen artean leudeke '*part of*', '*group of*', '*piece of*', etab. bezalako eraikuntzak.

Gure erlature bereziaren kontzeptua bada, azaldu berri ditugun "shunter" eta "linker"en hurreko litzateke. Adibideren bat ematearren, gurean erlature berezitat hartu ditugu izen-definizioetan ageri diren '*membre de*' tankerako eraikuntzak, '*faire*' faktitiboaren erabilera aditz-definizioetan, edo '*qui*' izenordainaz eraturakoak.

Elementu horien gaineko banakako ikerketak erakutsi du definizio-esaldietan zer-nolako egitura azaltzen duten elementuok; ikerketa hori bera izan dugu gidari erlatureoi dagozkien egitura semantikoaren zehaztapena egiterakoan. Definizio klasikoaren kasuan ez bezala, definizio-mota honi dagokion egitura semantikoa erlature berezi horien funtzioan eraikiko da.

Hizkuntza desberdinetako hiztegi desberdinak aztertuz egindako lanek antz nabariak dituzte. W7 eta LDOCE ingelesezko hiztegien gaineko ikerlanetan, N. Calzolariren taldekoek italierazko hiztegiekin egindakoetan edo, gurean bezala, frantsesezko hiztegien gainean egindakoetan (Pigamo, 90) lortutako metalengoaiazko eraikuntzen zerrendak hain

## II. kapitulua

dira hurbilak non eraikuntzon orokortasuna tradizio lexikografiko hutsa —azaleko antzeko errealizazioetan gauzatua— baino areago doan zerbait dela pentsatzen baitugu, batez ere ezartzen dituzten kontzeptuarteko erlazio lexiko-semantikoen aldetik.

Ez dugu hemen emango erlature berezi horien zerrenda osoa. Hurrengo azpiatalean hiztegiaren gaineko azterketatik erauzitako erlazio lexiko-semantikoen bilduma aurkeztuko dugu: erlazio horietako asko eta askoren errepresentazioa erlature berezi horien bitartez gauzatuko da. Hain zuzen ere III. kapituluan, hiztegi-ezagutzaren errepresentazioaz ariko garelarik, azalduko ditugu banan-banan erlature horien bitartez gauzatutako definizioak errepresentatzeko aukeratutako atributu bereziak (atributu definatorioak, DEF-ACTIOND, DEF-MEMBRED edo DEF-CEQUI bezalakoak). Irakurleak jo dezake halaber (Agirre *et al.*, 91) gure lanera, edota txosten honen A eranskinera, non aurkituko baititu definizioen analisi lexiko-semantikoaren azken fasean erabilitako hierarkia konkretuak. Hierarkia horietako patroi batzuetan argi asko ikus daiteke zer-nolako metalengoiatzko egitura den patroi horietako bakoitzak "ezagutu" nahi lukeena.

### II.4.6 LPPLko definizioetatik erauzitako erlazio lexiko-semantikoak.

Atal honetan laburki azalduko dugu LPPLko definizioetan edirendako kontzeptuarteko erlazio lexiko-semantikoen multzoa.

Sailkatzeko eskema moduko bat proposatu ondoren, erlazioak banan-banan azalduz joango gara, horretarako erlazio bakoitza gauzatzen deneko definizioen sintaxia eta LPPLtik zuzenean ateratako adibideak emanez. Batzuetan erlazioek dituzten zenbait propietate komentatuko ditugu, HEBaren gainean nolabaiteko portaera deduktiboa ezartzeko aukera eman diguten neurrian.

Hona hemen erauzitako erlazioak sailkatzeko erabilitako eskema:

- *Erlazio paradigmaticoak:*
  - a) Sinonimia eta antonimia.
  - b) Taxonomikoak: hiperonimia/hiponimia eta berezko taxonomia.
  - c) Meronimikoak.
  - d) Bestelakoak.
- *Erlazio sintagmatikoak:*
  - a) Eratorpena (erlazio morfologikoa).
  - b) Zerikusi morfologikorik gabeko elementuen artekoak.



c) Bestelakoak.

### Erlazio paradigmaticoak:

#### a) *Sinonimia eta antonimia.*

Hiztegian kontzeptuak azaltzeko erruz erabilitako teknika da *sinonimia*, eta areago LPPL bezalako hiztegi txikietan. Konkretuki, LPPLko izenen %22a, aditzen %39a eta adjetiboen %19a sinonimo edota kuasi-sinonimo diren kontzeptuen bitartez definiturik dator. Definiendumaren kategoria bereko hitz bakar batek —edo batzuek, komaz bereizirik— osatutako definizioak lotura sinonimiko implizitu gordetzen du gehienetan. Badira, hala ere, erlazio sinonimikoa esplizituki adierazten duten definizioak ("*syn. de*" *mot-quelconque*) baina, LPPLn behintzat, nahikoa urri dira horrelakoak. Definizio sinonimikoen egitura honelakoa litzateke:

*définition\_de\_nom= nom {"", " nom}.*

*définition\_de\_verbe= infinitif\_verbal {"", " infinitif\_verbal}.*

*définition\_de\_adjectif= adjectif {"", " adjectif}.*

edo —bestelako definizio batekin batera<sup>1</sup>—, izenen kasuan esaterako:

*définition\_de\_nom= nom {"", " définition\_de\_nom}.*

Erlazio honek aukera emango digu printzipioz, zuhurtzia handiz bada ere, simetria eta trantsitibotasuna bezalako propietateak ustiatzeko.

Adibideak:

***extrémité I I:*** *bout, fin*

***exhiber I I:*** *montrer, présenter*

Batez ere adjetiboen definizioetan edirenen dugun *antonimia* dela eta berriz, honako sintaxi hau betetzen duela ikusi dugu:

*définition\_de\_adjectif= ("non" / "pas") adjectif [{"", " définition\_de\_adjectif}.*

kasu honetan ustiagarria den propietatea simetria delarik (alde batera utziz oraingoz erlazio lexikalen arteko konposaketa, zeinaz III. kapituluaren ihardungo baitugu, HEBaren aberasketa dinamikoa jorratzean)

---

<sup>1</sup> Definizio misto hauen kasuan bigarren osagaia ez da izaten, noski, definizio sinonimikoa.

## II. kapitulua

Adibideak:

*difficile I I: pas facile*

*déloyal I I: non loyal*

### b) *Taxonomikoak.*

Bi taxonomi mota bereiztu ditugu: batetik *hiponimiaz*, eta beronen alderantzizkoa den *hiperonimiaz*, eraturakoak —delako "genus et differentia" definizio klasikoetan gauzatuak—, eta bestetik "berezko" taxonomiak edo esango diegunak, hau da, definizioetako '*sorte de*', '*genre de*', etab. bezalako formulak erabiltzearen poderioz eraturakoak.

Definizio klasikoetan definiensaren gunea (genusa) definitutako hitzaren (definienduma) kategoria gramatikal bereko hitz bat izaten da, diferentzi partea (differentia) osatzen duen modifikatzaile edo osagai baten ondoan etorri ohi dena. Genusa da definitutako kontzeptuaren hiperonimoa. II.4.3 azpiatalean emandako gramatikaxoetako lehen lerroetan aurki daitezke guneak, definitutako hitzaren kategoria berekoak direnak eta, normalean, differentia osatuko duten modifikatzaile edo osagaiez lagunduak etorriko direnak.

Taxonomi mota bi hauei begiraturuz gero, hierarkiak —gutxi edo gehiago mihiztatuak— eratzen dituztela ikusiko dugu; hierarkiotan kontzeptuen arteko ordena-erlazio bat ezarriko da. Erlazio taxonomiko hauei esker, hierarkia horietako elementuek heredatutako propietateak deduzitu ahal izango dira hiztegi-ezagutza ustiatzeko garaian.

Adibideak:

*débarbouiller I I: laver le visage*

*laver I I: nettoyer avec un liquide*

Izen-kontzeptuen artean lotura taxonomiko "berezkoak" edo ezartzen dituzten definizioen sintaxia honako hau litzateke:

*définition\_de\_nom= ("espèce de" / "sorte de" / "genre de") groupe\_nominal.*

Adibideak:

*brioche I I: sorte de pâtisserie*

*bolet I I: espèce de champignon*

c) **Meronymia.**

Erlazio meronimikoetan ezagunenetariko bat "partea-osoa" delakoa izango da. LPPLren azterketan, definizioetan gertatzen diren erlazio meronimikoen gauzatze desberdinak ediren ditugu. Kasu honetan hierarkia meronimikoak eraikitzea da kontua, ahaztu gabe erlazio horien trantsitibotasunak dituen arazoak (Winston *et al.*, 87). Hona hemen izen-kontzeptuen arteko erlazio horien azaleko gauzatzeen sintaxia:

*définition\_de\_nom= ("élément de" / "membre de") groupe\_nominal.*

*définition\_de\_nom= [adjectif] "partie" [adjectif] "de" groupe\_nominal.*

*définition\_de\_nom= "pièce de" groupe\_nominal.*

*définition\_de\_nom= ("ensemble de" / "réunion de" / "groupe de")  
groupe\_nominal\_pluriel.*

*définition\_de\_nom= nom ("à" / "avec") nom. (batzuetan)*

Adibideak:

*académicien I 1: membre d'une académie*

*racine I 1: partie enterrée de la plante*

*assiette I 2: pièce de vaisselle*

*orchestre I 1: ensemble d'instruments de musique*

*coucou I 2: pendule à sonnerie (baina **cymbale I 1: instrument à percussion**)*

d) **Bestelako erlazioak (paradigmatikoak).**

**Mailaketa.** Adjetibo eta aditzen definizioetan aurkitu ditugu mailaketa-erlazioak. Kasu horietan hiponimia erlazioa ere bertan ageri da. Hona sintaxia:

*définition\_de\_adjectif= "très" adjectif.*

*définition\_de\_verbe= infinitif\_verbal "de nouveau".*

Adibideak:

*charmant I 1: très agréable*

*rallumer I 1: allumer de nouveau*

## II. kapitulua

**Baliokidetza.** Aditz-partizipio baten bitartez definitutako adjetiboen kasuan dagoen erlazioari "baliokidetza semantikoa" esango diogu, izan ere horrelakoetan ezin baita esan sinonimia, hertsiki bederen, dagoenik.

*définition\_de\_adjectif= participe\_verbal.*

Adibidea:

**archaïque I 1:** *vieilli*

**Faktitibotasuna.**

*définition\_de\_verbe= "faire" verbe.*

Adibideak:

**abreuer I 1:** *faire boire*

**annoncer I 1:** *faire savoir*

**Erreflexibotasuna.** Parafrahirako informazio interesgarria ematen digu erlazio honek.

*définition\_de\_verbe= "se" verbe.*

Adibidea:

**voyager I 2:** *se déplacer*

**Gabezia.**

*définition\_de\_nom= "manque de" nom.*

Adibidea:

**bêtise I 1:** *manque d'intelligence*

**Erreferentziazkoa** (*aurreko adierarekiko*). Lexikografoari zenbait definizio laburtzeko aukera ematen dion formula.

*définition\_de\_nom= "son" nom.*

Adibidea:

**assiette I 3:** *son contenu* (**assiette I 2:** *pièce de vaisselle* adierari erreferentzia eginez)

### Erlazio sintagmatikoak:

Delako erlazio sintagmatiko hauek kategoria desberdinetako kontzeptuen artean ezarri ohi dira. Lehen aipatu ditugun "shunter" eta "linker" kontzeptuak ageri dira berriro hemen, gure *erlatore berezien* iturburu izan ditugunak alegia<sup>1</sup>.

- a) **Eratorpena (erlazio morfologikoa).** Hemen sailkatu ditugu elkarrekiko erlazio morfologikoren bat duten kontzeptuen arteko erlazioak (erlazio morfologikoa kontzeptuen horien adierazleen artean egongo da, noski), hau da, jeneralean eratorpenaz erlazonaturik daudenak. Hemen aurki daitezke aditzekin edota adjetiboekin erlazonatutako izenak, adjetiboekin erlazonatutako aditzak, izenekin edota aditzekin erlazonatutako adjetiboak, eta abar<sup>2</sup>.

**Aditzetiko izenak.** Izen-kontzeptu bat aditzeko batekin erlazonaturik.

- Agente orokorra (edo prototipikoa): definitutako izenak aditz-formak adierazitako egintzaren egilea denotatzen du.

*définition\_de\_nom= "ce qui" verbe\_PI3.*<sup>3</sup>

Adibideak:

*fermeture I 2: ce qui ferme*

*amusement I 2: ce qui amuse*

- Objetu orokorra (edo prototipikoa): definitutako izenak aditz-formak adierazitako egintzaren objektua (prototipikoa) denotatzen du.

*définition\_de\_nom= "ce qu'on" verbe\_PI3.*

Adibideak:

*apport I 1: ce qu'on apporte*

*dette I 1: ce qu'on doit*

- Bestelako aditzetiko izenak:

---

<sup>1</sup> Jakina, erlatore espezifikoko erlazio sintagmatikoetan ez ezik paradigmaticoen artean ere badira.

<sup>2</sup> Erlazioen sailkapena hemen eratorpenean oinarri harturik egiten badugu ere, guri interesatzen zaiguna, azken finean, kontzeptuen artean ezarritako erlazio lexiko edo semantikoa da.

<sup>3</sup> PI3: indikatiboko orainaldiaren 3. pertsona.

## II. kapitulua

*définition\_de\_nom*= [déterminant] "action de" infinitif\_verbal ["," ["de"] infinitif\_verbal].

*définition\_de\_nom*= "faculté de" infinitif\_verbal.

*définition\_de\_nom*= "manière de" infinitif\_verbal.

*définition\_de\_nom*= "état de ce qui est" (adjectif / participe\_verbal).

Adibideak:

**accentuation I 1:** action de accentuer

**sentiment I 1:** faculté de sentir

**fonctionnement I 1:** manière de fonctionner

**soulèvement I 1:** état de ce qui est soulevé

**Adjetibotiko izenak.** Izen-kontzeptua adjetibozko batekin erlazionaturik.

*définition\_de\_nom*= "état de ce qui est" (adjectif / participe\_verbal).

*définition\_de\_nom*= "qualité de" ["ce qui est"] adjectif.

*définition\_de\_nom*= "caractère de ce qui est" adjectif.

Adibideak:

**sécheresse I 1:** état de ce qui est sec

**blancheur I 1:** qualité de ce qui est blanc

**immortalité I 1:** qualité d'immortel (sic)

**sécheresse I 2:** caractère de ce qui est sec

**Adjetibotiko aditzak.** Aditz-kontzeptua adjetibozko batekin erlazionaturik, non adjetiboak jeneralean adierazten baitu aditzak denotaturiko egintzaren ondoriozko egoera edota kualitatea.

*définition\_de\_verbe*= "rendre" [graduateur] adjectif.

*définition\_de\_verbe*= "devenir" [graduateur] adjectif.

*définition\_de\_verbe*= "être" adjectif.

Adibideak:

**adoucir I 1:** rendre plus doux

**tiédir I 1:** rendre tiède

**tiédir I 2:** devenir tiède

**avoisiner I 1:** être voisin

**Izenetiko adjetiboak.** Adjetibo bat izenezko kontzeptu batekin erlazionaturik.

*définition\_de\_adjectif= ((( "relatif à" / "de" / "propre" ("à" / "de"))  
[déterminant] ) / "du" / "des" nom.*

*définition\_de\_adjectif= "conforme à" déterminant nom.*

Adibideak:

**agricole I 1:** relatif à l'agriculture

**ancestral I 1:** des ancêtres

**animal I 1:** propre de l'animal

**naturel I 1:** conforme à la nature

**Aditzetiko adjetiboak.** Adjetibo bat aditz-kontzeptu batekin erlazionaturik. Adjetiboak definitzeko maiz erabilitako formula da hau —baita izenentzat ere—, aditz-sintagmak denotaturiko egintza burutzen edota jasaten duenaren ezaugarri edo kualitate bat adieraziz.

*définition\_de\_adjectif= "qui" groupe\_verbal.*

Adibidea:

**fâcheux I 1:** qui fâche

**b) Erlazio morfologikorik gabeko kontzeptuen arteko erlazioak.<sup>1</sup>**

**Kasu-erlazioak.** Kasu-erlazio hauei esker kasuen betegarri ("filler") prototipikoak zein diren geratzen da bistan. Aurreko sailotako hainbat adibidetan ikusi ditugu kasu-erlazioak —agente eta objetu orokorrak, esate baterako—, nahiz eta adierazleak morfologikoki erlazonaturik dauzkaten kontzeptuen artean ezarritakoak izan. Beheko adibideetan *cautère* litzateke *brûler les chairs* egintzaren agente prototipikoa, eta *forer* egintza genuke *trépan* izeneko tresnaren helburu prototipikoa.

Adibideak:

*cautère I 1: ce qui brûle les chairs*

*trépan I 2: outil pour forer*

**c) Bestelako erlazioak (sintagmatikoak).**

**Atributiboa.** Aditz baten eta aditz-egintza jasaten duenaren egoera tipikoa adierazten duen adjetiboa erlazonaturik.

*définition\_de\_verbe= "être" adjectif.*

Adibidea:

*enrager I 1: être furieux*

**Gabeziakoa.**

*définition\_de\_adjectif= "sans" nom.*

Adibidea:

*banal I 1: sans originalité*

**Konformetasuna.**

*définition\_de\_adjectif= "conforme à" déterminant nom.*

Adibidea:

*normal I 2: conforme à la règle*

---

<sup>1</sup> Jakina, kontzeptuak denotatzen dituzten adierazleak dira elkarrekiko erlazio morfologikorik ez dutenak.



Eta honekin bukatuko dugu LPPLko definizioetan aurkitutako erlazioen zerrenda. Ondoko atalean azalduko ditugun analisi-prozesuetarako, erlazio horien azaleko gauzatzeen egituraz baliatu gara.

## II.5 Definizioen analisi sintaktiko-semantikoa.

Aurreko ataletan azaldutako prozesua, hiztegiko edukinen azterketa, ikuspuntu semantiko batetik egiteko, sortu eta jarraitu dugun metodologiaren lehen urratsa litzateke. Metodo honek LPPLko definizio-esaldien analisi sintaktiko-semantikoa egitera eraman gaitu, esaldion interpretazio lexiko-semantiko bat emango digun definizio-esaldien "parsinga" burutzerara. Horrela erdietsitako errepresentazio semantikoaren fruitu zuzena da HEBa. Analisi-prozedura bera zertan den azaldu aurretik, hiztegietako definizioen analisirako erabili izan diren metodo desberdinak aipatuko ditugu labur-labur.

Besteren artean N. Calzolarik (1984b), hiztegi-definizioez osaturiko corpus handien analisi-lanak ahotan hartuz, azpimarratzen du badirela, hiztegiko definizioen azpilengoiaren, sarri agertzen diren eskema lexikal edota sintaktikoak. Definizioen ezaugarri hori pattern-matching klasikoaren antzeko teknikez ustiatzeko aukeraz —kontzeptuarteko erlazioak erauzte helburu— hitz egiten du autore horrek.

(Calzolari & Picchi, 88)n berriz, lineango hiztegi bateko —testu forman— definizioak analizatzea helburu duen ikerketa-metodo bat zehazteko jarraituriko bidea azaltzen dute egileek. Metodo horrek definizio-metalengoiari dagozkion hitzak sailkatuzera eraman ditu. Hitz horien artean Kategoria Kontzeptualak deitzen dituztenak —*instrument, set, part, science, human*, eta abarren gisako hitzak— eta Erlazioak —*used, produced, in-the-form*, eta beste— aurkitu dituzte, horretarako irizpide "kuantitatiboak zein intuitiboak" erabiliz. Sailkapenaren ondorioz lortzen dituzten "etiketez" baliatuko dira gero, lema eta bestelako hitz-formekin batera, bilaketa eta analisirako patroietan. Erregela edo patroiak batetik, eta metalengoiari dagozkion zerak bestetik, aparte erazagutuak daudelarik, beren pattern-matching-ezko prozedura horrek dituen abantailen artean gehigarritasuna, malgutasuna, eramangarritasuna eta frogagarritasuna aipatzen dituzte.

Wilks eta bere taldekoen lanetan (Wilks *et al.*, 89; 90), *Longman Dictionary of Contemporary English* (LDOCE) hiztegiko kategoria "irekiei" —izenak, adjetiboak, aditzak eta adberbioak— dagozkien definizioen analisi sintaktikoa zertu dute, 100 erregela dituen testuingururik gabeko gramatika bat erabiliz. Analizatzailerak erabilitako lexikoa LDOCEko definizioetako oinarrizko hiztegitik hartua da. Analisia behetik gora egiten da —"chart" dun bottom-up eran—, ezkerretik eskuinerantz, beheerantzko filtroaz eta "katearik luzeena hobesteko" (osagai kopuru txikienekoa) estrategia baten arabera. Autoreen esanetan (Wilks

## II. kapitulua

*et al.*, 90), 1200 definizioko lagina analizaturik, erdietsitako arrakasta maila %95koa izan da. Analisiaren emaitza zuhaitz erako perpaus-egiturak dira, ondoren semantikoki interpretatuko dituztenak. Interpretazio horretarako pattern-matchingeko eta inferentzi teknikak erabiliko dituzte. Kontzeptuen adierazpide semantikoa frame erako egituretan oinarritzen da, eta dirudenez, definitutako kontzeptua eta bere genusaren arteko erlazioa besterik ez dute oraingoz errepresentatzen, differentiaren parteari dagokion informazio espezifikoaren erauzketa azterkizun utziz.

Lehenago ere esana dugunez, definizioak analizatzeko hartu dugun bidea H. Alshawik (1987; 89) erabilitakoari jarraitzen zaio hertsiki.

Analisi-mekanismo honek pattern-matching hierarkiko partzian du funtsa. Analizatzaileak patroï sintagmatikoz osatutako hierarkia batzuk erabiliko ditu; hierarkia horiek zuhaitz egitura izango dute, eta analizatu beharreko sarreren kategoria bakoitzeko zuhaitz bana eratu beharko da. Patroï bakoitza erregela bati dagokio, eta erregela osatzen duten elementuak hauexek dira: erregela-identifikadorea, patroï sintagmatikoa eta erregela umeen idenifikadoreen lista (erregelako patroïa baino "espezifikoagoak" direnak edukiko dituzten erregelak). Horretara, patroïak zuhaitzean beren espezifikotasun mailaren arabera daude kokatuak: espezifikoagoak orokorragoen azpian alegia.

### II.5.1 Alshawiren analizatzailearen implementazioa.

H. Alshawik deskribatu analizatzailearen (1987; 89) implementazio berri bat garatu dugu proiektuan, geure premia berezietara egokituz (Agirre *et al.*, 91; 92).

Analisi teknika honako hau da: erro-adabegiko erregelatik hasita —hierarkiako maila gorena— patroï-hierarkia sakoneran korritzen da, analizakizun den definizioa adabegi bakoitzeko patroïarekin parekatzen saiatuz. Parekatze edo matching-ak arrakasta duelarik, hierarkian erregela horren ume diren erregela bakoitzeko patroïekiko parekatzea lortzen saiatuko da analizatzailea ondoren. Lehen esan dugun bezala, erregela ume horietan patroï espezifikoagoak daude. Matching-prozedura hori behin eta berriro errepikatuko da —errekurtsiboki— harik eta analizakizun den definizioarentzat parekatzerik espezifikoena lortu arte. Hierarkiako adar bakoitzean, arrakasta izan dutenen artean patroïrik espezifikoena bakarrik hautatuko da.

Horrela iritsiko gara definizio bakoitzari analisi bat edo gehiago esleitzera. Analisi horien adierazpide, patroïko itemen eta item horiei dagozkien definizio-hitzen artean lortutako loturak izango ditugu.

Patroïen espresioan erabil daitezke nahiz kategoria sintaktiko bakunak —datu-basean dauden kategoria erreal konkretuen arabera beste nonbait definituak (ikus beherago)— nahiz

konplexuak (izen-sintagmak, aditz-sintagmak, etab.), zeinen egitura ere aparte definituko baita patroï lagungarri deiturikoak erabiliz. Orobat, definizioetako hitzen forma kanonikoekin parekatu beharreko karaktere-katea literalak ere erabiltzen ahal dira.

Patroïon sintaxia, pattern-matching-ean erabili ohi diren metakaraktere tipikoez hornitua, Alshawik berak definitutakoaren (1987; 89) antzekoa da oso eta, nahi izanez gero, txosten honen A eranskinean edota (Agirre *et al.*, 91)n azter daiteke, bertan ongi deskribaturik baitago.

Hierarkiako patroï sintagmatiko bakoitzarentzat erregela semantiko bat definitu da; erregela horrek bere patroïarekin parekatutako definizioari egokituko zaion —HEBa eraikitzeke garaian— egitura semantikoa zehazten du. Jakina, patroï bakoitzari dagokion erregela semantiko hori aukeratu ahal izateko, derrigorrezkoa izango da patroïaren umeekin arrakastarik lortu ez izana. Horrela, hierarkiari esker, definizioetako osagairik inportanteenak zehazteko aukera izango dugu, izan ere osagai inportanteenei patroï orokorrenak esleituko baitzaizkio normalean. Honela lortuko dira analisi ez-osoak —onargarriak, hala ere—, analisi zehatzagoak erdietsi ezin direnean.

Beraz, mekanismoak bereziki egokia dirudi hiztegi-definizioetako analisi partzialak lortu eta erabili ahal izateko.

Gure inplementazioan analizatzaileak baditu erabilinguruari, batik bat, zor zaion zenbait bereizgarri, hau da, hiztegi-informazioaren biltegi den HDBarekiko integrazioari alegia. Analizatzailea —COMMON LISPez idatzi eta LPPLa daukan HDBa zegoen makina berean ezarria— HDBarekin integratu zen, analizatu beharreko definizio-esaldiak datu-basetik zuzenean eskuratu eta parsing-aren emaitza desberdinak ere HDBan bertan erregistratu ahal izateko<sup>1</sup>.

Emaitzak gordetzerako analizatzaileak analisi desberdinak hierarkiako erregelen ordenaren arabera ordenatu egingo ditu, erregela berari dagozkion analisisien artean analizatutako testuaren zati luzeena hartzen dutenak lehenetsiz: honek ematen digu definizio baten analisiak sailkatu ahal izateko kontrol-mekanismo bat, hala-holakoa bada ere.

HEBak berak egin du lexikoarena analizatzailearentzat, definizioetako hitzen kategoria emanez aurretiko etiketarik ez zuten kasuetan. Horrela, kategoria desberdinetako sarrera bat baino gehiago zegokiolako hitz baten desanbiguazioa ezinezkoa izan eta etiketatzeke zirauen kasuetan, analizatzaileak datu-basetik bertatik —analizatzailearen benetako lexiko-sistema— eskuratu ditu kategoria horiek, parekatzea guztiekin saiatuz.

---

<sup>1</sup> Cz idatziko prozeduren bitartez lortzen da datu-basea eta analizatzailearen arteko interfazea, bai datuak eskuratzeko eta bai emaitzak gordetzeko.

## II. kapitulua

Sarrerren kategoria hiztegian era askotara ematen denez, bi kategori "mota" bereizi behar izan ditugu: *kategoria fisikoa* esango diogu LPPLk sarrerari ematen dion kategoria errealarari ('vt.', 'vi.', 'n.', 'f.', etab.); beraz, analizatzaileak eta guk —patroien eraketan— erabilitako kategoriak, *kategoria logiko* edo deituko ditugu. Hortaz, kategoria logiko bakoitzaren pean kategoria fisiko "zehatzagoen" multzo bat egongo da. II.7 irudian ikus daitezke kategoria logiko batzuk eta beroriei dagozkien kategoria fisikoen multzoak.

KATEGORIA LOGIKOAK	KATEGORIA FISIKOAK			
NOM	'f.'	'f. inv.'	'f. pl.'	'fpl.'
	'f. ou m.'	'm. et f.'	'm.'	'm. inv.'
	'm. inv. et adv.'	'mpl.'	'n.'	'npl.'
	'n. m.'	'pl.'	'QUAL-NOM'	'NOM'
ARTICLE	'art.'	'art. indéf.'		

**II.7 irudia.-** *Kategoria logiko* eta *fisikoen* arteko korrespondentziaren zenbait adibide.

Alshawik deskribatutako analizatzailearekiko, gure inplementazioak patroiko item sintaktikoei predikatu bat esleitzeko posibilitatea ere eskaintzen du. Horri esker, hierarkigileak patroia parekatzea murriztarazteko aukera izango du, murrizketa hori derrigor sintaktikoa izan behar ez duen informazioaren arabera egin daitekeelarik. Predikatudun item batekiko parekatzea arrakastatsua izan dadin, predikatuak itzultitako balioak egiazkoa izan behar du. Hala ere, esan behar da LPPLko definizioak analizatzerakoan ez garela aukera honetaz baliatu.

Inplementatutako analizatzailearekin LPPLko definizio ez diren bestelako esaldien analisi sintaktikoa ere egiten ahal da. Eginkizun horretan HDBak lexikoarena egingo luke eta, noski, definiziozko ez diren esaldi horientzako patro-i-hierarkiak eratu beharko lirake aldez aurretik. Dena den, bistakoa da analizatzaileak duen akatsik handienetakoa, analizatzaile orokor bezala erabili ahal izateko behintzat, analisi morfologikorik ez eskaintzea dela.

### II.5.2 Patroi-hierarkiak.

Alshawik (1987; 89) bere analizatzailearen aplikazio baten emaitzak ere ematen ditu, LDOCETik zoriz aukeratutako 500 definiziozko lagina harturik egindakoarenak hain zuzen. Lagungarriak barne 90 patro-i definituta, analizatzaileak definizio horietarik %77tan gunea ongi ezagutu zuela esaten du autore honek. Gainera, %61etan bestelako informazioa ere eskuratu zuela, zuzentzat jotako informazioa kasu horien %88tan. Alabaina, autoreak

aitortzen du definitutako patroï-sail horrekin nekez lortzen dela definizio osoa analizatzea (definizioeko hitz guztiak patroïko itemekin parekatzea). Bere ustez, 400 patroï edo definitu beharko lirateke analizatzailearen emaitzak onargarriak izateko (agi danez, LDOCEko definizioen analisi finagoa egiteko asmoa luke autore honek, halako patroï kopurua definitu eta gero).

Gure aplikazioari dagokionez, definizioen analisiaren azken jomuga definizio-esaldietan erabilitako lengoaia lexikografikoa ezaugarritzea dela azpimarratu behar, ezaugarritze horren nondik norakoak hauexek liratekeelarik:

- definizioetan agertutako *egitura sintaktiko usuenei* dagokien egitura semantikoa zehaztu;
- kontzeptuen arteko *erlatore bereziak* izan daitezkeen elementuak dituzten definizioen egitura semantikoa zehaztu, nahiz eta elementu horien maiztasun erlatiboa ez oso handia izan;
- erabat analiza ezin daitezkeen definizioetatik *informazio partziala* —baina zuzena—eskuratu.

Esan bezala, jarraitutako metodoak patroï-hierarkiak eratzeraz eraman gaitu, eta horretarako irizpide semantikoetan oinarritu gara. Esan daiteke aurreko ataletan azaldutako lan guzti horien emaitza direla, besteak beste, definizioen azken analisi sintaktiko-semantikoan erabiltzeko eratu ditugun hiru patroï-hierarkiak.

Izen-definizioak analizatzeko prestatu hierarkian 65 patroï daude, 49 aditzentzakoan eta 45 adjetiboenean. Patroï horien artean badira oso patroï orokorrak, izen-definizioentzako izen-sintagmadun patroïak edota aditz-sintagmadunak aditzenentzat, esaterako, baina baita § II.4.5en azaldutako azterketa enpirikoetatik eratorritako patroï espezifikoak ere.

Hierarkiotan definitu ditugun patroïen nolabaiteko irudia emateko, esan dezagun izenen definizioei begira prestatutako 37 patroï, aditzentzat egindako 21 eta adjetiboentzat prestatutako beste 37 patroï *espezifikoak* ditugula. Patroï espezifiko horien artean badira, besteak beste, ("*partie*" "*de*" (GNOMINAL I) (M&&)) bezalako patroïak izenen definizioentzat, ("*commencer*" "*à*" (GVERBAL I)) aditzentzat, edo ("*sans*" (GNOMINAL I)) adjetiboentzat. Espezifikotzat hartu ditugu —berehala azalduko ditugun emaitzen aurkezpenarako— definizio sinonimikoak ezagutzera zuzendutako patroïak ere, LPPL bezalako hiztegi txiki batean hainbesteko garrantzia, normala denez, dutenak.

Hierarkia horiekin batera 15 patroï lagungarri ere definitu ditugu, egitura sintagmatikoak deskribatzeko batik bat. Patroï lagungarri hauek oso baliagarriak dira definizioetako erlazio

## II. kapitulua

espezifikoko berberaren gauzatze diren egitura metalinguistiko desberdinak etiketa bakar baten pean biltzeko ere (Agirre *et al.*, 91).

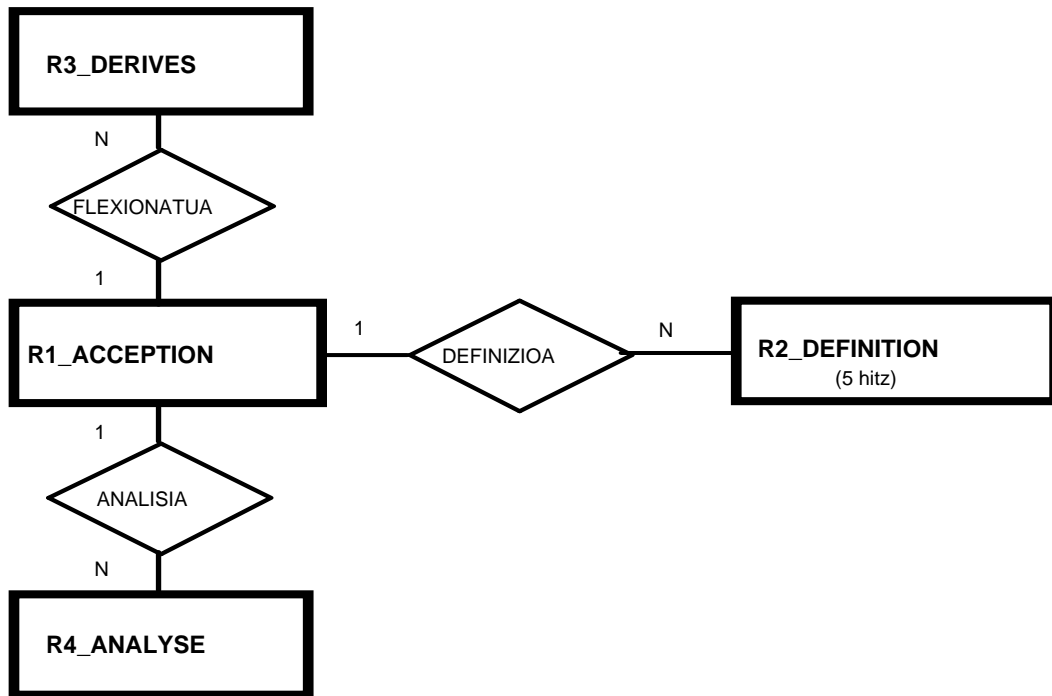
Gure helburua, LPPLko definizio-esaldiak labur samarrak zirela kontuan hartuz, analisi-prozesu honetatik "ahalik eta gehiena" erdiestea izan da. Beraz, pattern-matching-eko zernahitarako elementua —*M&&* metakarakterea— ahalik eta gutxien erabiltzen saiatu gara, ahalara patroien atzeko aldean soilik. Egindako analisi partzialen zuzentasuna oraindik erabat balioztatzeke badugu ere, esan dezakegu zernahitarako elementu hori patroietan axola gutxiz erabiltzeak, eta bereziki aukerazko itemen ondotik, eraman gintzakeela analisi sintaktiko —eta beraz, interpretazio semantiko— okerretara.

Irakurleak jo dezake berriro txostenaren A eranskinera prestatutako patroien zerrenda osoa —patroi lagungarriak barne— aztertu nahi izanez gero.

### II.5.3 LPPLko definizioen analisiaren emaitzak.

Sortutako patroi-hierarkiak LPPLko izen, aditz eta adjetiboen definizio guztiak analizatzeko erabili dira. Azpialde honetan, analisi horren emaitzak HDBan bertan erregistratzeko nola egin dugun kontatuko dugu lehenik, analizatzailea analizagai izan diren definizioei aplikatzeak izan dituen gorabeherak ere azalduz. Halaber, analizatzailearen aplikazio konkretu honen emaitzak emango ditugu.

Analisiaren emaitzak erregistratzeko taula berri bat erantsi genuen HDBaren eskema kontzeptualean: R4\_ANALYSE. Adiera-definizio bati dagozkion analisi guztiak ordenatu —ikus § II.7.1en analisi-anbiguotasunaren gainean dioguna— eta datu-baseko taula berri horretan geratu dira grabaturik. II.8 irudian ikus daiteke HDBaren eskema kontzeptual berria nola geratu den.



**II.8 irudia.-** HDBaren eskema kontzeptual berria, definizio-analisiaren emaitzak erregistratzeko prestatua.

Ikus dezagun, adibidez, *spatule 11* sarreraren definizioa —*sorte de cuiller plate*— analizatutakoan sortzen den emaitza. Definizio hori RN110532 patroiarekin — (*N"sorte""de"(GNOMINAL1)(M&&)*)— parekatuko da, eta, analisiaren emaitza gisa, ondoko egitura sortu eta grabatuko da HDBan:

(RN110532 ((GNOMINAL 1 2 2 ((NOM 1 2 1) (QUALIFICATIF 1 3 1))))))

Emaitza-egitura hori lista bat da, non, analisi-patroiaren identifikadore-zenbakiarekin batera, analizatu definizio-testuari dagokion egitura sintaktikoaren (edo sintaktikoen, anbiguetaterik dagoen kasuetan) errepresentazio bat ere ageri baita. Errepresentazio hori analizatutako definizioeko hitzen eta patroiko itemen artean ezarritako loturen bitartez eratuko da. Txosten honen B eranskinean aurkituko du irakurleak errepresentazio horren sintaxia eta berorri buruzko azalpen zehatzagorik.

LPPLko definizioen analisiaren emaitza zehatzak ematen hasita, lehen-lehenik esan behar da analisi partzialeko mekanismoa izanagatik izen-definizioen %57.76a, aditzen %79.8a eta adjetiboen %69.04a erabat analizatuak, hau da definizioeko hitz guztiak parekatuak zernahitarako elementua (*M&&*) erabiltzeke, suertatu direla aplikazio honetan. Zernahitarako elementu hori —definizio-esaldiko zero edo hitz gehiagorekin parekatzen dena— izen-definizioei egokitutako patroietan erabili da gehienbat.

## II. kapitulua

	IZEN-DEF.		ADITZ-DEF.		ADJET.-DEF.	
		%		%		%
<b>Porrota</b>	1086	7.9	105	2.00	527	16.35
<b>A taldea</b>	4657	33.88	2625	50.02	2370	73.51
<b>B taldea</b>	8001	58.22	2517	47.97	327	10.14
<b>Guztira</b>	<b>13744</b>		<b>5247</b>		<b>3224</b>	

### II.9 irudia.- Analisiaren emaitzak.

Egindako analisiaren emaitzak daude II.9 irudiko taulan. Hor ageri diren taldeak patroï mota desberdinen arabera egin dira. Parsing teknika partziala denez, arrakastatzat hartu ditugu analizatu esaldiaren hasieran egitura sintagmatiko bat bederen ezagutu deneko kasuak, egitura horrek definitutako kontzeptuaren genusa edo hiperonimoa edukiko duelakoan. Hori ez da izango, noski, lehenago aipatutako eraikuntza espezifikoen kasua; izan ere, era horretako eraikuntzei patroï espezifikoak egokitu baitzaizkie hierarkietan, beren itzulpen semantikoa ere modu berezi batez egingo delarik.

Ikus daitekeenez porrot maila nahiko baxua da, eta bereziki aditzen definizioen kasuan. Horrek adierazten du, gure iritziz, aditz-sarrereri dagozkien definizioak idazteko lexikografoak erabilitako egitura sintaktikoen multzoa murriztagoa dela beste kategorietakoak idazteko erabilitakoena baino. Porrot maila apal hori azalduko lukeen beste arrazoi bat litzateke aditz-definizioak laburragoak direla, batez beste, izenenak baino.

A taldekotzat hartu ditugu metalengoaia mailako patroïez ezagututako definizio-esaldiak, hau da, definizio-formula espezifikoak betetzen dituztenak. Gorago esan bezala, definizio sinonimikoak ere A talde honetan sartu ditugu, metalengoaiazko formulez eratutako definiziotzat hartu ditugunez: hitz bakarreko definizioak, sarreraren kategoria bereko hitz anitzez —komaz bereizirik— osatutako definizioak, etab. Lehen talde hau bereziki inportantea da guretzat, definizio horiei esleituko diegun egitura semantikoa patroï horien baitan dautzan erlatore berezien funtzioan eratuko baitugu zuzenean. Testuinguru honetan azpimarratu behar da adjetiboen kasuan zein maila altua lortu den A talde honetarako. Horrek erakusten du adjetiboak definitzeko formula sinonimikoak sarritan erabili ohi direla (%33.81), baina baita definizio-formula finkatuen erabilera adjetiboen kasuan bestetan baino ohizkoagoa dela: adjetiboen definizioen %39.7a mota honetako 27 patroïrekin analizatuak suertatu dira. Horrela iristen gara adjetibo-definizioen %73.51a (2370 definizio) A talde honetan sartzera.



B taldeko definizioak berriz, definizio aristotelikoak edo direla esan genezake, hau da, "genus et differentia specifica" tankeran egindakoak. Kasu horietan ateratzen zaio probetxu gehienbat analisi partzialaren posibilitateari. Analisi partzialez probetxatzeko ideia sinplea da oso: batzuetan differentiari dagokion definizio-partea "galdutzat" jota ere, genusari dagokiona behintzat beti eskuratzen dugula pentsatzean datza. Talde honetan sailkatutako definizio guztiei esleituko diegun egitura semantikoan, definizioaren hasierako egitura sintaktikoaren gunea —kasuan kasuko izen-, aditz- edo adjetibo-sintagma— definiendumaren genus edo hiponimotzat hartuko da.

#### II.5.4 Egitura semantikoen eraketa: HEBaren sorkuntza.

Analisi sintaktikoan erabilitako patroien itzulpen semantikoa egiteko, patroï bakoitzari *Egitura Semantikoa Eratzeko Erregela* (ESEE) bat esleitu zaio.

ESEE horien egitekoa, definizioa analizatu zaion kontzeptu bati errepresentazio semantikoa ematean datza. Esan bezala, analizatutako definizioa eta lortutako emaitza loturik daude HDBan; analisisa errepresentatzeko patroïaren erreferentzia bat eta definizio-hitz edota -egituren eta patroï-itemen artean ezarri loturen funtzioan adierazitako espresio bat erabiltzen dira. HEBa eraikitzerakoan, definizio-analisi bikote bakoitza analisisiko patroïari —erregistratutako analisi-espresioan ageri dena, HDBko R4-ANALYSE taulan— dagokion ESEEaren bitartez prozesatuko da, egitura semantikoa lortzearen. Egitura semantiko hori HEBan unitate berriak sortuz eta unitateen arteko loturak —nahiz berri horien berorien artean nahiz berrion eta lehendik daudenen artean— ezarriz gauzatuko da.

II.10 irudian ikus daiteke patroï sintaktiko bat —izenen definizioetarako prestatu hierarkiako RN110532 patroïa— eta dagokion egitura semantikoa eratzeko erregela. Patroï horrek ezagutzen dituen definizio-egiturak '*sorte de*' hitzekin hasi eta ondotik izen-sintagma bat dutenak izango dira, izen-sintagmaren ondoren *M&&* zernahitarako elementuarekin parekatuko litzatekeen edozein hitz-sekuentzia etor daitekeela. Patroï honen bitartez ezaguturiko definizioentzat sortuko den egitura semantikoa bere ESEEn adierazten da. ESEE horietan bi parte bereizten dira argi eta garbi: lehen partea, *unitateen sorkuntza*, non sortu beharreko unitateak zehazten baitira, hiztegiko kontzeptuak errepresentatzeko erabiliko den frame-sareko (THESAURUS) nodoen funtzioan; bigarren partean *balio-esleipenak* adierazten dira, sortutako unitateon atribuetan egin beharreko balio-asignazioak direla medio unitateen artean loturak ezarriz.

Irudiko adibide konkretuan hiru kontzeptu-unitateren sorkuntza adierazten da (gogoratu II.5.3n ikusi berri dugun analisi-emaitzaren adibidea): DEFINIENDUM unitatea, analizaturiko definizioak definitzen duen kontzeptuaren adierazle, CONF1, *GNOMINAL I* egitura sintaktikoaren errepresentazio gisa sortutako konfigurazio-unitatea, eta REF1,

## II. kapitulua

patroiko (*GNOMINAL 1 ((QUALIFICATIF 1))*) itemari dagokion adjetibo-kontzeptua errepresentatuko duena. CONF1 hierarkikoki bere gunekontzeptuaren —patroiko (*GNOMINAL 1 ((NOM 1))*) itemarekin parekatutakoaren— menpeko den izen-konfigurazio bezala sortzen da —NOMINALES klasearen barneko (ikus III. kapituluan)—. Konfigurazio-unitatea eta bere guneari dagokion kontzeptu-unitatearen arteko lotura hierarkiko horiek automatikoki ezartzen dira konfigurazio-unitateok sortzean. REF1 unitatea QUALITES klasearen azpiklase gisa sortuko da, dagokion forma kanikoaren deiturapean.

Balio-esleipenaren parteak, adibidean eta besteak beste, DEF-SORTED erlazio espezifiko definitu kontzeptua —DEFINIENDUM— eta CONF1 definienseko izen-sintagmaren errepresentazio den konfigurazio-unitatearen artean ezartzen dela adierazten du. Kontzeptuarteko erlazio definitorio hori DEF-SORTED atributu edo "slotaren" bitartez gauzaten da, atributu horren sorburua '*sorte de*' espresioan datzala. Halaber, CONF1 eta REF1 unitateen artean CHARACTERISTIQUE erlazio baten ezarpena ere adierazten du ESEEk bigarren parte honek.

```
;;; RN110532 patroia (izenen definizioak)
```

```
;;; (N "sorte" "de" (GNOMINAL 1) (M&&))
```

```
;;; Patroiari dagokion ESEEA:
```

```
`(RN110532
```

```
;;; Sortu beharreko unitateak:
```

```
( (DEFINIENDUM)  
  (CONF1 CONF T ,(can '(GNOMINAL 1 ((NOM 1)))) (NOMINALES))  
  (REF1 REF ,(can '(GNOMINAL 1 ((QUALIFICATIF 1)))) (QUALITES)))
```

```
;;; Balio-esleipenak:
```

```
( (,DEFINIENDUM DEF-SORTED ,CONF1)  
  (,CONF1 CHARACTERISTIQUE ,REF1))
```

```
)
```

**II.10 irudia.-** Patroi sintaktiko baten adibide sinplifikatua eta dagokion Egitura Semantikoa Eratzeko Erregela.

Ikus daitekeenez, LISPeke backquote delako sintaxia erabiltzen dugu ESEE horien adierazpidean. HEBaren hasierako eraikuntzaren fasean, espresio horiek<sup>1</sup> erabiliko ditugu

---

<sup>1</sup> Izan ere, espresio horien izaera erazagutzaile-prozedurala dela esan dezakegu: irudian ikus daitekeenez, ESEE horietan LISP funtzio-deiak egin daitezke zuzenean. Esate baterako, hortxe dugu *can* funtzioari eginiko deia (funtzio hau patroiko item bati aplikatzen zaio, eta berorri dagokion hitzaren forma kanonikoa itzultzen du).

zuzen-zuzenean THESAURUSeko unitateak sortu eta berorien artean ezarri beharreko loturak ezartzeko.

ESEEn syntaxian literalak ere onartzen dira. Literal horien bitartez adieraz daitezke, besteak beste, ESEEarri esker sortarazitako kontzeptu baten hiperonimo konkretuak, edota fazetak, errepresentazio semantikorako hautatu dugun ereduko atributuak hobeki zehazteko erabili ohi direnak (ikus III. kapituluan). Esan dezagun, adibide bat ematearren, fazetok analisi-fase honetan erabat desanbiguaezenak diren erlazio semantikoen zehaztapenean erabiltzen direla. HEBaren hasierako eraikuntzaren fasean maila sintaktikoan edo geratuko diren erlazioen kasua litzateke; erlazio horiek ESEEn hobeto zehazten ahal dira, horretarako ziurtasun-baliodun fazetak erabiliz (sakonean egon daite(z)keen erlazio semantiko(a) zein d(ir)en adieraziz)<sup>1</sup>. Txosten honen C eranskinean dago HEBa eraikitzeke erabili ditugun ESEE guztien zerrenda beren syntaxiaren zehaztapenarekin.

Esan dezagun, bukatzeko, HEBaren eraikuntza ESEE horien interpretazioan datzala — HEBaren programa "eraikitzaileak" burutzen duen interpretazioa—, eta interpretazio hori kasuan kasuko definizioari, bere egituraren errepresentazio den analisia dela medio (R4-ANALYSEn gordea), aplikatuko zaiola.

## II.6 Hiztegi Ezagutza-Basaren prototipoaren eraikuntza.

### II.6.1 LPPLtik aukeratutako azpimultzoa: Le Français Fondamental.

HEBaren prototipoa eraikitzeke garaian LPPLko sarreren azpimultzo bat aukeratu beharra egon zen, LPPLa handitxo baitzen bertan ezarri nahi genituen eragiketak ezarri ahal izateko. Bestalde, prototipo bat egiteko nahikoa iruditu zitzaigun azpimultzo bat hartzea. Dударik gabe, hiztegi oso bat hartzeak dauzkan abantailen artean multzo itxia lortzearena aipatu behar da, hau da, definizioetan parte hartzen duen hitz oro (bere forma kanonikoa) agertuko dela hiztegiaren sarrera gisa definitua. Zaila da benetan —ezinezkoa ez bada— hori betetzen duen hiztegirik aurkitzea, eta jakina, ez da LPPLren kasua inondik ere.

Alabaina, azpimarratu beharra dago LPPLren gainean egindako azterketa eta analisi guztiak hiztegi osoaren gainean burutuak izan direla —izen, adjetibo eta aditzen definizioak— eta, hori dela eta, gure ustez, ezagutzaren errepresentazioari buruz hartutako erabakiak ez dira, inola ere, gutxiestekoak prototipoaren izaria txikiagoa izate honengatik.

---

<sup>1</sup> Adibidez, *pour* preposizioak ezarritako erlazioa —sintaktikoa— zehatzagotu egin daiteke *OBJECTIF* izeneko fazetarekin, ziurtasun-balio bat emanez. HEBaren aberasketa-fasean informazio probabilitistiko horiek erabiliko dira, horrela ezarritako erlazioen desanbiguaketa burutzearren.

## II. kapitulua

Azpimultzoaren hautaketarako erabili irizpideak honako hauek dira:

- a) *Gramatikala*: "HEBaren sarrera-kontzeptu" gisa izenak, adjetiboak edo aditzak besterik ez ditugu hartu (HEBan definizio bat edukiko duten kontzeptuak, egindako analisi sintaktiko-semantikoaren emaitza)<sup>1</sup>.
- b) *Lexikografikoa*: zein hitz hartu pentsatzen hasita, *Le Français Fondamental* (G. Gougenheim. Paris: Didier, 1958) hiztegiko sarrerak —goian aipatu kategorietakoak— aukeratzea erabaki genuen. Horrela, prototipoaren eraikuntzan *Le Français Fondamental* (FF) hiztegiko sarreren zerrenda erabili dugu LPPLko sarreren bahe gisa: LPPLko izen-, aditz- eta adjetibo-sarrerei dagozkien adiera guztiek osatzen dute, hasiera batean, HEBaren prototipoa, beti ere FFn sarrera gisa agertzen baldin badira.

Hasieran hirugarren irizpide bat ezartzea pentsatu genuen: *egiturari* dagokiona. Ez genituen onartuko —HEBaren prototiporako— analisi sintaktiko semantikoan arrakastarik izan ez duten definizioak (ikus § II.5.2). Hala ere, irizpide hori alde batera utzi eta definizio horiei dagozkien adierak ere onartu egin ditugu prototipoan. Horrela, definizio-analisan gertatutako porrotak ez du esan nahi adiera horiek baztertzen ditugunik, baizik eta HEBra "defektiboki pasatzen" ditugula, hau da, definizio-errepresentaziorik gabe. Adiera horiei dagozkien nodoak, HEBan duten hasierako definizio-errepresentazioaren aldetik (ikus III. kapituluan), "hutsak" daudela esan dezakegu.

Prototipoaren neurri zehatza, aurreko irizpideon arabera hautatutako adieren kopuruaren aldetik, honako hau da:

- 1461 izen-adiera.
- 844 aditz-adiera.
- 302 adjetibo-adiera.

### II.6.2 Prototipoaren erabidea.

Hemen beheko II.11 irudian ikus daiteke eskematikoki HEBaren hasierako erabidea zernolakoa izan den (II.1 irudiko 2. urratsari dagokio).

---

<sup>1</sup> Izango dira bertan beste kategorietako kontzeptuak ere —adberbioak, gehienbat—, baina beste kontzeptuekin dituzten erlazioen arabera errepresentatuak, hau da, HEBan "hasierako definiziorik" gabeak. Kontzeptu horiek beste kontzeptu batzuen definizio analizatueta agertuak izango dira eta, analisi horien ondorioz, patroi-itemekin lotuak. Kontzeptu analizatu horien errepresentazioaren parte gisa azalduko dira, beraz, HEBan.

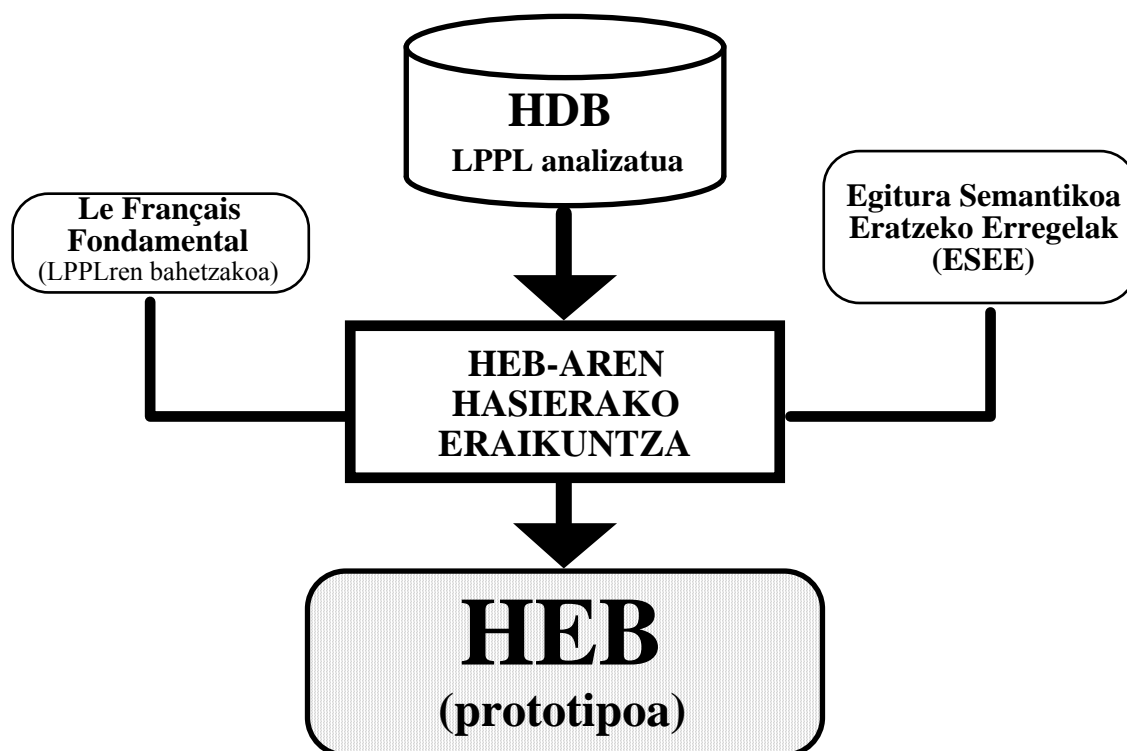
Hasteko esan behar da, definizioak analizatu ondoren eta HEBaren hasierako eraikuntza hori burutu duen programa abiarazi baino lehen, "errepresentazio-aldaketa" hori nozitu behar zuten HDBko datuen gainean (FFko sarrerei dagozkienak) zenbait prozesu —eskuzko zein erdiautomatiko— burutu dela. Prozesu horiek honakoak izan dira:

- a) Maiz erabilitako zenbait hitzen desanbiguaketa lexikala: zenbait hitz-agerpen eskuz desanbiguatu ditugu, homografo- eta adiera-zenbakiak erantsiz. Bistakoa da hori egitea ez dela samurra kasu guztietan, eta beraz, dudazko kasuetan anbiguetatea hor geratu da.
- b) Zenbait forma kanonikoren zuzenketa: jakina denez, hasiera batean hitz-agerpenei esleitutako forma kanonikoa agerpenaren berdina zen erabat, anbiguetate morfologikoren bat edo maizeneko agerpenen kasuan izan ezik, kasu horiek aurreko urrats batean zuzenduak izan baitziren (ikus § II.4.2). Behin definizioak analizatu eta gero, oker zirauen zenbait forma kanoniko zuzentzeko egokiera hobea zen, definizioei esleitu patroiek berek erakutsi baitigute errakuntza horiek non dauden edo non egon daitezkeen<sup>1</sup>.

HDBtik bertatik (ikus II.11 irudia) eskuratzen da HEBko parte izango diren sarrerei dagozkien informazioa, bahetzat erabiliko dugun sarrera-listaz baliatuz (FF). Hasierako eraikuntza hau burutzen duen programak ESEEak ere hartzen ditu sarreratzat; ESEE horiek zehazten dute HDBtik datorren informazioarekin HEBan eraiki beharreko egitura semantikoak nolakoa izan behar duen. HDBko definizioak analizaturik eta analisi-emaitez bertan grabatuta daudela, prozesua zuzena da: programak banan-banan hartzen du sarrerei dagokien informazioa eta, sarreraren analisisian ageri den erregela-identifikadorearen arabera, dagokion ESEEA aplikatzen dio. Horrela doaz sortuz HEBa osatuko duten unitateak, beren arteko loturak ere ezarriz.

---

<sup>1</sup> ("*qui*" (*M+ VERBE/PI3 1*)) bezalako patroik batekin pareka zitezkeen —definizio-eskemak aztertzean erabilitako irizpide heuristikoak direla medio (ikus § II.4.3)— bigarren hitzaren forma kanoniko gisa aditzaren forma kanonikoa gabe forma flexioduna (indikatiaboko orainaldiko hirugarren pertsona) duen definizio batekin. Kasu horretan, patroik hori definizioarekin lotu duen analisi-emaitez berak erakusten digu forma kanoniko horretan egon daitekeela errore bat.



**II.11 irudia.-** HEBaren eraikuntzaren hasierako urratsa.

Prozesu horren emaitza bezala, HEBaren prototipoaren lehen bertsio bat lortu dugu. Bertsio honek nozitu beharko ditu oraindik, hain zuzen ere HEBaren egituraketari esker posible izan diren, aldaketak. Horri buruz jakin nahi dezanak jo beza III. kapitulura, HEBaren aberasketa dinamikoa jorratzen den atalera<sup>1</sup>.

II.12 irudian dagoen adibidea —*agricole 11* hiztegi-sarrera, LISP eran— HDBtik eskuratzen den irteeraren itxura nolakoa den erakusteko emana da. Ikus daitekeenez, irteera-egitura hori kabiitze maila anitzeko lista asoziatibo bat da, eta LPPLko jatorrizko —edo birmoldaturiko— informazioa dauka bere baitan.

---

<sup>1</sup> Automatikoki exekutatzen diren prozedurak —iratxoak— ere presente egon dira HEBaren prototipoaren lehen bertsio hau eraikitzean, kontzeptuarterko zenbait erlazio automatikoki ezarriz (ikus III.2 atala). Horrela, lehen bertsio hau HDBtik etorri ezagutzarekiko nolabait "aberastua" dela esan daiteke.

```
( (CLEF "agricole I 1")
(CATEGORIE "adj.") (GROUPE-CATEGORIEL "ADJECTIF") (USAGE "")
(TYPE-DEF "A") (EXEMPLE "machine -")
(DEFINITION
  ( (MOT-DEF "relatif") (FORME-CANONIQUE "relatif I ?")
    (CAT-MOT-DEF "adj.") (ALT-MOT-DEF ""))
  ( (MOT-DEF "a`") (FORME-CANONIQUE "a` I 1")
    (CAT-MOT-DEF "pre'p.") (ALT-MOT-DEF ""))
  ( (MOT-DEF "I") (FORME-CANONIQUE "Ie I 1")
    (CAT-MOT-DEF "art.") (ALT-MOT-DEF "APS"))
  ( (MOT-DEF "agriculture") (FORME-CANONIQUE "agriculture I 1")
    (CAT-MOT-DEF "f.") (ALT-MOT-DEF "")))
(ANALYSES
  ( (REGLE RQ114111)
    (LIENS ((GNOMINAL-QQ 1 2 2 ((DETERMINANT 1 2 1) (NOM 1 3 1)))))))
```

**II.12 irudia.-** LPPLko sarrera analizatu bat (HDBtik irteten den bezala), HEBaren hasierako eraikuntzan erabilia izateko moduan<sup>1</sup>.

Irudiko lista asoziatibo horretan, gako-hitzaz eskura daitekeen zenbait azpilista ageri da. Esate baterako, CLEF gako-hitzak ematen digu *agricole I I* sarrerari dagokion informazio guztiaren atzibidea. CATEGORIE, USAGE, TYPE-DEF edo EXEMPLE izeneko gakoien bitartez berriz, LPPLko jatorrizko zenbait informazio atzi daiteke. GROUPE-CATEGORIEL gakoari esker eskuratuko dugun informazioa sarreraren kategoria sintaktikoa da, HDBan egindako zenbait eragiketaren ondorioz lortua (izenak, adjetiboak, aditzak eta adberbioak izan dira hemen egindako sail handiak). Sarrerari dagokion definizioa —*relatif à l'agriculture*—, hitzez hitz ageri da MOT-DEF gakoien pean, DEFINITION gakoak indexatutako azpilistaren barnean kabiaturako azpilista zenbaitetan (hitzeko bana). Hor ditugu, definizio-hitz bakoitzari lotuak, haren forma kanonikoa —desanbiguatze lexikalaren ondorioz lortutako zentzu-adierazgailuekin batera—, etiketatze-prozesuetan lortutako kategoria sintaktikoa eta, batzuetan, aldakuntza orto-morfologikoa<sup>2</sup> (itzul II.3 irudira). Azkenik, analisiaren emaitza ere hor ageri da: definizioarekin parekatu patroari dagokion erregela-identifikadorea alde batetik —RQ114111, parekatutako patroien arteko espezifikatzailea<sup>3</sup>—, eta patroiko itemen —

<sup>1</sup> GROUPE-CATEGORIEL gakoak analisisan erabili kategorian "logiko" bat errepresentatzen du (§ II.4.4). '?' ikurrak berriz —homografo- edo adiera-zenbakiaren tokian— hitz-agerpen horretan anbiguetate lexikal bat dagoela adierazten du.

<sup>2</sup> Ikus daitekeenez, zenbait informazio osatzeke dirau. Irudiko kasuan, definizio-hitz guztiak beren kategori etiketa baldin badute ere, *relatif* hitzaren desanbiguetate lexikal ez du erabateko arrakastarik izan (*relatif I ?*).

<sup>3</sup> RQ114111: ((M\* RPA I) (M\* GNOMINAL-QQ I)). RPA I eta GNOMINAL-QQ I aparte definitutako patroia lagungarriak dira. M\* metakarakterea da, hain zuzen ere, ondoko kategoria kategoria lagungarria dela salatzen diguna.

RPA: ( "relatif" "à" / "propre" "à" / "propre" "de" )

GNOMINAL-QQ: ((M+0 DETERMINANT I) (M+ NOM I))

Oharra.- Patroi lagungarriaren sintaxia zertxobait aldatu egin da, argitasunaren izenean. M+0 metakarakterek ondoan datorren kategoriako hitz batekin parekatzeko aukera ematen du. M+ metakarakterek berriz, literalki adierazitako hitz batekin edota ondoko kategoriako hitz batekin (hitz bakar batekin derrigorrez).

*GNOMINAL-QQ 1* eta bere barneko *DETERMINANT 1* eta *NOM 1*— eta definizioeko hitzen artean ezarritako loturak bestetik.

## II.7 Erabilitako prozeduren mugak.

(Boguraev & Levin, 88)n, LNPrako lexikogintzan MRDak erabiltzearen abantailez dihardutela, autoreek, jokabide horrek dituzkeen alde txarrak ere azpimarratu nahiz, "giza adimena eta ergeltasun artifiziala elkarganatzeko premiaren" aurrean gaudela diote. MRDetako ezagutza eskuratzeko lanetan azal daitezkeen bi arazo-iturri nagusi aipatzen dituzte:

- Iturriaren (MRDa) *fidagarritasun eza* ("unreliability"): tekleatze-hutsak, errakuntza "sintaktikoak" (inkonsistentziak, elipsiak, zirkulartasuna, domeinu-hedadurari dagozkion akatsak, etab.) eta "semantikoak".
- Iturriaren *ez-osotasuna* ("incompleteness"): ez-eginezko hutsak eta datu implizituak.

Ezinbestean, gu ere era horretako arazoekin topatu gara LPPLrekin egindako lanetan<sup>1</sup>. Ondoren arazo guzti horien ikuspegi orokorra eman eta erabilitako metodologiaren akatsak azaldu nahi genituzke.

### II.7.1 Analisi sintaktiko-semantikoaren mugak eta arazoak.

Hiztegi-definizioen analisiak izan ditzakeen arazoez ari direla, Boguraev eta Levin-ek (1988) garrantzi handiko informazioa galtzeko arriskua aipatzen dute —analisi partzialeko teknikekin litekeena, dudarik gabe—, arazook definizio-esaldietan lokuzioak erabiltzetik datozenak, erreferentziarekin loturikoak edo definizioen zirkulartasunari legozkiokeenak direlarik.

Esana dugu jadanik gure metodoa bereziki egokia iruditzen zaigula LPPLko bezalako definizio-esaldi labur samarrak analizatzeko. Ordea, horrek ez du inola ere esan nahi analisia kasu guztietan arrakastatsua izan denik eta, are gutxiago, oraintxe ahotan hartu nahi genituzkeen arazo motak izan ez ditugunik.

Horietako problema batzuk aztertu ditugu dagoeneko kapitulu honetan bertan. Hasteko, aurretiko *analisi morfologiko baten faltak* porrotera eraman du analisi sintaktiko-semantikoa zenbaitetan, baita etiketatze heuristikoa (§ II.4.3) egin ondoren ere. Egia da forma flexionatuen agerpen maila hiztegi-definizioetan —eta batez ere LPPLkoetan, definizioak

---

<sup>1</sup> Ez dakiguna, ordea, da, asmatu ote dugun ezarritako prozeduren "ergeltasun artifizialaren" kontrako borroka horretan!



labur eta "eskematiko" samarrak izaki— dexente apala dela, pluraleko formak, adjetiboen femeninoak eta aditz-forma jokatu batzuetatik at (indikativoko orainaldiko hirugarren pertsonari dagozkionak eta partizipioak, gehienbat) horrelako gauza gutxi ageri baita berorietan. Hala ere, bistakoa da analisi morfologikoa alde aurretik eginez gero emaitza zuzenagorik irits litekeela<sup>1</sup>.

Sintaxi mailako arazoak direla eta, *koordinazioari* dagozkionak ez ditugula tratatu esan behar. Estreina-estreinako maiztasun-azterketetan ikusi genuen jada LPPLn konjuntzioen agerpenen maiztasuna ez zela oso handia (%0.51 *et*, eta %0.29 *ou*). Baina, hori baino inportanteago, konjuntziodun definizio-eskema sintaktikorik ez dugu aurkitu maizenekoen artean. Hori dela medio, koordinazioari dagokion problematika —problematika interesgarria izan daitekeela ukatu gabe, hiperonimo koordinatuen kasuan den bezala, esate baterako— alde batera utzi dugu.

Bazterrean utzitako beste arazo bat lokuzio edo nolabait finkatutako edo lexikalizatutako espresioen kasua dugu; ezaguna da kasu horietan interpretazio semantikoa ezin daitekeela "konposizionalki" egin.

*Erreferentzi* arazoak berriz, nahiko modu berezian aztertu dira. LPPLko definizioetan ageri diren izenordain usuenak erlatibozkoak dira. Horrek eraman du gure ikerketa izenordain horiek definizio-metalengoaian betetzen duten funtzioaren aldetik aztertzerara, honela zenbait eskema espezifiko aurkituz. Tratatu dugun beste kasu bat —definizioetatik ateratako erlazioak azaldu ditugunean aipatua (§ II.4.6)—, hain zuzen ere "erreferentziatzko erlazioa" deitu duguna da.

Beste arazo-iturri bat analisirako patroietan *zernahitarako itema erabiltzeak* —*M&&* metakarakterea— dakar. Esana dugu gure aplikazioan item hori ahalik eta gutxien erabiltzen ahalegindu garela. Adjetiboen definizioak analizatzeko prestatu hierarkian, item hori duten 5 patroi besterik ez daude, eta bat bakarra aditzenean. Aitzitik, izenentzako hierarkian maiz samar erabili dugu —patroi bitatik batean (32tan) ageri da—<sup>2</sup>. Item honen erabilerak ekar ditzakeen problemak, aukerazko parekatze-itemekin batera erabiliz gero (bereziki *M+0* itemarekin, ondoko kategoriako zero edo hitz gehiagorekin parekatzen dena, eta *M\*0*rekin, portaera berdina duena baina kategoria konplexuekin) sortzen dira, batik bat. Arriskua egitura semantiko oker bat asignatzean dago, eta are handiagoa da zernahitarako elementua

<sup>1</sup> Etiketatze-fasean egindako kategori esleipen okerrekin ere —batez ere analisi morfologikoak emaitza anbiguoak emango lizkigukeen kasuetan; hala nola *produit* edo *pratique* hitz-formen kasuak, zeintzuk izen-agerpen edo *produire* eta *pratiquer* aditzen forma, hurrenez hurren, izan daitezkeen, "automatikoki" izen gisa etiketatutako izen direnak—, eraman gaitzakete sintaktikoki eta/edo semantikoki okerrak diren analisietara.

<sup>2</sup> Hiru hierarkietako erroa (*M&&*) patroia da, eta ez dugu hemen kontatu, beste ezein patroik arrakastarik ez duenean analisiak aurrera egin dezan baitago hor kokaturik. Patroi horiekiko parekatzea izan da, hain zuzen ere, porrotzat hartu duguna, § II.5.2n emaitzak eman ditugunean.

## II. kapitulua

aukerazko item horietako baten ondotik erabiltzen denean, item hori ezein hitzekin parekatzeke gera baitaiteke. Adibide bat ipintzearren, ikus dezagun *éligible I I* adierari dagokion definizioaren kasua —*qui peut être élu*—, RQ12 patroiarekin —("qui" (M\* ADJS-GVERBAL/PI3 1) (M&&))— parekatua, non zernahitarako itemak 'être élu' sekuentzia "harrapatu" baitu. Horrek esan nahi du definizioan ezagutu den zatia 'qui peut' dela, argi eta garbi analisi akastuna, okerra, emanez<sup>1</sup>.

H. Alshawik berak (1987; 89) esaten du definizioak era ez-osoaz analizatzen direnean gerta daitekeela adiera bati dagokion *informazioaren mamia edo muina bera galtzea*. Areago esango genuke, zenbait kasutan egiaztatu dugunez, ez dela funtsezko informazioaren galtze soila gertatzen, baizik batzuetan, "atzemandako" informazioak ez duela ziurtatzen bere zehaztasuna edo zuzentasuna. Berriro esan, gure lanaren zuribide gisa, LPPLko definizioen laburrari esker hiztegiko 14350 definizioen kasuan (hiztegiko %64a) lortutako analisisia osoa izan dela.

Egiten ahal da, noski, printzipioz emaitza hoberik emango lukeen analisi saiorik. Izan ere, erabilitako analisi-teknikak berez eramaten baikaitu, lortutako emaitzak aurrean izanik, prestatu patroi-hierarkiak hobetzera. Bide horri jarraitu gatzaizkio lehen ere, gorago azaldu ditugun hierarkiak prestatze lanean. Hala eta guztiz ere, hierarkiok hobetu nahi izanez gero badira berrikustaldi bat mereziko luketen kasuak: konparazio batera, ezein definizioarekin parekatu ez diren patroi espezifikokoak edota osatzeke geratutako analisisien kasuak. Bi kasu horietan aise hobe daitezke gauzak, porrot egindako patroi espezifikoena zerbait aldatuz (porrota, definizioaren egitura ezagutzeko ezgai den erregela sintaktiko baten erruz sortua da eta).

ACQUILEX proiektuan definizioak analizatzeko teknika berbera erabili dute, baina, eremu mugatuak hartu dituztelarik, patroi-hierarkia "finagoak" egiteko aukera izan dute: analizatutako definizioen multzoa SUBSTANCE, FOOD, DRINK, PERSON, PLACE eta INSTRUMENT peko taxonomiei dagokie izenen kasuan, eta MOVEMENT eta FOOD taxonomiekin zerikusia duten aditzei (Ageno *et al.*, 92). Fintze horri esker lortzen ahal da, seguruenik, informazio osoago eta zuzenagorik.

Gure lanaren helburu nagusia hiztegi-ezagutzaren errepresentaziorako eredu bat proposatu eta, eredu hori oinarritzat hartuz, hiztegi-sistema urgazlearen egingarritasuna aztertzea zenez, analisi sintaktiko-semantikotik lortutako emaitzak aski zuzen eta aberatsak

---

<sup>1</sup> Analisi txar horren zergatia *ADJS-GVERBAL/PI3* patroia lagungarriaren definizioan datza, zeinek ez baitu kontuan hartzen definizio konkretu honetan ageri den aditz-sintagmaren egitura. Baina arazoa *ADJS-GVERBAL/PI3* kategoriaren egitura sintaktikoan bertan dago, atzeneko tokian aukerazko izen- edo preposizio-sintagma bat baitu. Bistakoa da aukerazko azken sintagma hori ez dela ezein hitz-sekuentziarekin parekatu aztergai dugun definizio horretan eta, aldiz, zernahitarakoa bai.

zirela iruditu zitzaigun aurrera egiteko. Alderdi semantikoari dagozkion problemak, hala ere, hurrengo azpiatalean azalduko ditugu.

Bukatzeko, analisi-emaitzen *anbiguotasunari* buruz bi hitz esan nahi genuke hemen. Anbiguotasun hori, batez ere, patroï-hierarkiari berari zor zaio: analisi bat baino gehiago lortzeko, edo zuhaitzaren adar desberdinetako erregelekin parekatze zuzenak suertatzen dira edo patroï bakar batek parekatze era desberdinak onartzen ditu (azken kasu horretan erregela berari dagozkion analisi bat baino gehiago izango ditugu).

II.13 irudiko taulan LPPLko izen, aditz eta adjetiboen definizioen analisisian izandako anbiguetate maila ikus daiteke, zenbaki absolutu zein erlatibotan. Taulako zenbakiak adierazten dute zenbat definizio parekatu den gehienez lau patroïekin (bat, bi, hiru edo laurekin), sail bakoitzeko bigarren zutabearen portzentaia metatua ere ageri delarik.

Parekatutako patroien kopurua	IZEN-DEF.		ADITZ-DEF.		ADJET.-DEF.	
		%		%		%
<b>1</b>	8033	58.45	4322	82.37	2596	80.52
<b>2</b>	4183	88.88	810	97.81	579	98.48
<b>3</b>	1003	96.18	32	98.42	29	99.40
<b>4</b>	464	99.56	75	99.85	20	100

### II.13 irudia.- Analisi-emaitzen anbiguetate maila.

Arestian esan dugunez, analisi aniztasun hori bi arrazoirengatik suerta daiteke: edo kategoria konplexuek —patroï lagungarrien bitartez definituak— patroï bakarrarekin analisi desberdinak eman ditzaketelako, edo hierarkiako adar desberdinetan kokaturiko patroï desberdinak definizio-esaldi berarekin parekatzen direlako. Ez dago anbiguetaterik, noski, definizio bat adar bereko patroï desberdinekin parekatzen den kasuetan, kasu horietan emaitza patroïrik espezifikoa baita (beste guztiak arbuiatuz).

Datu-basean sintaktikoki zuzenak diren analisi guztiak gordetzen badira ere, HEBaren eraikuntzan lehenengo analisisia bakarrik hartzen da kontuan (lehenengo lekuan dagoen analisisia izango da beraz, ikuspuntu semantikotik zuzenetsiko dena). Jokabide honek behartu gaitu hierarkiako maila berean kokatu beharreko patroïak kontu handiz ordenatzera, izan ere analizatzailea, datu-basean analisiak erregistratzeko orduan, patroï-hierarkia korritu ahala joango baita analisi zuzenak erregistratuz (hierarkian dauden ordena berean, beraz). Hala ere, emaitzak eskuz egiaztatu dira (HEBaren prototiporako hautatutako azpimultzoko sarreren kasuan bakarrik eta analisi aniztasunik baldin bazegoen) eta, beharrezko zenean,

## II. kapitulua

analisien kokapena aldarazi egin da datu-basean, horrela interpretazio semantikorako aurrezuzenketa bat eginez.

### II.7.2 Modu honetan eraikitako HEBaren mugak.

Atal honetan, kapituluan zehar deskribatu erabidea izan duen HEBaren mugak azaldu nahi genituzke, ikuspuntu semantikotik dituzkeenak batez ere.

Hasteko, berriro aipatuko ditugu HEBa sortu duen eraikitze-programa martxan jarri aurretik egin ditugun eskuzko desanbiguaketa lexikala eta forma kanoniko usuenen zuzenketa prozesuak (ikus § II.6.2). Esan bezala, prozesu horiek oso kasu gutxitan egin dira. Konponketa lan horien interesa bistan dago: HEBrako ahalik eta informazio fidagarriena eskuratu. Ez baitakigu, jakin, nora irits gintezkeen hiztegia bezalako iturri ez-oso —eta, agian, inkonsistente— batetik abiatuz, iturri hori, ezinbestean, partzialki eta modu ez erabat seguruan tratatuz<sup>1</sup>. Alabaina, egindako ikerketaren arloa nagusiki lengoia naturalaren tratamendu *automatikoarena* izanik, ez dugu bide horretan sakondu nahi izan, ezagutza ez-oso, eta agian ez-zuzena, daukan hiztegi-ezagutza base baten erabileraren erronkari aurre eginez. Gainera, eta ez da hau gutxienekoa, HEB horren gainean ezarritako dedukzio-prozedurei legokieke —aberasketa dinamikoa esaten diogun prozesuan— akats horiek, ahal den neurrian, zuzentzea<sup>2</sup>.

Analisi-patroiei dagozkien ESEEak idazterakoan izandako oztopoak azaltzeko, aipa dezagun desanbiguaketa lexikalaren berezko zailtasuna, erlatore berezizat hartutako kasuetan ere izugarri gaitza gertatzen baita maiz. ESEE horietan ageri diren erlazio espezifikoek —DEF-PARTIED, DEF-ACTIOND, DEF-RENDRE, eta abar bezalako atributu definitorioen bitartez gauzatuak, eta erlatore berezien ideari zuzenean lotuak— anbigutasun lexikaleko arazoak aurkezten dituzte.

Ikus dezagun '*partie de*' hitzekin hasitako definizioen kasua. LPPLn *partie* sarrerak bost adiera dauzka, eta horietako bik bere zentzu "meronimikoarekin" zerikusia dute:

*partie I 1: portion*

*partie I 2: élément d'un ensemble*

eta *portionen* lehen adiera (bigarrena ez zaigu orain interesatzen) LPPLn hauxe da:

---

<sup>1</sup> Hain zuzen ere, egindako tratamenduen akats batzuk zuzen baitaitezke honela.

<sup>2</sup> Erabiltzailea izan ohi da hiztegiaren akatsak —hiztegia erreferentzi tresna betetingabea denez— nolabait zuzentzen dituena, kontsultatzen duenean. Horra hor hiztegi-sistema urgazle *adimenduna* egin nahiak duen beste erronka bat, sistema horrek izan beharko bailuke, printzipioz behintzat, oztopo horiek gainditzeko gauza, erabiltzaileari modu "adimentsuago" batez lagunduko badio.

*portion I 1: partie d'un tout*

*nef I 1* adieraren definizioan —*partie d'une église*—, nekez jakin daiteke *partieren* zein adierari dagokion definizioko agerpen hori. Eta hori '*partie de*' definizio-formulaz egindako definizio batean gertatzen da (printzipioz ongi zehaztutako formula, esanahi ez-anbigua behar lukeena). Antzeko beste kasu bat '*élément de*' hitzekiko formulez egindako izen-definizioetan gertatzen da, *élément* hitzaren adiera desberdinen artean bereiztea batzuetan zaila gertatzen baita.

Kontzeptu "huts" deitu ditugun kasua litzateke HEB honen beste muga bat. HEBan unitate batez errepresentaturik dagoen baina bere definiziotiko ezagutzarik ez duen kontzeptuari esaten diogu "kontzeptu hutsa". Kontzeptu horiek beren adierazleak besteren definiziotan agertu direlako daude hor —definizio horiek analizatu eta beraiei dagokien unitatea sortu baita orduan—, eta hutsak baldin badaude da HEBaren gune izateko aukeratutako LPPLren azpimultzoko ez direlako. Kontzeptu horien artean daude, esate baterako, besteren definiziotan agertu adberbioak —beren definizioak ez dira eta analizatu—. Horrela, HEBa multzo itxia dela esan dezakegu, izan ere edozein kontzepturen definizioan agertutako edozein kontzeptu-erreferentziak bai baitu bere "errepresentazioa" den unitatea ezagutza-basean (nahiz eta, zenbaitetan, errepresentazio hutsa izan hasieran). HEBan kontzeptu huts horiek edukitzearekin hiztegi arruntetan hain normala den itxidura ezaugarriaren falta<sup>1</sup> arindu nahi genuke nolabait.

Interpretazio semantikoari, oro har, gagozkiolarik, mugarik inportanteenetako bat —H. Alshawik berak ere adierazia (1987; 89)— erabilitako analisi-teknikari berari dagokio: patroiei esleituriko egitura semantikoak definizioen azaleko egiturari estuki lotuak dira. Ikus ditzagun, konparazio batera, bi kasu hauek: (1) preposizioek ezarritako sakoneko erlazio semantikoak —genusaren eta differentiaren artekoak, adibidez— ezin dira bestelako informaziorik gabe azaldu eta, beraz, ESEEtan hortxe geratzen dira gehienbat sintaktikoak diren atributuen bitartez errepresentatuak<sup>2</sup>; eta (2) gune eta modifikatzailearen arteko erlazio batzuk ere —izen eta adjetiboaren artekoa, adibidez— ESEEtan CARACTERISTIQUE (edo antzeko) atributu "zehazkabe" batez errepresentaturik geratuko dira maiz, bestelako informaziorik ezean interpretazio sakonagorik egiteko aukerarik gabe.

<sup>1</sup> Gogoratu LPPLri berari buruzko datuak, sarrera gisa bertan falta direnei buruzkoak (§ II.4.2). Hiztegien itxidurarik eza hitzen mailara ere iritsi ohi da, adiera edo zentzuen mailan askoz ere inportanteagoa delarik, noski.

<sup>2</sup> Esan behar da hala ere, ESEE horietan esplizituki adierazten dela, batzuetan, geroagoko interpretazio semantikorako baliagarri izan daitekeen informaziorik (interpretazio posible bakoitzarentzat —edo batzuentzat bederen— jotako ziurtasun-balioak emanez).

## *II. kapitulua*

Ondoko gogoeta hau egitera eramaten gaitu honek: posible ote da, aldez aurretik domeinua, edo oro har mundua, formalizatu gabe, lengoaia naturalezko esaldien errepresentazio egoki batera iristea? Begira zer dioen honetaz B. Boguraev-ek (1986:15):

Finally, dictionaries are not encyclopaedias. Definitions, however detailed they may be, still assume some basic, general knowledge about the world. Attempts to extract the meaning of a word sense from its description in a dictionary and to convey this through an encoding in a formal knowledge structure require before anything else the formalisation of this general knowledge, without which no useful representation of any particular definition could be achieved. This poses the question of how to establish the common ground assumed by the lexicographers during the process of writing definitions.

Nolabait, munduari buruzko formalizazio bat egin egiten dugula esan dezakegu, hiztegi-definizioetatik erauzitako erlazio taxonomiko eta meronimikoak esplizitu bihurtzen ditugularik. Baina, dirudienez, definizio-esaldien errepresentazio semantiko guztiz zuzena eta osoa ez da posible halako formalizaziorik gabe. Domeinuari buruzko ezagutza, mundu erreala duguna azken finean, formalizatu egin beharra dago. Hori dela eta, HEBan kodetu egitura semantikoa, oraingoz, maila anbiguo batean —hasierako fasean, behintzat— geratu izana ez da harritzekoa; maiz kontzeptuarteko erlazioak azaleko erlazio sintaktiko hutsak izan daitezke (definizio-testutik zuzenan eratorriak), eta gehienez ere, benetan sakonean dagoen erlazioa halako edo halako erlazio semantikoa izateko dagokeen ziurtasunaren ohar moduko batekin aberastuak egongo dira. HEBaren aberasketa-fasean (ikus III. kapitulua) egingo zaio aurre errepresentazio semantiko "finago" bat lortzearen arazoari, eta horretarako ezagutza-baseaz beraz —bere akats eta gaitasun guztiekin— baliatzeko asmoa dugu.

### **III. Ezagutzaren errepresentazioa. Aberasketa dinamikoa eta alderdi inferentziala HEBan.**

Hirugarren kapitulu honen helburua HEBrako asmatu dugun ezagutzaren errepresentaziorako eredua azaldu —aurreneko bi ataletan, § III.1 eta III.2— eta HEBaren hasierako eraikuntzaren ondoren bertan burututako aberasketa dinamikoaren prozesua zertan den deskribatzea da. III.3 atalean berriz, alderdi inferentzialaz arituko gara, hau da, hiztegi-ezagutza horren gainean ezarritako —edo proposatutako— dedukzio-mekanismoez, eta azkenik, § III.4en azalduko dugu eraiki dugun HEBaren prototipoaren egungo egoera zein den.

Aurreko kapituluan azaldu bezala, HEBaren prototipoaren "lehen bertsioa" LPPLko definizioen analisi-prozesuaren emaitza dugu. Prozesu horrek, azken buruan, interpretazio semantiko batera eraman gaitu, hasierako eraikuntzan HEBko kontzeptu-unitateen — nolabait erlazionatuak, etapa horretan— sorkuntza gauzatu. Estreinako bertsio hori sortutakoan —non kontzeptuarteko egitura taxonomikoa esplizitu bihurtua baita, kontzeptu-unitateak sortu ahala exekutatu diren prozedura espezifikokoak direla medio—, bertako kontzeptu-unitateen artean beste erlazio esplizitu batzuk egiteko xedeaz, zenbait eragiketa burutuko dira, horretarako sortu prozedura bereziak martxan jarritz. Halaber, aberasketaren bigarren urratsa den horretan erlazio lexikalen zenbait propietate ustiatuko dira, kontzeptuarteko erlazio berriak sortzeko aukera emango duen desanbiguaketa lexikal eta sintaktikoa eginez. Erabiltzailearen eskariei erantzutean HIZTSUAK erabiltzen dituen dedukzio-mekanismoen artean, erlazio lexikalen propietateen ustiaketa eta erlazioen arteko konposaketa aipatu behar dira, errepresentazio-ereduak berak ematen dizkigunez gain (herentzia taxonomikoa, bereziki).

Aurrera baino lehen, azpimarratu behar da kapitulu honetan azalduko ditugun dedukzio-mekanismoak gehiago direla hiztegi-ezagutzaren errepresentazioari eta inferentzi mekanismoei buruzko proposamenak, prototipoan (benetan) dagoen ezagutzaren gainean erabat inplementatu eta exhaustiboki egikaritutako prozedurak baino.

#### **III.1 Ezagutzaren errepresentazioa.**

Ezagutzaren errepresentazioa funtsezkoa da adimen artifizialaren eremuan. Lengoia naturalaren prozesamenduaren arloan hizkuntz ezagutzaren errepresentazioa ere funtsezkoa dugu.

### III. kapitulua

Barr eta Feigenbaum-ek *ezagutzaren errepresentazioa* honela definitzen digute (*Handbook of Artificial Intelligence*, Pitman: Palo Alto, 1982:143)<sup>1</sup>:

En I.A., un système de représentations de connaissances est une combinaison de structures de données et de procédures d'interprétation qui, si elles sont correctement utilisées dans un programme, conduiront à un comportement cognitif. Les travaux sur la représentation des connaissances ont conduit en I.A. à la conception de plusieurs types de structures de données pour stocker l'information dans des programmes d'ordinateur, ainsi qu'au développement de procédures permettant une manipulation "intelligente" de ces structures de données pour réaliser des inférences.

Logika, sare semantikoak, objektu egituratuak —frame, eskema, prototipo, eta abar—, produkzio-erregelak, eskema prozeduralak, eredu konexionistak, eta beste, dira bereziki Adimen Artifizialean garatu eta/edo erabilitako formalismoak, ezagutzaren errepresentazioaren arazo desberdinei nolabaiteko irtenbideak ematearren. F. Rastier-ek dioen bezala baina (1987:80), *“Aucun de ces formalismes n'a été élaboré pour permettre spécifiquement de représenter le contenu lexical; cependant, tous ont été utilisés peu ou prou pour le faire.”*.

Ezagutzaren errepresentazioaren arloan, eta batez ere hizkuntzaren errepresentazioari dagokionean, komunitate zientifikoak izan duen arazo sonatuenetarikoa zenbait ikerlari formalismo logiko edo "semantika formalen" aldekoa den bitartean beste batzuek beren lanetarako "sare" gisako notazioak hautatu izana izan da.

Ikus dezagun, esate baterako, Robert Kowalski-ren epai itxurako pasarte hau, (Guenthner & Sabatier, 87:103)tik hartua bera ere:

Il y a un seul langage pertinent pour représenter l'information —qu'il soit déclaratif ou procédural— c'est la logique des prédicats du premier ordre. Il y a une seule façon intelligente de traiter l'information, c'est celle qui consiste à appliquer des méthodes d'inférence déductive. La communauté en Intelligence Artificielle aurait dû réaliser cela plus tôt si elle n'était pas si insulaire. La communauté travaillant sur les Bases de Données, par exemple, a appris la leçon très tôt.

Dudarik gabe, ongi definituriko semantika formal garbi bat eta berorri esker ezartzen ahal den dedukzio-mekanismoa lirateke, erabili ohi den notazioaren sinpletasunarekin batera, logikari errepresentazio-formalismo bezala interes handiena damaiotenak. Aitzitik, ezagutza-base bat osatzen duten egitateak antolatzeko printzipiorik eza eta prozedurazko ezagutza<sup>2</sup> eta

---

<sup>1</sup> Definizio hori (Guenthner & Sabatier, 87:104)etik hartua da.

<sup>2</sup> PROLOGek bilduko lituzke bere baitan, nolabait, eskema logiko eta prozeduralen abantailak.



heuristikoak errepresentatzeko zailtasunak lirateke jokabide horren kontrako arrazoi nagusienak (Mylopoulos & Levesque, 84).

Hirurogeitamarreko hamarkadaren azken aldera grafoei logikaren beste notazio bat izatea —nolabait baliokidea— ezagutu bazitzaien ere, beren baliagarritasuna eztabaidagai da oraindik: alde batetik logika sinbolikoa arbuiatzen zutenak, ez-errealista eta linguistikoki baliogabetzat jotz; bestetik berriz, "sare-zaleak", onenean ere, lehen mailako logikaren aldaera hutsen egile edo, txarrago oraindik, inongo denotaziorik gabeko notazio nahasien sortzailatzat zituztenak. Sowa-k dio (1984), arrazoia, nolabait, bi aldeetakoei emanez, logikazale "garbiek" arrazoia dutela "narrasei" egin ohi dizkieten kritika batzuetan baina, beren aldetik ez diotela behar bezalako arretarik eskaintzen lengoaia naturalarekiko korrespondentzia leunari ("mapping") edota eskemen balio heuristikoari. Aurrerago Sowa, logikaren aitzindari batez ari dela —mende hasierako Charles Sanders Peirce— ondokoa gogorarazten du, alegia Peirce "grafo" tankerako notazio batera lerratu zela formalismo linealari ekarri inportante batzuk egin ondoren. Peirce-ren "grafo esistentzialak" dira, hain zuzen ere, beherago xehetasun gehiagoz aipatuko ditugun Sowa-ren grafo kontzeptualen formalismoari oinarri logikoa ematen diotenak.

Ahlswede-k (1985) sareen bidea hartzen du lexikoa —lexiko erlazional gisa harturik ("relational lexicon")— errepresentatzeko, bere iritziz sare tankerako errepresentazioak eredu erlazionalak lexikoari damaion interkonexioaren propietatea nabarmenarazten baitu. Alabaina, bi jokabideok elkarren osagai baino alternatiba gisa ikusiak izan direla eta damuz, predikatu-kalkuluak hizkuntz espresioak dotoreago errepresentatzeko bidea eskaintzen duela aitortzen du, eta bide formal horri lexikoaren muineko hiztegiaren zehaztapenean funtzio inportante bat ezagutzen dio.

Hala eta guztiz ere, hizkuntza baten lexikoaren errepresentazioa, jeneralean, kontzeptu- eta erlazio-multzo bezala egin izan da, era deklaratio batez eta, gehienetan, sare semantikoen (tankera bateko zein bestekoak) bitartez formalizatua. M.W. Evens-ek argitara emandako liburuan (1988), lexikoaren errepresentaziorako ikuspegi "erlazionala" hartzen duen zenbait lan aurkezten da, ikuspegi hori nagusiki sare semantikoen bidez gauzatzen delarik.

Hizkuntzaren unibertso lexikala kontzeptu- eta erlazio-multzotzat hartzen baldin badugu, sareen aukera da, guretzat ere, unibertso horren errepresentaziorako erarik "naturalena"; horrek ez du inola ere esan nahi, noski, eredu hori hautatzeak leharzkeen arazoak alde batera utziko ditugunik.

Arazoari aurre egiteko beste ikuspegi bat lengoaia naturala bera ezagutza errepresentatzeko lengoaiatzat hartzea da. Jokabide horren oinarrian, berorri jarraitu zaizkion autoreen esanetan (Binot & Jensen, 87:709), ondokoa legoke: *"The basic assumption of this*

### III. kapitulua

*research is that natural language itself is a knowledge representation language that can be conveniently accessed and richly exploited.*”. Autore horien lanean, konkretuki, linean duten hiztegi bateko definizioak erabiltzen dituzte desanbiguatze sintaktikoa —interpretazio semantiko anitzeko preposizio-loturen kasua, batik bat— egiteko gauza den "aditu semantiko" baten oinarritzat, horretarako testu eran gordetako definizioak denbora errealean analizatuz. Beren hitzetan, beraiek erabilitakoak bezalako teknikei esker LNPko sistema arruntetan egin ohi den informazio semantikoaren eskuzko kodeketaren premiarik —astuna eta, maiz, ez-oso— ez izateko esperantza piz daiteke.

Bistakoa da azken lan hori gure problematik gertu dagoena, izan ere —baina ez horregatik bakarrik— autore horiek ere hiztegi-definizioekin baitihardute lanean. Gurean egin dena, aurreko kapituluan azaldu bezala, definizio-testuetako hitzen arteko —edo, hobeto esanda, kontzeptuen arteko— erlazioak erauzteari begira definiziook sintaktiko-semantikoki analizatzea izan da, analisi horretan metalengoaia lexikografikoari buruz aurretik egindako azterketetan eskuratu ezagutzaz baliatu garelarik. Kapitulu honetan ikusiko den bezala, kontzeptuarteko erlazio horiek errepresentatzeko hartutako bidea ez da bere jatorria den lengoaia naturaletik gehiegi urruntzen (eta zuzenean parafraseagarri edo gertatzen dela esan daiteke). Aipatu berri ditugun autoreek ere (Binot & Jensen, 87) azpimarratzen baitute, beren artikulua azken aldera, definizioak aldiro "berranalizatzen" ibiltzeak dakarren eraginkortasun galera, beren sistemarako ikasteko era simple baina baliagarri bat —analisiaren ezagututako erlazio lexikalak era "formalizatuago" batean gordez, geroago ostera erabiltzekotan— proposatuz.

Jarrera horiek indartzera dator, errepresentazio-formalismo bezala logikaren alde dihardutenen aurka, (Wilks *et al.*, 90:103)n irakur daitekeen hau:

Our position on the inseparability of knowledge and language is that this goes further than is normally thought and that particular language structures —text structures— are in fact a paradigm for knowledge structures (Wilks, 1978) or, to put it very crudely, knowledge for certain purposes should be stored in text-like forms (as opposed to, say, predicate calculus-like ones).

Aurreraxeago (104. orr.), bigarren kapituluan aipatu dugun MTDaren ("Machine Tractable Dictionary") formatuaz ari direlarik, formatu horrek "*as language-like as possible*" izan behar duela diote.

Gogoratu, bukatzeko, gure sistema eta bertan egindako hiztegi-ezagutzaren errepresentazioa giza erabiltzaileari begira sortuak direla, ohizko hiztegiak giza erabiltzaileen tresnatarako —eta ez LNPko sistematarako— sortuak izan diren gisa berean. Beraz, hizkuntza edo lengoaia naturaletik hurbil dagoen adierazpideak "naturalago" dirudi sistemaren helburu diren erabiltzaileen aldetik begiratuta ere.

### **III.1.1 Sare semantikoak eta frame-sistemak.**

Azpiatal honetan eman nahi dugu, beste ezer baino lehen, sare tipoko errepresentazioen bideari jarraikiz —zentzu zabalean harturik— egindako lan nabarmenenen ikuspegi orokorra, horiek baitira gurean hiztegi-ezagutza errepresentatzeko hautatu ereduaren oinarri teorikoa. Bukatzeko, HEBaren prototiporako usatu dugun softwarea deskribatuko dugu: software hori dela medio, ezagutza errepresentatzeko frame egiturak erabiltzen dira, erlazioak adierazten dituzten atributuen bitartez estekatuak, honela hiztegiarentzat nahi dugun eredu erlazional hori sortuz.

Deitura desberdinekin —sare semantikoak, menpekotasun kontzeptualeko grafoak, partiketadun sareak, herentzia egituratzeko sareak— bada ere, "grafoak" adimen artifizialeko hainbat eta hainbat sistematan erabili izan dira (Sowa, 84). Sare semantikoetan, kontzeptu baten "esanahia" kontzeptuak sarean duen posizioak emana dator, sareko beste kontzeptuekiko dituen loturen bitartez, alegia.

Lengoaia naturala errepresentatzeko sare semantikoak usatzen aitzindari R.F. Simmons (1973) dugu. Hara zer dioen hizkuntza bateko kontzeptuak errepresentatzeko sareak erabiltzeari buruz (63. orr.):

A semantic network purports to represent concepts expressed by natural-language words and phrases as nodes connected to other such concepts by a particular set of arcs called semantic relations. Primitive concepts in this system of semantic networks are word-sense meanings. Primitive semantic relations are those that the verb of a sentence has with its subject, object and prepositional phrase arguments in addition to those that underlie common lexical, classificational and modificational relations. A complete statement of semantic relations would include all those relations that would be required in the total classification of a natural language vocabulary.

M.R. Quillian (1968) izan zen hitzen esanahiak eta hitzen arteko elkarketak errepresentatzeko *sare semantikoak* erabili zituen lehena<sup>1</sup>. Lehenbiziko lan hori —psikolinguistikaren arloan egina, Quillian-en helburua giza memoria semantikoa errepresentatzea zen eta— emankorra izan zen oso sare tipoko errepresentazioen alderditik ikusita, baina errepresentazio-era horren azpiko semantikaren benetako formalizazio saiorik ez zen W. Woods-en lana (1975) iritsi arte egin. J.F. Sowa-ren grafo kontzeptualen sistema (1984) bide honetan egindako lanen arteko adibiderik berrienetakoa litzateke. Hori baino lehen ezagutzaren errepresentaziorako zenbait lengoaia garatu ziren, sare semantiko eta frameetan oinarri harturik: aipa dezagun, adibide gisa, KRL (Bobrow & Winograd, 77).

---

<sup>1</sup> Sare semantiko batean bi ezugarriren bilketa aurkituko dugu: batetik kontzeptuen hierarkizatze bat —kontzeptuak sareko korapilune edo nodoek errepresentatzen dituzte—, "is a" arkuen bitartez gauzatua, eta bestetik, nodo bakoitzetik irteten diren arkuen bitartez errepresentaturik, kontzeptuen ezaugarrien deskribapena.

### III. kapitulua

Sare semantikoen bilakaera kontatzen denean aipatzeke utzi ezin den lan bat (Brachman, 79) da, non ordura arte erabilitako sare semantikoak maila desberdinetan sailkatzen baitira, berorietan definitutako loturen ezaugarrien arabera. Maila horien artean, autoreak maila "epistemologikoa" —edo ezagutzaren egituratze-maila— bereizten du, aurreko notazio asko eta askotan inplizitua izanik ere aipatu artikuluan esplizitu bihurtu eta, handik aurrera, sare tipoko formalismo berri baten sorrera bideratzen duelarik. Formalismo berri horren ezaugarri ditugu, besteak beste, barne-egitura kontzeptuala —kontzeptuarteko rol eta elkarrerlazioen bidez emana— eta herentzia egituratua. Honen ondorioz sortu ziren bada Herentzia Egituratuzko Sareak (SI-Nets, Structured Inheritance Networks), KL-ONE (Brachman & Schmolze, 85)<sup>1</sup> ezagutzaren errepresentaziorako lengoaiarik sonatuen eta aztertuenetarikoaren oinarritzko paradigma.

KL-ONE ezagutzaren errepresentaziorako lengoaia bezala definitzen da, edo, hobeto esanda, sistema bezala, inplementazioak gehigarritzko aukerak ere ematen baititu, hala nola sareak gordetzeko aukera, errepresentatu egitateei buruzko galderak egitekoa, eta abar. Sare semantikoen eta frameen tradizioetik etorritako lengoaia da funtsean, eta hierarkikoki egituratutako *kontzeptu generikoen* bitartez osatutako sareak dira ezagutza errepresentatzeko ematen duen bidea. Maila epistemologikoan oinarritutako formalismoa izanik, egitateak deskribatzeko bi egitura primitibo mota eskaintzen ditu: kontzeptuak eta *rolak*. SI-Net direlakoan baitan datzalarik, alderdi inferentzialetik begiratuta funtsezkoa den kontzeptua ekarri zuen lengoaia honek: *sailkapen-mekanismoa* alegia, zeini esker deskribapen bat —kontzeptu berri bat— hierarkian dagokion lekuan kokatzea posible baita, berak subsumitzen duen ororen gainean eta bera subsumitzen duten guztien azpian alegia. KL-ONE da ezagutzaren errepresentaziorako bi formalismo desberdin bereizten dituen lehen sistema, bata ezagutzen alderdi deskribatzailearen adierazpide izatera zuzendua, eta bestea proposizio- edo asertzio-atalari dagokiona. Bi osagai horien zirriborro bat besterik ez dago hasierako KL-ONE horretan, eta urtetan interesa mantendu zuena lehenengoa izan zen. Hala ere, KL-ONEren ondoren etorritako lengoaietan bigarren osagaiaren ideia askoz ere garatuago dago; sistema edo lengoaia horiei *hibrido* esaten zaie, eta horien artean ezagunenak KL-TWO, KRYPTON, etab. (Brachman *et al.*, 83; 89) ditugu.

Sistema hibridoetan bi lengoaia diferente egon ohi dira, bata terminoen eta beren definizioen errepresentaziorako, normalean *osagai terminologikoa* edo TBox esaten zaiona, eta bestea, *osagai asertzionala* edo ABox, munduari buruzko asertzio edota axiomak errepresentatzeko erabilia. Labur esanda, TBox-eko kontzeptuen definizioek *terminoak*

---

<sup>1</sup> KL-ONE 1977an sortua da estreinakoz, R.J. Brachman-en doktorego-tesiko errepresentazio-paradigmaren lehen inplementazio bezala.

definitzen dituzten bitartean, ABox-eko *asertzioak* termino horiez baliatzen dira munduan egia denari buruzko proposizioak adierazteko.

Errepresentazio-tresna terminologiko bat egokia litzateke, gure ustez, hiztegiko kontzeptuen errerepresentaziorako (Beck, 89). Erlazio lexikal hierarkiko nagusia hiponimia/hiperonimia izaki, erlazio hori sistemak eskainitako erlazio taxonomikoarekin bat egin dugu gure HEBan. Hori dela medio, kontzeptuen beren hiponimoetarako propietate-herentzia errerepresentatzeko sistemaren herentzi zerbitzuez baliatu ahal izan gara, hierarki eratzaileraren bestelako erlazio lexikalen kasurako irtenbide espezifikokoak programatu behar izan ditugularik. Sailkapenaren arazoa ez da gure kasuan hil ala bizikoa, zeren gehienetan "bai baitakigu" non kokatu kontzeptu bat bere definizioa analizatu eta genusa zein den jakinik; hargatik, hiztegi-ezagutzaren gaineko hainbat prozesu deduktibotan hizkuntz kontzeptuen sailkapen-algoritmo baten premia nabaritu dugu (kontzeptuen artean "subsuntzioa" dagoenentz determinatu beharra dakarrena). Gure proiektuan erabilitako softwarearen errepresentazio-lengoaiak ez du benetako arrazonomendu terminologikorik eskaintzen, herentzi era desberdineko hierarkiak errerepresentatu eta ustiatzeko egokia bada ere.

J.F. Sowa-ren lanera itzuliz (1984), esan dezagun grafo kontzeptualak ezagutzaren errerepresentaziorako (deklaratiboa) notazio bat direla, konputazionalki erabilgarriak. Grafo kontzeptualok sare semantikoaren formalismoari zenbait gehigarri ekarri diote, lengoia naturalaren aberastasuna eta konplexutasuna adierazteko era errazago eta egokiagorik eskainiz (Pazienza & Velardi, 87). Grafo kontzeptualak hizkuntz errerepresentaziorako erabiltzen direnean, esaldi edo hitz baten esanahia kontzeptuz eta erlazio kontzeptualez (bi nodo mota) osatutako grafo orientatu baten bitartez errerepresentatuko da. Hitzen errerepresentaziorako grafo mota desberdinak erabiltzen dira, hala nola mota-definizioak ("type definitions"), non hitz bat bere genusaren eta hiponimokideetarik bereizten duten ezaugarri diferentzialen bitartez errerepresentatzen den, eta grafo kanonikoak ("canonical graphs"), kontzeptuari lotutako erlazio kontzeptual desberdinen betegarrien ("filler") murriztapen semantikoak edo adierazteko erabiliak.

Logikarako notazio naturalago baten bila hasita sortuak, grafo kontzeptualak lehen mailako predikatuen logikarako notazio osoa, eta logika modal eta goimailakoetarako zuzeneko hedapenak dituen dira (Sowa, 84:20-21). Sortzailearen esanetan, aurreko hainbat formalismoren sintesi gisakoa lirateke (Sowa, 84:324-325), bereziki Heidorn-en (1972, 75) Augmented Phrase Structure Grammar direlakoak, Schank-en (1975) Menpekotasun Kontzeptuala, Hendrix-en (1979) partiketadun sareak, KL-ONE —esan bezala Herentzia Egituratzeko Sareetan oinarritua— eta Shapiro-ren (1979) SNePS izenekoak (Semantic Network Processing System).

### III. kapitulua

Gurean erabilitakoaren antzeko sistemek, frameetan oinarritutako sistemek, alegia, KL-ONE, KRL, etab.en aurreko berberak dituzte, eskaintzen dituzten herentzi mekanismoak arrazonamendu terminologikoa esaten zaion horren tankerakoak direlarik. Izan ere, logika terminologikoa lehendik sare semantikoetan eta frameetan zegoen zenbait ideia formalizatzeko saio bat da (Patel-Schneider, 91). Frameetan oinarriturik softwarearen merkatuan dauden sistemak, prototipogintzarako eta sistema aditu edo adimen artifizialeko programak egiteko tresna gisa dira aurkeztuak; sistema hauek biltzen dituzte, ezagutza deklaratioa errepresentatzeko partearekin batera, programazio-mekanismo desberdinak, horrelako lanetarako biziki interesgarri egiten dituztelarik. Gure kasuan, errepresentatutako hiztegi-ezagutzaren gainean hainbat dedukzio-mekanismo ezartzeko asmoa genuenez, egoki iruditu zitzaigun tresna desberdinak ingurune bakar batean biltzen zituen sistema baten aukera: hori dela eta, HEBaren prototipoa inplementatzeko horrelako sistema bat hautatu genuen.

Proiektu honen inplementaziorako egindako aukera bada, ezagutzaren ingeniartzarako pentsatutako software-tresna batean oinarritzen da: Intellicorp enpresaren KEE™. KEE adimen artifizialerako ingurune misto bat da, zenbait programazio-paradigma eta inferentzi mekanismo biltzen dituena, ezagutzaren errepresentaziorako frameetan oinarritu lengoaiarekin batera<sup>1</sup>. LISPen gainean eraikitako ingurune honek objetuen ezaugarriak — objetuei orientatutako programazioaren zentzuan—, frameenak, erregelenak eta beste biltzen ditu, ezagutza-basearen eraikitzaileari tresna-multzo integratu bat eskainiz, ezagutzan oinarritutako sistema adituak (KBES, Knowledge-Based Expert Systems) eraikitzean sortzen diren programazio-beharrak bereziki kontuan dituena.

TellAndAsk da ingurunean integratutako lengoia bat, egitate berriak erazagutzeko edota ezagutza-baseko egitatei buruz bilaketa deduktiboa abiarazteko aukerak ematen dituena: sistemaren lengoia horren bitartez Prolog tipoko eskariak adieraz daitezke. Erregelen interpretaziorako ere Prolog erako programazio logikoa erabiltzen da. Erregela-sistemak eskaintzen du tresnarik zorriketarako ("debugging") zein prozesu deduktiboen azalpenerako (Ramamoorthy *et al.*, 87).

Guzti horrez gain, KEEren beste aukera bat objetuei orientatutako programazioarena da, frameei lotutako metodoak —hierarkian beheragoko mailetan berezituak, edota baita desgaituak ere, izan daitezkeenak— mezu-igortze delako mekanismoaz exekututzen direlarik; edo arrazoiketa hipotetikoa egitekoa —domeinu batean suerta daitekeen hainbat

---

<sup>1</sup> KEEen frameak objektuak zein erregelak errepresentatzeko erabiltzen dira eta hierarkikoki antola daitezke. Frameek slotak dituzte objetuen atributuak deskribatzeko; atributu horien balioen gaineko murriztapenak ezar daitezke, sloten eta balioen herentzi modu desberdinak zehaztu, atributu horiek eguneratzean martxan jarriko diren prozedurak esleitu, eta abar.

egoeraren modelizazioa eta alderaketa egiteko aukera—, horretarako "mundu" edo testuinguru desberdinak erabiliz (KEEworlds). KEEworlds egiaren mantenimendurako sistema batean oinarritzen da (TMS, Truth Maintenance System edo Assumption-based Truth Maintenance System); sistema hori, bere aldetik, oso dedukzio-mekanismo interesgarria da, egitateen arteko menpekotasunak ezarri eta automatikoki kudeatzen baititu (aldaketak izanez gero, adibidez).

### **III.1.2 Hiztegi-ezagutzaren errepresentazioa.**

Aristotelerengandik datorkigu (Sowa, 84) hizkuntz entitate edo unitate kontzeptualak errepresentatzeko "mota" hierarkia bat erabiltzeko ideia<sup>1</sup>. Nozio hori nonnahi agertuko zaigu hiztegiko kontzeptuen errepresentazioaz aritzean: kontzeptuok mota-hierarkia edo hierarkia kontzeptual bat eratzen dute, zeinen ardatza klase/azpiklase erlazioa baita.

M.R. Quillian-ek (1968), giza "memoria semantikoaren" eredu baten bila ziharduela, hiztegiko definizioak hartu zituen kontzeptu artean dauden lotura semantikoak iturri eta adierazpidetzat. Honela, hiztegi-definizioen errepresentaziorako formalismo bat proposatu zuen: sare semantikoak. Quillian-i zor diogu "type node" eta "token node" direlakoan arteko estreinako bereizketa: "type node" horiek lirake Aristoteleren moten errepresentazioa, "token node"ak moten testuinguru desberdinetako agerpenak errepresentatzeko erabiliz (Quillian-engan (1968), "token"ak mota-kontzeptuen hiztegi-definizioetako agerpenei dagozkie). Honi buruz, M.W. Evans-ek bere liburuaren sarreran (19. orr.) ondokoa dio:

Quillian was the first to recognize the need for a type-token relation connecting token nodes representing words used in definitions to the type node representing the concept.

Aurreko azpiatalean aipatu terminologiko-asetzional bereizketa type-token erlazio honetatik hurbil dago.

Hiztegi-errepresentaziorako gure proposamenean ere ageri dira mota (type) eta agerpen (token) kontzeptu horiek. Mota-hierarkiaren adierazpide dugu hiperonimia/hiponimia erlazioek eratutako kontzeptu-hierarkia, kategoria desberdinetan kontzeptu lexikalen arteko ordena-erlazio bat ezartzen duena. Agerpenaren nozioa ere —mota-kontzeptuen definizioetako agerpenen errepresentazio bezala— badarabilgu, nahiz eta gurean ikuspegi

---

<sup>1</sup> Aristotelerentzat (384-322 K. a.) definitu beharrik ez zuten bederlatzi mota primitibo besterik ez zegoen: GAIA, KOPURUA, KUALITATEA, ERLAZIOA, DENBORA, LEKUA, EGOERA, IHARDUERA, eta PASIBOTASUNA. Gainerako mota guztiak definitu egin behar dira, goimota baten azpian sailkatuz eta beren mailako besteengandik elementu espezifikoen bitartez bereiziz (genus et differentia).

### III. kapitulua

asertzionaletik baino errepresentazio-ereduaren osagai terminologikoaren elementutzat hartu<sup>1</sup>.

Atal honetan hiztegi-ezagutzaren, eta bereziki hiztegi-definizioen, errepresentazioaren inguruan egindako hainbat lan aipatu eta aztertuko ditugu laburki.

Hiztegiaren egituraketaz eta bertan kodetu beharreko informazioaz, A. Zampolli-k (1980) honako hau dio:

A machine dictionary is seen as an archive which is suitable for registering the body of knowledge produced by linguistic research in its dynamic development. This knowledge can be incorporated, for example, under the form of syntactic/semantic features, relationships of synonymy and antonymy, codes of semantic fields, case structures, possible constructions, rules of selection and co-occurrence, and must be able to be modified allowing both for the evolution of linguistic theories and the incorporation of additional data coming from the corpus.

Eredu erlazionalaz berriz, ezagutza lexiko-semantikoaren adierazpide gisa hartuta, (Iris *et al.*, 88:263)n ondokoa irakur daiteke:

A relational model represents human semantic memory as a network in which each node is a concept and concepts are linked together by a variety of semantic and lexical relations.

Izan ere, aurrerantzean ahotan hartuko ditugun eredu guztiak erlazionalen sailean koka ditzakegu, hizkuntz lexikoaren errepresentazioaren aldetik ikusita behintzat. Beren arteko aldeak erabiltzen dituzten formalismoetan edota definitutako erlazio edo primitiboen multzoan bilatu beharko ditugu.

(Yoshida *et al.*, 82)n, japonierazko hiztegi arrunt batetik informazio semantikoa erauztera zuzendutako lan bat deskribatzen da. Hiztegi-definizioen errepresentaziorako —definiziook analizatu ondoren— autoreek "grafo semantikoak" deiturikoak usatzen dituzte, funtsean ohizko sare semantikoak direnak. Kontzeptuak lau kategoriatan sailkatzen dituzte: objektu konkretuak, gertakariak, atributuak (egoerak) eta izaki abstraktuak, nolabait izen, aditz, adjetibo/adberbio eta hitz abstraktuen pareko edo direnak. Hitz-klase bakoitza errepresentatzeko frame desberdin bat erabiltzen dute.

N. Calzolari-ren lanetan (1983; 84a; 86; 90) ere lexikoa kontzeptuarteko erlazioen unibertso konplexu gisa hartzen da. (Calzolari, 83:49)n esaterako, “. . . *we should arrive at representations of the complex interrelations, existing between concepts and codified in our linguistic system, . . .*” dio. Kontzeptu bakoitzak, bere esanahia, elkarrerlazio-unibertso honetan duen kokalekutik eskuratzen duela —"erlazional" tipoko errepresentazio hauetan— esaten du (Calzolari, 84a:460)n:

---

<sup>1</sup> Ikus § III.1.3.4en, *kontzeptu definitu* eta *konfigurazioei* buruzkoa.



In a "relational organization" of a computerized dictionary with complex interlinked structures, each word acquires its meaning as a result of its position in some of the partitionings created by the formalized relations.

Lexikoaren errepresentaziorako eredu erlazionalaren egokitasuna ostera azpimarratuz, (Calzolari, 86; 90)en autoreak datu-base erlazionalak eta sare semantikoen erabilera aholkatzen du, ezagutza lexikala errepresentatzeko biderik egokienak direlakoan. Aipatu artikuluotan halako errepresentazio batek esplizitu eman beharko lukeen zenbait erlazio aipatzen du: kasu-erlazio edo erlazio tematikoak, hiponimia, argumentu desberdinen hautapen-murritzapenak, eratorpen lexikaleko erlazioak, eta abar.

G.A. Miller eta bere taldekideen lanen helburua (Miller, 85; Miller *et al.*, 88), psikolinguistikaren mundutik abiatuz, giza burmuineko ezagutza lexikalaren antolaketa ulertzea litzateke. (Miller, 85)en oso eredu sinplea proposatzen da kontzeptuen esanahia errepresentatzeko: kontzeptu baten esanahia kontzeptu sinonimoen multzo batek errepresentatuko luke. Autoreak "hiztegi-matrizea" deitutakoaren bitartez errepresentatzen ditu hitzen arteko erlazioak —matrizeko zutabeetan— eta hitzen adierak —matrizeko errenkadetan kokatuak—, horrela sinonimia ardatz horizontalean errepresentatuz eta polisemia bertikalean. Lan berriago batean, (Miller *et al.*, 88), egileek WordNet azaltzen dute, ingelesaren erreferentzia lexikaleko sistema elektronikoa edo. WordNet horretan ere, kontzeptuak errepresentatzeko sinonimo multzoak usatzen dituzte, oraingoan bestelako erlazio semantikoen adierazpide diren erakusle etiketatu batzuk gehituz: hiponimia, meronimia, antonimia, eta beste. Artikulu horren autoreek diotenez, hiztegi-definizio tipikoen akats bat kontzeptuen arteko sinonimia "ezkutatzea" litzateke; sinonimi ezkutatzeko horren arrazoia, maiz, lexikografoak kontzeptu horiek definitzeko era desberdinak erabili izana izan ohi da.

G. Arangoren tesia (1983) gure lanaren aski hurbila da. Hiztegiko definizio batetik abiatuz, Arangoren programa "analizatzaileak" —datu bezala definizioa analizatzeko moduan hedaturiko ATNa hartuz— kasu-egiturazko frame bat eratzen du, sare semantikoen ezaugarri nagusiak berekin dituen errepresentazioa osatuz. Behin definizioa "ulertuz" gero, Arangoren sistemak "ulerkuntzaren frogapena" edo egiten duen bigarren parte bat du, non galderak analizatu —"ulerkuntza" faseko formalismo berean errepresentatuz— eta erantzunak ematen dituen. Sistema probatzeko espainolezko Larousse hiztegi bateko definizio-multzo murriz samarra erabili zen<sup>1</sup>.

---

<sup>1</sup> "Le système a été testé avec une vingtaine de mots qui représentent les mots générés à partir du concept 'luz' (l'équivalent de lumière), et sur un ensemble d'une cinquantaine de concepts générés à partir du concept 'fuego' (l'équivalent de feu)." (Arango, 83:30).

### III. kapitulua

(Ahlsweide, 85)en ere erlazio lexiko-semantikoen adierazpidetzat sare bat erabiltzen da, halako errepresentazio bati esker eduki daitekeen ahalmen inferentzialaren aldetik begiratuta erlazio horien propietateek duketen garrantzia azpimarratzen delarik. MTQ<sup>1</sup> ereduan hiru erlazio primitibo besterik ez dira erabiltzen, beste erlazio guztiak hiru horietara errendi daitezkeelarik: Modifikazioa, Taxonomia eta "Queuing" delakoa. Eredu honetan, sinonimia erlazio taxonomikoaren muturreko kasua edo litzateke (*a* eta *b* sinonimo dira: *a is a b* eta *b is a a*), eta parte-osoa erlazioa ere taxonomiaren funtzioan eman daiteke: "*Behatza is a eskuaren-parte*" (non, *eskuaren-parte* espresioan *eskua* eta *partearen* arteko modifikazio-erlazio bat bailegoke)<sup>2</sup>.

Lexikoaren errepresentaziorako ikusmolde erlazionalaren inportantzia indartua gertatzen da (Ahlsweide *et al.*, 86:66)n, ondokoa diotelarik:

First, the relational links between an entry and the entries for the words used to define it, give indirect access to far more world knowledge than that contained in the immediate definition.

J.F. Sowaren grafo kontzeptualak ere erabili izan dira hiztegiaren ezagutza semantikoaren errepresentaziorako, nahiz eta, kasu honetan hiztegia bereziki lengoia naturalaren tratamendu automatikoari begira hartu izan den (Pazienza & Velardi, 87; Sowa, 88). Aipatu lan horien aurrenekoan, definizioak errepresentatzeko —Sowaren ereduaren arabera— kontzeptu eta erlazio kontzeptualen grafo bat erabiltzen da. Autoreen esanetan, beren sistemak, ezagutza semantikoaren errepresentazioaren arloko panazea ez bada ere, *ikusmolde semantiko sakoneko* baten bideragarritasuna —baita mugatu gabeko domeinuetan ere— erakusten du.

E. Chouraqui eta E. Godbert-en lana (1989) lehenago aipatu ditugun Herentzia Egituratzeko Sare Semantikoen (SI-Net) tipoko errepresentazioen formalizaziorako proposamen bat da, objektu-unibertsoak objektu arteko erlazioen bitartez deskribatzeko aukera ematen duena. Lan hori hemen aipatzen badugu definitutako formalismoa segituan hiztegiko definizioen errepresentazioan erabiltzen dutelako da. Egileek ez dute esaten, egia esan, hiztegiko definizioak nola bihurtu beren formalismora, baina ikusmoldea interesgarri zaigu bere baitan duen formalizazio saioagatik bereziki. Formalizazio horren oinarrian, klase eta multzoen teoriatik eratorritako propietate formalekin batera, sare semantikoetako oinarritzko elementuak eta objektuei zuzendutako errepresentazioetakoak leudeke: kontzeptua, instantzia, erlazioa, eta abar.

---

<sup>1</sup> MTQ eredua (Modification-Taxonomy-Queuing) Oswald Werner-ek proposatu zuen antropologiaren arloan, kontzeptuen arteko erlazio primitiboen multzo ideal bezala. Multzo hori, unibertsal izateko asmoarekin jaioa, kalkulu proposizionalatik ekarritako zenbait erlazio bolearrekin aberastua izan da.

<sup>2</sup> "queuing" erlazioak ez bide du inongo zereginik definizioen errepresentazioan.

Y. Wilks-ek eta bere taldekideek berriz (1989, 90), MTD (Machine Tractable Dictionary) delakoaren eraikuntzan dihardute. MTD horren helburu nagusia lengoia naturalaren tratamendu automatikoan erabiltzea bada ere, hizkuntz semantikaren ikerketa teorikorako ere baliagarri nahi lukete. Sistemaren iturburua euskarri magnetikoan dauden hiztegi arruntak lirateke (LDOCEa bereziki), eta beren artikuluetan, hain zuzen ere, MRDak MTD bihurtzeko eskala handiko hiru metodo deskribatzen dituzte (Wilks *et al.*, 90). Hiztegiko hitzen esanahien errepresentazioari dagokionean, esanahi edo zentzu horien alderdi intentsionalaren adierazpide diren frameak erabiltzen dituzte, sare hierarkikoetan antolatua (Wilks *et al.*, 89:220):

The output of the dictionary parsing program is a lexicon of word-sense frames, with each frame explicitly or implicitly positioned in multiple, pre-existing, hierarchies.

Zentzuen adierazpide diren frameok ("sense-frames") diseinatzeko —Y. Wilks-en taldeko proiektu horietan— ezagutzaren errepresentazioari, testuen egitura semantikoari eta, bereziki, hiztegien egiturari dagozkion printzipio komunak izan dituzte oinarri. Frame horiek beste hitz-zentzu batzuek osatzen dituzte, zeintzuek, bere aldetik, beren framea izango baitute, Quillian-en "planoen" edota benetako hiztegi baten antolamendu zirkularren tankeran. Zentzu-frameek bi atal nagusi dituzte, arkuak eta nodoak. Arkuon artean, zentzua bere genusarekin (beste hitz-zentzu bat, bere framea edukiko duena) lotuko duen arku etiketatua topatuko dugu. Zentzu-frameen arteko arkuak hitz-zentzuen sare semantiko "usu egituratua" osatzen dute, egileek "sense-network" deritzatena. Egileek dioten legez, arkitektura orokor hau, nodoetan frameak dituen sare semantiko baten egitura duena, frameetan eta sare semantikoetan oinarritu ezagutzaren errerepresentaziorako ereduen antzekoa da oso<sup>1</sup>.

Hiztegiko zentzu edo adieren errerepresentaziorako gure eredua ere honen kidea da: alde batetik (1) kontzeptuen adierazpide ditugun frameek kontzeptuon intentsionalitatea errerepresentatzen dute, (2) kontzeptuen arteko erlazioak atributuaren bitartez —arku etiketatua— gauzatzen dira, eta (3) atributu horien balioak erakusleak dira benetan, atributuak denotatzen duen erlazioaren bitartez loturiko kontzeptuen frameen erakusleak alegia.

F. Pigamo-k bere tesian (1990) —lengoia naturalaren tratamenduan maila "makroskopiko" bateko analisi semantikorako tresnen premia nabarmenak arduraturik— egiaztatu du ezagutzaren errerepresentaziorako eredu guztien oztopoa errepresentazio desberdinetan egindako analisi semantikoaren eraginkortasuna frogatzeko zailtasunean datzala, horren arrazoa izaki hizkuntzaren parte aski zabala hartuko duen sistema bat

---

<sup>1</sup> Eskemen teoria aipatzen duten hemen: KRL, FRL, KL-ONE eta Frail (Wilks *et al.*, 89:223).

### III. kapitulua

garatzeko ezintasuna. Egilearen hitzetan, “. . . *l'alternative consiste à établir un modèle permettant de procéder à la construction des descriptions de manière automatique, ceci à partir de la seule source disponible d'informations sémantiques: les dictionnaires de langue.*” (Pigamo, 90:2). Hori dela eta, Pigamok Réseaux Sémantiques micro-structurés (RSmS) izeneko formalismo simple eta egitura txikikoa ("faiblement structuré") diseinatu eta garatu du. Autoreak berak dioenez, ezagutza semantikoa sare semantikoen bitartez errepresentatu badu eredu asoziatibo baten premiak eraginda izan da, egitura-baterakuntzan oinarritu eredu batekin kontrajarriz. Bestalde, aldez aurretik zenbait erlazio "primitibo" hautatzearen aldekoen kontra, eraikitze-prozesuari "uzten" dio zer-nolako loturak erabiliko diren erabakitzen, (Chouraqui & Godbert, 89)n idaroki moduan edo.

Gurean hurbil gaude, Pigamorenetik, errepresentazio-formalismoaren asoziatibotasunari dagokionean, aukeratutako sare semantikoen eredu tankeran, tratatutako lexikoaren orokortasunean —domeinu mugatu baten aukeraren kontra—, eta, aurrekoaren ondorioz, errepresentagai den ezagutza semantikoaren iturritzat euskarri informatikoan dauden hiztegiak erabiltzeari dagokionean. Aitzitik, gure abiapuntua erlazio lexiko-semantikoen multzo bat da —primitiboak, nahi bada—, aurreko kapituluan azaldutako definizioen azterketatik eratorria eta azterketa horrexek zilegiztatua.

Bukatzeko, aurreko kapituluan zabal samar azaldu dugun ACQUILEX proiektuan, grafoetan oinarritu baterakuntza motaduna ("typed graph-based unification") darabilen lengoaia sortu dute ezagutza lexikalaren errerepresentaziorako, herentzia lehenetsiarekin ("default inheritance") aberastua eta ezaugarri-egituren ("feature structures") baterakuntza lehenetsiaren funtzioan formalizatua (Copestake, 92). Formalismo hori erabili dute proiektuan sortutako ezagutza-base lexikalaren (LKB) garapen eta inplementaziorako.

Esan beharra dago, ACQUILEX proiektuaren helburua LNPrako ezagutza-base lexikalen eraikuntza denez, LKB horretan badela lexikoko elementu desberdinei buruzko informazio sintaktiko eta semantikorik. Proiektu horretan ere Alshawiren analizatzailea (1987, 89) erabili dute hiztegietako definizioetatik informazio semantiko partziala erauzteko. (Alshawi, 87; 89)n lortutako analisisien emaitza diren egitura semantikoak analizatutako definizioetan ageri diren modifikatzaile desberdinen eta erlatiboetako perpausen isla dira, beti ere definizioaren gune semantikoaren erreferentzia esplizitua egiten dutelarik (definitu kontzeptuaren genusa). Egitura semantikoon forma, sintaxi aldetik, atributu-balio bikotezko lista kabiaturia da<sup>1</sup>.

Txosten honen II. kapituluan esan bezala, eskuz eraikitako kontzeptu-hierarkia — "types"— bati esker egindako finketak posible egin du, ACQUILEXen, definizio-analisitik

---

<sup>1</sup> Erabilitako atributu batzuk honako hauek dira: CLASS, PURPOSE, HAS-PART, PROPERTY, MATERIAL, etc.

emaitza hoberik eta zuzenagorik eskuratzea, hobekuntza horretarako analisi-patroien hierarkiak mota-kontzeptu (type) desberdinentzat zehaztu behar izan direlarik.

Mota-hierarkia hori da ACQUILEXen errepresentazio-ereduaren oinarria. Mota-hierarkiaren eraikuntzan Carpenter-en (1990) ideiei jarraitu zaizkie. Mota bakoitza hierarkian zehar heredatuko den ezaugarri-egitura baten jabe da. Lexiko-sarrerak errepresentatzeko ezaugarri-egitura horietako bat erabiltzen da, hiztegiko definizioaren analisitik ateratako eta besterik ezean heredatutako informazioarekin gutxi edo gehiago osatua<sup>1</sup>. Sarrera lexikal baten genusa LKBan errepresentatzeko beste sarrera bat erabiltzen da, aurrenekoak azken honetatik heredatuko duelarik (besterik ezean). Herentzia lehenetsia inplementatzeko baterakuntza lehenetsiaren bertsio bat erabili da<sup>2</sup>. ACQUILEXen, mota-hierarkiaz gain ezaugarri-egituren artean beste hierarkizazio bat ezartzen dute, "psort"-hierarkiak sortuz. Egileek, mota-hierarkia lexikoan antolamendu bat inposatzeko beharrezkotzat jotzen duten bitartean, "psort"-hierarkia horiei deritzete egokiago hiztegitik erdiautomatikoki eratorritako taxonomiak errepresentatzeko (Copestake, 92).

### **III.1.2.1 Ezagutza lexikalaren errepresentazio semantikoa.**

Azpiatal honetan aipatu nahi ditugun lanak semantika lexikalaren arlokoak dira, gure errepresentazio hautuan nolabaiteko eragina izan dutenak, hain zuzen.

HIZTSUAren xedea giza erabiltzailearentzat laguntza-tresna izatea denez, gure arreta semantika lexikalak —hizkuntza bateko elementu lexikalen esanahiaren errepresentazio gisa harturik— eta munduari buruzko jakintza orokorraren hurbiltasunak hartu dute. Horren inguruan (Miller, 85)en irakur daiteke, hizkuntz ulerkuntzarako sistema orok eduki behar lukeen informazio lexikalaz ari delarik, informazio horrek eta sistemak munduari buruz daukan ezagutza orokorrak oso loturik egon behar dutela.

D.A. Cruse-ren liburuan (1986) alderdi teorikotik aztertzen da semantika lexikala, ikuspegi orokor bat emanez. Semantika lexikala hitzen esanahiari dagokio: hitzen portaera semantikoa, izan ere, oso utzita egon da literaturan, esaldien semantikarantz eta semantikak logikaren sistema formalekiko dituzkeen harremanak aztertzerak makurtua batik bat. Cruse-ren liburu horretan erlazio lexikalei buruzko egitate deskribatzaile eta jeneralizagarriak ezartzen dira, semantikaren edozein teoria formalek bere baitan behar lituzkeenak, mintzo-

---

<sup>1</sup> Izenen egitura semantiko lexikalaren adierazpidetzat "Relativised Qualia Structure" direlakoak erabili dituzte ACQUILEXen, J. Pustejovsky-k (1989, 91) deskribaturiko "qualia" nozioan oinarrituak.

<sup>2</sup> Baterakuntza lehenetsian, ohizko baterakuntzan ez bezala, informazio inkonsistentea baztertu egiten da, horregatik porrotik eragin gabe ordea.

### III. kapitulua

esaerak, anbiguotasun lexikala, sinonimia, hiponimia eta meronimia bezalako erlazio hierarkikoak, eta abar sakonki aztertuz.

Gure proiektuan hautatu errepresentazioaren gaira itzuliz, C.J. Fillmore-ren (1968) lan sonatua, kasu-gramatiken sorrera ekarri zuena, izan da aditzentzat hartutako errepresentazio lexikalaren oinarri. Hainbat aditz-definizio, aditz-sintagmen bitartez eratuak izanik, oso egoki dira definizioaren gunea —eta definituaren genusa, beraz— den aditzak bere aldameneko izen-sintagmek denotaturiko kontzeptuekiko dituen kasu-erlazioen bidez analizatuak izateko. Jakina, kasu-erlazioen adierazpide izango ditugun atributuen "filler"ak rola era "prototipiko" batez beteko duten kontzeptuen erreferentzia izango dira, izan ere horixe baita hiztegi-definizioon ezaugarrietako bat (ikus § II.2.2n honi buruz esandakoa). Talde honetako definizioen analisiaren azken helburua sakoneko kasu-egituraren funtzioan errepresentatzera iristea bada ere, proiektuaren egungo egoeran aditz hauen errepresentazioak, maiz, maila azalekoago batean dirau (ikus hemen behean).

Interesgarria ere, hemen azaltzeko, kontzeptuen errepresentaziorako J. Pustejovsky-k (1989, 91) garatutako "qualia" direlakoan teoria. Qualia-egituren teoria ("Qualia Structure") izen-adierazlea duten elementu lexikalentzako errepresentazio-lengoaia bat da. Izen-adiera baten qualia-egitura adierari dagokion informazio lexikalaren parte batek osatuko luke, hain zuzen lexikoaren herentzidun sare moduko antolaketa batean atzigarriena edo litzatekeen informazioak<sup>1</sup>. MRDetatik lor dezakegun informazioa izango litzateke (Briscoe *et al.*, 90), eta horrexegatik aipatzen dugu hemen. Gainera, HEBan izen-kontzeptuak biltzeko sortutako ENTITES klasean definitu dugun zenbait atributu qualiaren nozio horretatik hurbil dago.

Hitz baten qualia-egiturak haren esanahiaren lau alderdi adierazten ditu (Pustejovsky, 91), izen-kontzeptu baten semantika aditzen argumentu-egitura tipikoaren tankeran ezaugarritzen dutenak:

- kontzeptua eta bere atalen edo osagaien arteko erlazioa: osagai-rola ("constitutive role")<sup>2</sup>;
- eremu zabalago batean bereizten duten ezaugarri fisikoak: rol formala ("formal role")<sup>3</sup>;
- bere helburua eta funtzioa: ("telic role")<sup>4</sup>;

---

<sup>1</sup> Kontzeptuei buruzko bestelako informazioa, orokorragoa edota "entziklopedikoa" esaterako, ez bide litzateke hain aise atzitzeko modukoa sare honetan (Briscoe *et al.*, 90).

<sup>2</sup> Gaia, pisua, atalak eta osagaiak.

<sup>3</sup> Orientazioa, magnitudea, forma, neurriak, kolorea eta posizioa.

<sup>4</sup> Agente batek, ekintza bat burutzen diharduela, duen helburua, eta zenbait iharduera zehazten duen berezko funtzioa.

- bere jatorria edota kausa: agente-rola ("agentive role")<sup>1</sup>.

J. Pustejovsky-ren arabera, semantika lexikal konputazionalako edozein teoriak kontuan hartu behar lituzke errepresentazio lexikaleko maila desberdin batzuk: argumentu-egitura, gertakari-egitura ("event structure"), qualia-egitura eta herentzia lexikaleko egitura. Azken mailari buruz, (Briscoe *et al.*, 90)en diote herentzi egitura horrek determinatzen duela hitz bat lexikoko beste kontzeptuekiko nola erlazionatzen den, eta maila semantiko horri dagokiola ezagutza lexikala eta munduari buruzko jakintza orokorraren arteko lotura (zentzu komuna).

Erlazio semantikoei buruz berriz, (Winston *et al.*, 87:418)n irakur daiteke erlaziook ere memoria semantikoa daudela gorderik, kontzeptuak bide daudenaren antzera edo:

Current theories of the structure of the lexicon generally assume that knowledge of semantic relations is stored in semantic memory in a structured and interrelated fashion.

Gurean egin dugu zertxobait zentzu horretan, erlazio lexiko-semantikoei dagokien guztia metaezagutza gordeta dagoen basean errepresentatuz (ikus § III.1.3.2): erlazio bat eta hiztegiko kontzeptuen arteko lotura, erlazio lexikal bat beste batzuekin konposatzeko moduak, etab. Honetaz esan behar dugu erlazioen hierarkizazio bat, errepresentazioa hobetu eta portaera deduktibo arinago bat ezartzeko, ikerkizun geratutako zerbait dela.

### **III.1.3 Ezagutzaren errepresentazioa HIZTSUAren HEBan.**

Atal honetan azalduko dugu gure proiektuan proposatutako ezagutzaren errepresentazioa zertan den. Errepresentazio hori dela eta, azpimarratu beharreko puntuak hauexek lirateke: (1) THESAURUSEko (errepresentazioaren osagai diren hiru ezagutza-basetariko bat) kontzeptu-thesaurus edo elkarrerlacionaturiko zentzuen sarea, eta (2) STRUCTURES ezagutza-basea, HEBko lexikoaren antolamendu orokorra islatu eta bertako kontzeptu-egitura eta erlazioei buruzko metaezagutza daukana.

Aukeratutako errepresentazio-eskemak —kontzeptuen thesaurusari dagokionez— *frame erako egituretan*<sup>2</sup> du bere funtsa. Frameok kontzeptu edo lexikoko zentzuen adierazpide

---

<sup>1</sup> Sortzailea, "artifact", "natural kind" eta kausazko katea.

<sup>2</sup> Esan bezala, frame bat egitura bat da, errepresentatzen dituen objektuak deskribatzeko atributuak edo slotak dauzkana. Slot horien bitartez objektuen arteko erlazioak adieraz daitezke: izan ere, slot baten balioa beste objektu baten erreferentzia denean erlazio baten aurrean gaude, bi objektuen arteko lotura baten aurrean alegia. Frameetan oinarritu sistema batean aurkituko ditugun erlazio nabarmenenak objektuen arteko hierarkia bat eratzen dutenak dira, hierarkia horrek herentzia bitartezko dedukzioak egiteko aukera eskainiko duelarik: klase/azpiklase eta klase/elementu erlazioak dira. Herentziaz hitz egiten dugunean balio-herentziaz dihardugu —sloten balioen herentziaz, hain zuzen—, hierarkian beharagoko mailetara heredatzen diren balioez alegia (azpiklase eta elementu edo instantziei pasatzen zaizkienak), baina baita informazio-egituren herentziaz ere, hau da, hierarkiako objektuen artean gauzatzen den sloten eta sloton ezaugarrien intsuldatzeaz.

### III. kapitulua

dira, eta kontzeptuarteko erlazio lexiko-semantikoen errepresentazio diren atributuen bidez elkarrerlazio naturik daude. Hauxe iruditzen zaigu, lehenago ere esana dugunez, hiztegian dagoen ezagutzaren —definizioen bitartez ezarritako elkarrerlazio kontzeptualen multzo konplexu bezala ikusita— adierazpiderik "naturalena"<sup>1</sup>.

Zentzu-thesaurusari dagokionean, aurrerago alderatuko dugu gure proiektuan hartutako bidea sareetan oinarritu beste errepresentazio-formalismo zenbaitekin (§ III.1.3.5). Zehazkiago esanik, etsenplu bidez erkatuko dugu gurea sare-errepresentazio klasikoekin, J.F. Sowaren (1984) grafo kontzeptualekin, eta hiztegi-definizioen errepresentazioa gaitzat hartu duen beste zenbaitekin (Arango, 83; Chouraqui & Godbert, 89).

Errepresentazio-eredua zehatz-mehatz deskribatzeari ekin aurretik, berriro esan errepresentazio hori zilegiztatzeko duena hiztegi arrunt baten gainean egindako lan enpirikoaren emaitzak direla, aurreko kapituluan azaldutakoak hain zuen. Hiztegi horretako definizioetan aurkitutako egitura sintaktiko-semantikoak eta definizio horien bitartez gauzatutako erlazio esplizitu zein inplizituak dira errepresentazioaren oinarria. HIZTSUAK eskaini beharreko funtzionalitateak ere, hiztegi-laguntzarako sistema den heinean, izan du, noski, bere pisua errepresentazioaren diseinua egiterakoan (ikus IV. kapitulua).

Halaber, ez dago esan beharrik tesi honetan ezagutzaren errepresentazioaren aldetik proposatutakoa aldakizun eta, jakina, hobekizun geratzen dela. Horregatik, HEBa diseinatu dugunean, litezkeen aldaketa edota hedapenak ahalik errazen egiteko modua izan dadin ahalegindu gara.

#### III.1.3.1 Ezagutzaren egituraketa.

HIZTSUAren HEBko ezagutza hobeki egituratzeko helburuarekin, ezagutza hori hiru ezagutza-azpibasetan (azpi-EB) banatzea erabaki genuen: THESAURUS, STRUCTURES eta DICTIONNAIRE izenekoak.

---

<sup>1</sup> Hori dela eta, B. Boguraev-ek (1987:12-13) ez du inolako zalantzarik agertzen hierarkikoki egituratutako sareen erabilgarritasunaz, nahiz eta bere kasuan hizkuntzaren ulerkuntzarako sistema baten lexikoaren errepresentazioaz aritu:

“While there is no consensus on the kind of representation scheme best suited for capturing the knowledge required for language interpretation and understanding, it is nonetheless possible to identify the distinct classes of propositional systems and type-hierarchy systems. The latter utilise, broadly speaking, general notions of frame-like concepts with slot-like role descriptions, organised in an inheritance hierarchy along generalisation/specialisation axes. Most of these recent work on knowledge representation, whether representative of the strict type-hierarchy approach (exemplified by, e.g. FRL, KRL, NETL, UNITS, KL-ONE) or of the hybrid style of KRYPTON and KL-TWO, for example, can be cast into this general mould.

For natural language processing, where the general world knowledge is just as important as specialised domain- and task-dependent knowledge, the utility of hierarchically structured networks of concepts need not to be emphasised. And while it is not entirely clear whether all of the structured information required by the system functions can be derived in a systematic and consistent way from dictionary sources, such sources offer a particularly good and convenient starting point for initial compilation of taxonomically structured knowledge about the world.”



THESAURUSen kontzeptu-sarea errepresentatuko dugu, hots, hiztegiko zentzuak erlazio lexiko-semantikoen bitartez lotuak. THESAURUSeko unitateak STRUCTURES azpi-EBan definitutako klaseen instantzia dira<sup>1</sup>.

STRUCTURESen definitzen dira HEBko egitura desberdinak. Horrela, esan daiteke STRUCTURES dela HEBan lexikoak duen antolamendu orokorraren isla: bertan sortzen eta definitzen dira beste azpi-EBetako unitateek heredatuko dituzten slot edo atributuak, bertan zehazten dira slot horiek fazeten bitartez<sup>2</sup>, bertan adierazten da slot horietako bakoitzaren herentzi modua edo slotean onar daitezkeen balioen klasea, eta abar. STRUCTURESen, bada, erazagutzen da DICTIONNAIRE eta THESAURUSeko unitate mota desberdinei buruzko eta THESAURUSeko unitateen arteko erlazioei buruzko metaezagutza, horrela HEBko ezagutzaren kudeaketa erraztuz. Beste aldetik, azpi-EB honetan erazagutzen dira eta definitzen prozedura desberdinak —iratxo eta metodoak, batik bat—, eta erlazioen propietateen ustiapena posible egingo duten erlazio-ezaugarriak; hori dela eta, STRUCTURES HIZTSUAren Dedukzio-Moduluarekin estuki lotua da, azpi-EB hori baita, nolabait, ezagutza deduktiboa erazaguturik daukan biltegia.

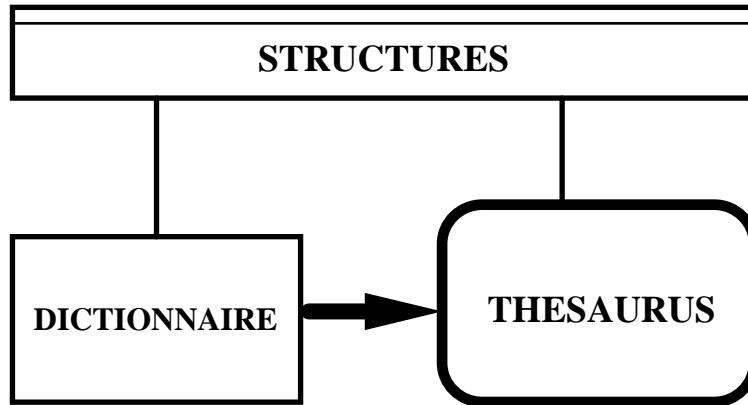
DICTIONNAIRE azkenik, erabiltzaileari hitzetik adierarako atzibidea eskaintzeko sortua da. Azpi-EB honetako unitateek hitzak errepresentatzen dituzte eta hitz horien adieraziak diren zentzu edo adierekin zuzenean loturik daude. DICTIONNAIREren funtzio nagusia beraz, HIZTSUA kontsultatzen denean beteko da: HEBko ezagutza lexikala zentzu edo adieretan oinarritzen bada ere, DICTIONNAIRE azpi-EBari esker erabiltzaileak bere eskariak hitzen funtzioan egin ahal izango ditu.

III.1 irudian ikus daiteke HEBa osatzen duten hiru azpi-EBen eskema simple bat. STRUCTURES da nolabait eskema osoaren "gainean" dagoena, THESAURUS eta DICTIONNAIREko unitateak azpi-EB horretan definitu eta zehaztutako klaseen azpiklase edo instantzia direna adierazi nahian edo. DICTIONNAIRE eta THESAURUSen arteko gezia DICTIONNAIREk THESAURUSeko kontzeptu-unitateetara iristeko atzibidea — hitzetik adierarako bidea— errepresentatzen duela adierazteko jarri da

---

<sup>1</sup> CONCEPTS-TYPE klasearen eta beronen azpiklaseen kasuan izan ezik, zeintzuekin THESAURUSeko unitateek klase/azpiklase loturak edukiko baitituzte.

<sup>2</sup> Fazetak slot edo atributuen atributu modukoak lirateke. KEEK ematen dio aukera erabiltzaileari bere fazetak definitzeko, atributuen zenbait aspektu hobeki zehaztearren.



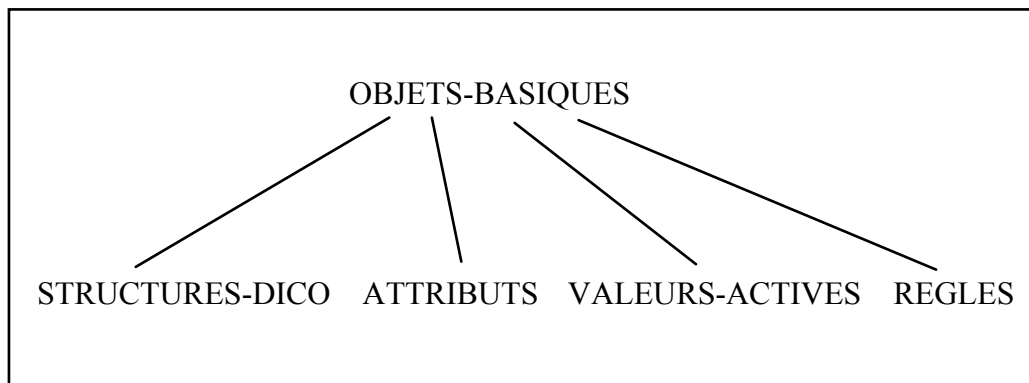
**III.1 irudia.-** HEBaren eskema orokorra.

Ondoko ataletan deskribatuko ditugu banan-banan eta zehazki azpi-EB hauek, beren ezaugarriak, unitateetan definituriko slotak, etab., errepresentazioari, oro har, dagokion mailan zein inplementatutako prototipoari dagozkion xehetasunen mailan.

### **III.1.3.2 Metaezagutza: STRUCTURES azpi-EBa.**

STRUCTURES azpi-EBa HEBko ezagutzen antolamendu hierarkikoaren isla da.

OBJETS-BASIQUES izeneko klase orokor baten azpian lau azpiklase aurkituko ditugu: STRUCTURES-DICO, ATTRIBUTS, VALEURS-ACTIVES eta REGLES (ikus III.2 irudia).



**III.2 irudia.-** OBJETS-BASIQUES klasea eta bere azpiklaseak STRUCTURES azpi-EBan.

STRUCTURES-DICO da THESAURUS eta DICTIONNAIREko unitateen zehaztapenak biltzen dituen klasea. Klase honetan —eta bere azpikoetan— deskribatzen dira HEBko egitura desberdinak eta egituron portaera objektuei zuzendutako programazioaren ikuspegitik. Esan daiteke STRUCTURES-DICO datu-mota definizio multzo bat dela.

ATTRIBUTS klaseak STRUCTURES-DICOn eta bere azpiklaseetan definitutako atributuen sailkapenari buruzko informazioa biltzen du. ATTRIBUTS klase honen azpian daude sailkaturik, besteak beste, erlazio lexiko-semantiko desberdinak.

VALEURS-ACTIVES klaseak berriz, STRUCTURES-DICOn eta bere azpian definitutako slotei loturiko "iratxo" erako prozedura-instantziak<sup>1</sup> edukiko ditu.

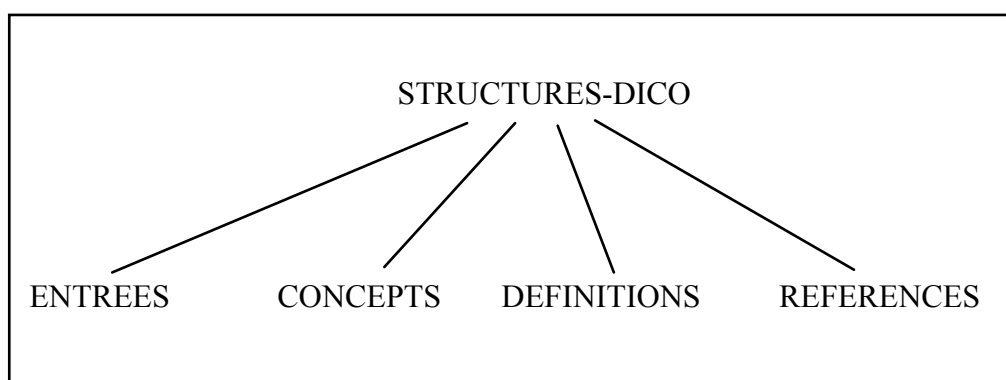
Azpi-EB honetan kokatu ditugu dedukzio-prozesu desberdinetan erabili beharreko erregelak ere<sup>2</sup>. REGLES klasean bada, sailkatzen dira erregelak azpiklase desberdinetan, azpiklase horien instantziak, azken finean, erregela banakoak direlarik.

Azpi-EB honen KEE listatu osoa txosten honen D eranskinean ikus daiteke.

### III.1.3.2.1 Hiztegi eta thesauruseko egituren klasea: STRUCTURES-DICO.

STRUCTURES-DICO klaseak islatzen du DICTIONNAIRE eta THESAURUS azpi-EBetako unitateen antolamendu eta hierarkizazioa.

STRUCTURES-DICO horren azpiko klaseetan inportanteena CONCEPTS delakoa da (ikus III.3 irudia). Klase honen azpian CONCEPTS-TYPE, CONFIGURATIONS eta CONCEPTS-AMBIGUS azpiklase lagungarria daude (ikus III.4 irudia). Horiexek dira HEBko klase "kontzeptualak", klase horietako edo berorien azpiklaseetako elementuak kontzeptuen errepresentazio diren *kontzeptu-unitateak* baitira (hiztegiko adiera edo THESAURUSen sorturiko bestelako unitate kontzeptualak).



**III.3 irudia.-** STRUCTURES-DICOn azpiko egitura hierarkikoaren partezko ikuspegia.

<sup>1</sup> Hemendik aurrera "instantzia" zein "elementu" erabiliko ditugu klase/elementu edo instantziario erlazioaz ari garenean (multzoetako barnekotasunaren pareko), "azpiklase" berriz, klase/azpiklase erlazioaren kasurako utziz (partekotasuna).

<sup>2</sup> Egia esateko, erregelaren sailkapenak merezi lezake aparteko azpi-EB berezi bat. Base berri horrek edukiko lukeen ezagutza —logikoki Dedukzio-Moduluari legokiokeen ezagutza deduktiboa— STRUCTURES azpi-EBtik fisikoki berezia geratuko litzateke honela.

### III. kapitulua

Klase horiez gain, STRUCTURES-DICoren azpian ditugu ENTREES, bere elementuak DICTIONNAIRE azpi-EBan dituen hiztegi-sarreraren klasea, adierak beren kategoria sintaktikoaren arabera sailkaturik dituen DEFINITIONS klasea, eta REFERENCES, hasiera batean definiziorik ez izan arren beste kontzepturen baten definizioan agertu direlako THESAURUSen sortutako kontzeptuak biltzen dituena.

CONCEPTS azpiko CONCEPTS-TYPE klasea Quillian-en (1968) "type nodes" direlakoei dagokie. CONCEPTS-TYPE klase horretan lau azpiklase ageri dira hasiera batean —geroago findu beharreko sailkapena—: ENTITES, ACTIONS/EVENEMENTS, QUALITES eta ETATS, izen, aditz, adjetibo eta adberbioei, hau da, kategoria sintaktiko "irekiei" gutxi gorabehera dagozkienak (ikus III.4 irudia). THESAURUSen errepresentatutako kontzeptu oro azpiklase hauetakoren batean egongo da —klase/azpiklase zuzeneko lotura baten bitartez, edo zeharka, konfigurazioen kasuan bezala—, nahiz eta, agian, beste klaseren bateko elementu ere izan (bereziki, egon daitezke CONCEPTS-AMBIGUS edo REFERENCES klaseko elementuak direnak). Beren definizioa HDBan analizaturik daukaten kontzeptuak DEFINITIONS klaseko lau azpiklaseetariko batean egongo dira gainera.

Horretara, THESAURUSeko kontzeptu-unitateak, CONCEPTSeko azpiklaseren bateko edo batzuetako elementu izateaz gain, beste "ikuspuntu" batetik kontzeptu definitu edo erreferentzia bezala ere sailkatzen dira (DEFINITIONS edota REFERENCES klaseetan), thesaurusean egoteko arrazoia HEBaren iturburu-hiztegiko azpimultzoko sarrera izateagatik den ala beste adieraren baten definizioan agertu izanagatik den (berezko definiziorik gabe baina). Kontzeptu-unitateei buruzko "ikuspegi" desberdin horiek —klase bat edo gehiagoren kide izate horrengatik— ezagutzaren egituratze eroso eta dotorea egiteko bidea ematen digute. Esate baterako, azpiklase eta elementuek heredatuko dituzten atributuak —slotak— sortu eta definitzerakoan, edota klase bakoitzeko elementuen portaera —metodo, slotei lotutako iratxo, eta abarren bitartez— zehazterakoan.

III.4 irudian ikus daiteke CONCEPTS klasearen azpiko hierarkia. Esan bezala, THESAURUSeko kontzeptu-unitate guzti-guztiak daude, zuzenean nahiz zeharka, CONCEPTS-TYPE klasean (ikus aurrerago mota-kontzeptuoi eta konfigurazioei buruz dioguna). Ikus dezagun, CONCEPTS-TYPE klasera itzuli baino lehen, CONCEPTS-AMBIGUS klasean egindako kontzeptu anbiguo edo partez desanbiguatuen sailkapena.

CONCEPTS-AMBIGUS azpiklase lagungarria HDBan burutu analisi-prozesuetatik desanbiguateke iristen diren kontzeptuak biltzeko erabiltzen dugu. THESAURUSeko zenbait kontzepturen beste "ikuspegi" bat edukitzea lortzen dugu honela. Lexikalki anbiguo diren atributu-balioak —kontzeptuen erreferentzia, azken finean— kontzeptu anbiguoen

sorkuntzarako mekanismo baten bitartez tratatu ditugu: sarreraren zein adierari dagozkion garbi ez dagoen hitzen agerpenen kasua litzateke<sup>1</sup>. Horrelakoak THESAURUSen "desanbiguatzeke" uztea erabaki dugu —kontzeptu anbiguoak sortu eta (Vossen & Serail, 90)en egiten denaren antzera identifikatuz—, ondoren HEBan egingo litzatekeen desanbiguatze lexikaleko prozesu baten zain. Adibidez, HEBan gero eta aberatsago izango bide den ezagutzan oinarrituz, erreferentzi balio anbiguo horien desanbiguaketan parte hartuko duten irizpide heuristikoak definitu eta aplikatu ahal izango dira.

Kontzeptu hauek hiru azpiklasetan sailkatu ditugu, daukaten anbiguetate mailaren arabera:

- **HOMOGRAPHE**: azpiklase honetan homografo mailako anbiguetatea duten kontzeptuak egongo dira eta, identifikadore gisa, kontzeptuaren deitura adierazten duen sinboloan bi galdera-marka jarriko ditugu. Adibideak: |entretenir ? ?|, |faculté ? ?|<sup>2</sup>.
- **ACCEPTION**: klase honek adiera mailako anbiguetatea daukaten kontzeptuak bilduko ditu, kontzeptu horiek identifikatzeko adiera-zenbakiari dagokion lekuan galdera-marka bat erabiliko dugularik. Adibideak: |plaie I ?|, |panser I ?|, |ouvrage I ?|.
- **COMPLEXES**: dudazko desanbiguaketa jasan duten kontzeptuak egongo dira azpiklase honen barruan, hau da, bi edo adiera gehiagoren arteko zalantzaren bat dagoeneko kasuak. Anbiguetatea errepresentatzeko kasu hauetan, kontzeptuaren sinboloan adieren arteko barra okerrak erabiliko ditugu: |instrument I 1/2|, |faire I 2/1|, |donner I 5/6|.

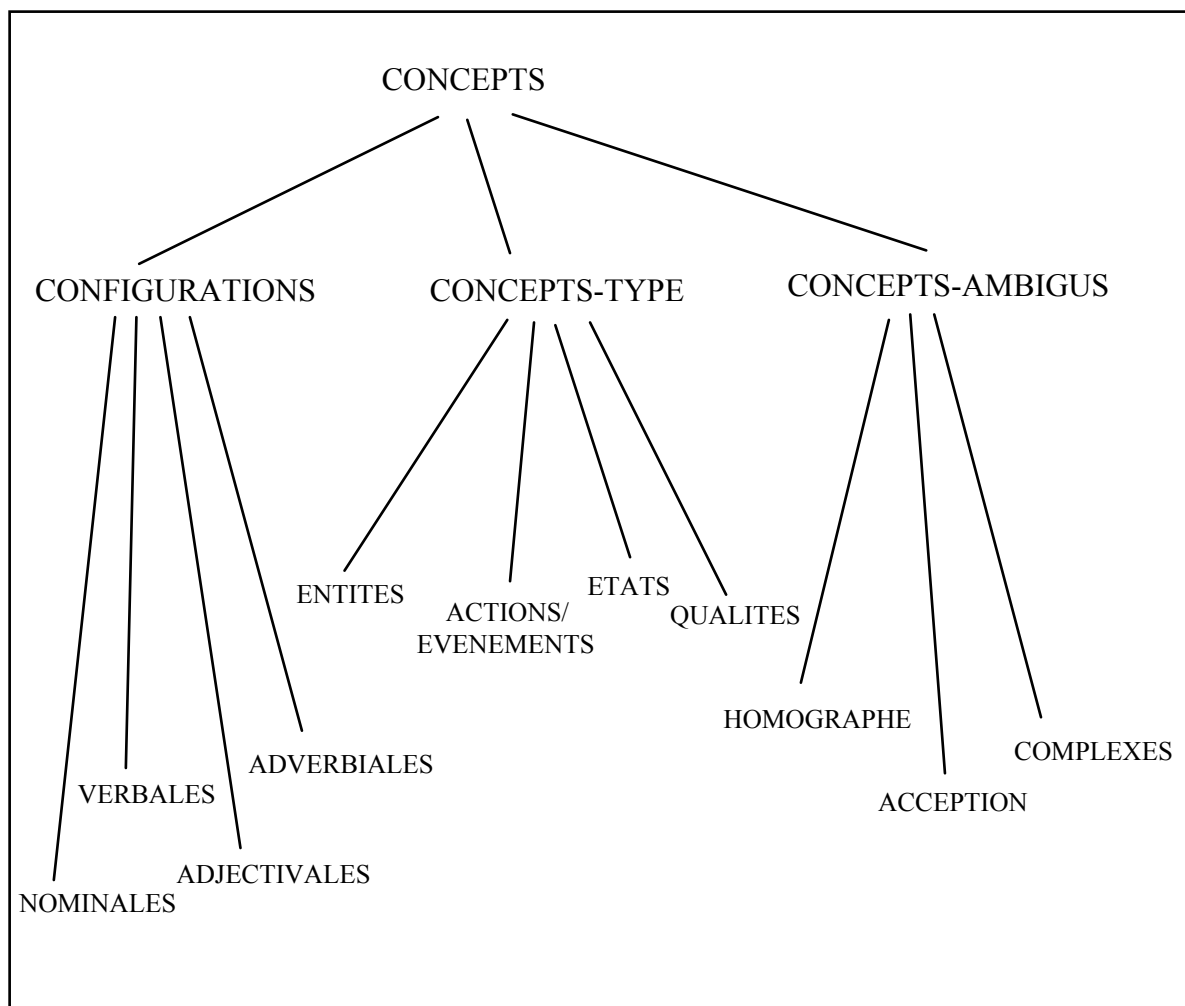
Honela, THESAURUSeko unitate kontzeptual batzuk, anbiguetateren bat dutenak alegia, kontzeptu anbiguoen klase honetako elementu izango dira<sup>3</sup>. Horrek emango digu aukera gero, HIZTSUAKo prozesu deduktiboak martxan jartzeko garaian, horrelako kontzeptuen portaera berezia zehaztu ahal izateko.

---

<sup>1</sup> Zenbait lanetan (Vossen & Serail, 90), kasu anbiguo horien gehiengoan lehen adiera izan ohi dela esaten dute. Definizio batean anbiguo geratutako hitz-agerpen guztiei lehen adiera egokitzea ez da, guk LPPLn ikusi dugunez behintzat, inondik ere egokiena: oso maiz gertatzen baita hitz baten lehen adiera ez dela beste definizioetan gehien erabilitakoa, askotan bigarrena edo hirugarrena izaki (eta hori "metalingoaiazko" espresiotzat jotakoen artean ere egiaztatu dugu).

<sup>2</sup> Hemendik aurrera kontzeptuak errepresentatzeko dagokien LISP sinboloa erabiliko dugu. Adibidez, **chien I 1** hitz-adierari dagokion sinboloa hauxe da: |chien I 1|.

<sup>3</sup> Kontzeptu anbiguok sailkatzen dira, beste kontzeptu guztiak bezalaxe, CONCEPTS-TYPE klasearen azpian ere, baina, anbiguo diren bitartean, kontzeptu anbiguoen klase hauetakoren baten elementu ere izango dira.



**III.4 irudia.-** CONCEPTS klasea STRUCTURES azpi-EBan.

CONCEPTS-TYPE klasera itzuliz, esan dezagun berriro Quillian-en "type node" direlakoan kide ditugula. Hasiara batean, CONCEPTS-TYPE klasea lau azpiklase handitan dago banatua, banaketa hori nahiko "lañoki" egina delarik. ENTITES klaseak adierazletzat izena duten kontzeptuak biltzen ditu; ACTIONS/EVENEMENTS aditz-kontzeptuentzat dago pentsatua, QUALITES adjetibozko kontzeptuentzat eta ETATS adberbiozko adierazlea duten kontzeptuentzat. Orain deskribatu klaseetan ez bezala, THESAURUSEko kontzeptu-unitateak klase kontzeptual hauen azpiklase ditugu —eta ez elementu—. Honi esker lortzen dugu, datu-egituren herentzia (slotak, balio-eremuak, eta abar) ez ezik, THESAURUSEko kontzeptuen artean erlazio taxonomikoen arabera —klase/azpiklase lotura beraren bitartez implementatutako hiperonimia eta hiponimia— ezarritako hierarkian balioak goiko mailetatik beheragokoetara pasatzea. Horrela, esan genezake CONCEPTS-TYPE nodoa THESAURUSEko kontzeptu guztien gainklase moduko bat dela.

CONCEPTS-TYPE klase honetan eta bere azpiklaseetan definitzen dira *atributu erlazionalak*, hau da, sareko kontzeptuen arteko erlazio lexiko-semantikoen implementazioa

diren atributuak. CONCEPTS-TYPE klasean egindako banaketari esker azpiklase bakoitzean atributu desberdinak sortzen ahal ditugu, atributu bakoitzaren balio-eremua —mota, edo atributuak adierazi erlazioaren kodomeinua—, herentzi rola eta portaera prozedurala zehaztuz. Esate baterako, *kasu-atributuak* aurkituko ditugu ACTIONS/EVENEMENTS klasean definiturik, edo CARACTERISTIQUE eta ORIGINE bezalakoak ENTITESen. Azpiklase guztietan aurkituko ditugu *metalengoaiazko atributuak*, hiztegiko definizio-era desberdinetatik eratorriak direnak: PARTIE-DE, CE-QUI edo MANIERE-DE ENTITESen, RENDRE edo CESSER-DE ACTIONS/EVENEMENTS klasean. Atributu erlazional guzti hauei loturik erlazioaren alderantzizkoa automatikoki kudeatzen duen iratxoa dago (erlazio bakoitzari dagokion alderantzizko erlazioa zehazteko atributua sortzen den lekuan egin behar da, horretarako bereziki definitutako fazeta batean).

*Mota-kontzeptuen hierarkia* hori beste lan batzuetan azaltzen diren mota-hierarkien kidakoa litzateke (Copestake, 92). Hala eta guztiz ere, bistakoa da gure mota-hierarkia ez dela inola ere hierarkia konplexu bat, baizik, esan dugun bezala, nahiko hierarkia "lañoa". Mota-hierarkia bat eskuz eraikitzea —aipatu berri ditugun lanetan bezala— ez da, izan ere, txantxetako: (Copestake, 92)n —ACQUILEX proiektua— autoreak 450 motako hierarkia azaltzen du, mota horietan 80 ezaugarri desberdin definitu direlarik; hierarkia horrek aukera ematen dio, noski, ezagutza lexikala era "finago" batez adierazteko. Ez da ziur, hala ere, gure ustez, horrelako hierarkia bat hiztegi arrunt batetik ateratako informazioarekin soilik eraiki daitekeenik: argi dago lan hori burutzeko premiazkoa dela nolabaiteko ezagutza orokorra. Alabaina, guk pentsatzen duguna zera da, alegia, gure proiektuan eraikitako hasierako hierarkia simple hori aurrerago fintzeko aukerak izango direla —era erdiautomatikoz, seguruenik—, hiztegi-ezagutzarekin egindako inferentziei esker, eta batez ere, ezagutza horren alderdi taxonomikoaren gainean egindakoen bitartez.

Bai (Yoshida *et al.*, 82)n eta bai (Wilks *et al.*, 89)n ere, kategoria sintaktiko ireki bakoitzerako —izen, aditz, adjetibo eta adberbioak— frame-egitura desberdin bat aukeratzen dute<sup>1</sup>. (Basili *et al.*, 92)n aukeratutako kategoria semantikoen multzoak ere sinpletasun handia erakusten du, 13 kategoria besterik ez baititu<sup>2</sup>. Kasu horretan, corpus bateko hitzak eskuz etiketatu dira, hautapen-murriztapenei buruz ("selectional restrictions") egindako azterketa —hitzen arteko elkarketan oinarritua— izanik helburu.

---

<sup>1</sup> (Wilks *et al.*, 89)n oso garrantzizkotzat hartzen dituzte kasu-rolak aditzen frameetan; izenitarako berriz, erreferenteen propietateak eta osagaien zehaztapena jotzen dituzte ezinbestekotzat. Lehen ere esana dugu (ikus III.1.2) (Yoshida *et al.*, 82)n kontzeptuak lau klasetan sailkatzen dituztela: objektu konkretuak, gertakariak, atributu edo egoerak, eta izaki abstraktuak, bakoitzarentzat frame desberdina definituz.

<sup>2</sup> PHYSICAL\_ACT, MENTAL\_ACT, HUMAN\_ENTITY, ANIMAL, VEGETABLE, MATERIAL, BUILDING, BY\_PRODUCT, ARTIFACT, MACHINE, PLACE, QUALITY eta MANNER.

### III. kapitulua

III.4 irudira itzuliz, CONFIGURATIONS izeneko klasea geratzen zaigu azaltzeko. *Konfigurazioen* sorburua Quillian-en (1968) "token"ak —mota-kontzeptuen agerpenak definizioetan— errepresentatzeko egitura baten premia izan da. Gurean, konfigurazioak *zenbait kontzepturi dagokion eduki semantikoa duten eraikuntza sintagmatikoen adierazpide* ditugu. Horren arabera, HEBaren hasierako eraikuntzan, konfigurazio berri bat sortuko da definizioaren analisisan eduki semantikodun kontzeptu bat baino gehiagori dagokion sintagma egitura bat ezagutzen denean: izena eta adjetibo zenbait duen izen-sintagma, osagaiekiko aditz- edo adjetibo-sintagma, etc. Konfigurazio bat beti bere hun-kontzeptu edo buruaren azpiklase —hiponimo— bezala sortuko da (izen-sintagmako izena, aditz-sintagmako aditza, etc.). Determinazioa edota generoak, izen-sintagmen kasuan, edo erreflexibotasuna, modua, aspektua, denbora eta pertsonak aditzezkoetan, ez dituzte berez konfigurazioak sortaraziko<sup>1</sup>, loturak zuzenean mota-kontzeptuarekin egingo direlarik.

Konfigurazio mota desberdinetan slotak berezitu ahal izatearren konfigurazioak ere lau azpiklasetan banatu dira: NOMINALES, VERBALES, ADJECTIVALES eta ADVERBIALES.

NOMINALES klasean A, AVEC, CARACT/ETAT, CARACTERISTIQUE, DE, EN, POUR, PREDICATION eta SANS *atributu sintagmatikoak* eta DCEQUI eta PARTIED *metalengoiatzkoak* definitu dira. Konfigurazio mota hau izen-sintagmen errepresentaziorako dago pentsatua.

VERBALESen —aditz-sintagmen adierazpide izateko sortua— izenezko kontzeptu-unitateekiko erlazioak errepresentatzen dituzten atributu sintagmatikoak definitu dira: A, ADVERBE, ATTRIBUT, AVEC, DE, EN, OBJET eta POUR. Horiekin batera metalengoiatzko atributu bat: PARENTHESE.

ADJECTIVALESen GRADUATEUR eta PARTICIPED atributu sintagmatikoak, eta ADVERBIALESen GRADUATEUR atributu sintagmatikoa besterik ez da definitu.

Konfigurazioen ideia honek badu bere erreferentzi puntua P. Vossen-en lanean ere (1989), hiztegi-definizioen sistemak ezarritako erlazioak sarrera eta definizioetako hitzen artekoak baino sarrera eta hizkuntz espresioen artekoak direla dioenean. Eta hori, esateko, inferentzia sistematikoak egin nahi baldin badira, definizioetako hitzei berei adina hitz horiek definizio-espresioetan betetzen duten funtzioari egin beharko diogula kasu. Konfigurazioak dira, gure eremuan, hizkuntz espresio horien errepresentaziorako sortutako baliapiderik nagusiena.

---

<sup>1</sup> Hala ere, definiziotik datorren informazio hori ez da galduko HEBko adierazpidean: informazio hori gordetzeko slotetan propio definitutako fazetak —atributu edo sloten edukia hobeki zehazteko modua— erabiltzen ditugu.



Azpi-EB honen deskribapenari bukaera emateko, hiru azpiklase geratzen zaizkigu oraindik hobeto azaltzeko: DEFINITIONS, REFERENCES eta ENTREES.

DEFINITIONS klaseak, bere azpiklase diren NOMS, VERBES eta ADJECTIFSeKin, THESAURUSen LPPLtiko definizioa errepresentaturik daukaten kontzeptuak kategoriaren arabera atzitzeko aukera ematen digu. Klase horretan definitu dira *informazio orokorreko atributuak*, HDBtik datorren informazioa edukitzeko sortuak: CATEGORIE-SYNTAXIQUE, GROUPE-CATEGORIEL, REGLE-ANALYSE, TEXTE-DEFINITION, TEXTE-EXEMPLES eta UTILISATION. Atributu horiek guztiak DEFINITIONSen instantzia diren unitateek klase horren azpiklaseen bitartez heredatuko dituzte; horrela atributuok fintzeko aukera, balio lehenetsien mailan adibidez, egongo da.

DEFINITIONS klase honetan eta berorren azpiklaseetan definitu ditugu, baita ere, *atributu definitorioak*, LPPLko azterketan aurkitutako definizio-era desberdinak errepresentatzeko sortuak: definizio-era orokorren adierazpide diren DEF-CLASSIQUE edo DEF-SYNONYME, adibidez, DEFINITIONS goiklasean definiturik dauden bitartean, azpiklaseetan definitu dira era espezifikagoak errepresentatzen dituzten atributuak.

NOMSen bada, atributu definitorio hauek daude: DEF-ACTIOND, DEF-ARTD, DEF-CARACTEREDCQE, DEF-CEQUI, DEF-CEQUON, DEF-EGSD, DEF-EMD, DEF-ERGD, DEF-FACULTED, DEF-MANIERED, DEF-MPPD, DEF-PARTIED, DEF-PIECED, DEF-QUALITEDCQE eta DEF-QUI, identifikadoreek berek adierazten dituzten definizio-era bereziei dagozkienak<sup>1</sup>.

VERBES klasean, honako hauek: DEF-CESSERD, DEF-COMMENCERA, DEF-DEVENIR, DEF-DNOUVEAU, DEF-ETRE, DEF-ETREN, DEF-FAIRE, DEF-METTREN, DEF-NEPAS eta DEF-RENDRE.

Eta ADJECTIFSen berriz, honako hauek: DEF-ANTONYME, DEF-CONFORMEA, DEF-CONTRAIREA, DEF-DONT, DEF-PEU, DEF-QUI, DEF-QUIAD, DEF-QUIEST, DEF-QUINE, DEF-RELATIFA, DEF-SANS eta DEF-TRES.

REFERENCES klaseko elementuak ere THESAURUS azpi-EBko unitate dira eta, gorago esan bezala, elementuok CONCEPTS-TYPEren azpiklase baina DEFINITIONSen elementu ez diren kontzeptuak dira<sup>2</sup>: iturburu-hiztegitiko definizio-informaziorik ez duten

---

<sup>1</sup> DEF-EGSD: '*espèce de*', '*genre de*' edo '*sorte de*' formularekiko eskema duten definizioetarako.  
DEF-EMD: '*élément de*' edo '*membre de*' formularekiko eskema duten definizioetarako.  
DEF-ERGD: '*ensemble de*', '*réunion de*' edo '*groupe de*' formularekiko eskema duten definizioetarako.  
DEF-MPPD: '*manque de*', '*perte de*' edo '*privation de*' formularekiko eskema duten definizioetarako.

<sup>2</sup> Ongi bereizi beharra dago alde batetik "erreferentzia" hauek, besteren definizioetan agertu izanagatik THESAURUSen sortutako kontzeptuak direnak, eta bestetik, porrotezko analisia izan duten kontzeptuak, hauek ere "hutsak" THESAURUSen. Azken horiek, nahiz eta ez eduki berez definiziotiko informaziorik HEBan, ez baitira REFERENCES klaseko izango baizik DEFINITIONSekoak (berorren azpiklaseren batekoak).

### III. kapitulua

erreferentzia edo "kontzeptu hutsen" kasua dugu. Interesgarria litzateke ikustea ea HIZTSUA tresna baliagarria den THESAURUSera "hutsik iristen" diren kontzeptu hauei buruzko ezagutzak deduzitzeko garaian.

ENTREES klaseko instantzia diren unitateak THESAURUSekoak gabe DICTIONNAIRE azpi-EBko elementu ditugu, orain arteko klase guztietakoak ez bezala. Oso egitura sinplea duen framea erabiliz —izan ere ENTREESeko instantzia baten framean SENS izeneko slota besterik ez baitago, eta slot horrek frameak errepresentatzen duen hitzari dagozkion adieren lista dauka— lortzen da hiztegi-sarrera den hitzaren eta bere zentzu edo adieren arteko lotura esplizitu egitea.

STRUCTURESen definituko dugu HEBko klaseei eta beren instantziei dagokien portaera. STRUCTURESeko klaseek beren slot bereziak dituzte: slot horiek sortzerakoan beren balio-eremua eta azpiklase edota instantziei balioak nola heredatuko zaizkien zehaztu behar da; era berean sloten balio lehenetsiak eman, eta balioen irakurketa edota eguneratze unean exekutatu diren iratxoen esleipena ere egin behar izaten da. Goian aipatu atributu erlazionalek errepresentatutako erlazio lexiko-semantikoen propietateak ere slotek dituzten fazetetan deskribatzen dira.

D eranskinean ikus daitezke STRUCTURES azpi-EB honetako klase bakoitzean definitutako atributuak eta beren zehaztapenak.

#### III.1.3.2.2 Atributuen sailkapena: ATTRIBUTS. Erlazio lexiko-semantikoei buruzko metaezagutza.

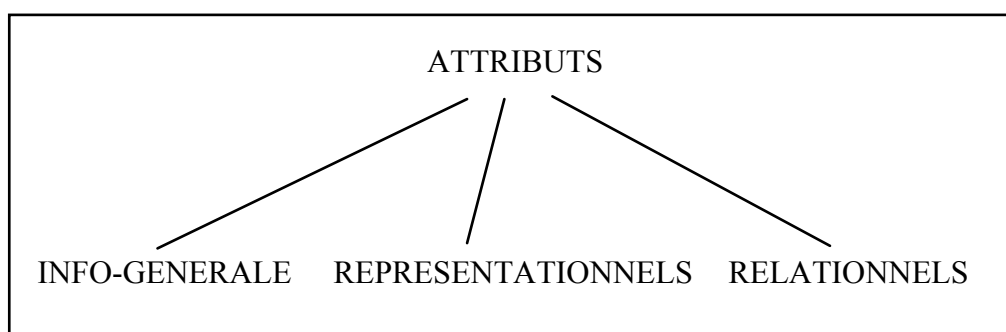
STRUCTURES azpi-EBko ATTRIBUTS klasean sailkatu ditugu HEBan erabilitako atributuak (azpi-EB honetako STRUCTURES-DICO klasearen azpian erabili eta definituak batez ere). Atributu mota desberdinei buruzko metaezagutza den hori informazio-iturritzat erabiltzen da (atributu bat erlazionalen klasekoa denentz jakin, etab.). Atributu oro da ATTRIBUTS klaseko azpiklaseren baten barnekoa; barnekotza horren xehetasunak atributu bakoitza definitzerakoan eman behar dira, horretarako *CLASSE-ATTRIBUT* izeneko fazeta erabiliz.

Frame erako egiturak darabiltzan ezagutzaren errepresentaziorako sistema batean, slotak edo atributuak —hemendik aurrera *atributu*— unitate desberdinen ezaugarriak deskribatzeko erabiltzen dira. Atributuok eduki ditzakete zenbakizko balioak zein testuzkoak, baina baita askoz informazio konplexuagoak ere: taulak, irudi-egiturak, programak eta, guretzat interes handiena daukana, frame-sistemako beste unitateen erreferentziak. Klase batean zenbait

atributu definitu eta sortzeak<sup>1</sup> klase hori bere atributuak bere azpiklase eta instantziei heredatzen dizkien eskematzat hartzeko aukera ematen digu.

HEBko kontzeptuen errepresentaziorako hiru atributu mota nagusi definitu ditugu: INFO-GENERALE, REPRESENTATIONNELS eta RELATIONNELS izenekoak (ikus III.5 irudia). Aurrera baino lehen esan behar dugu HEBan egiten den errepresentazioa, adiera edo zentzuei dagokienean batik bat, *bi mailatako errepresentazio* bat dela<sup>2</sup>:

- **Maila definitorioa:** maila honetan errepresentatzen da, horretarako espreski sortutako atributuen bitartez, adiera edo zentzuari dagokion definizioa, atributu horien fazetetan gordez (ikus aurrerago) determinazioa, modua, denbora, pertsona eta abar bezalako xehetasun morfosintaktikoak. Maila honetan, zentzuaren errepresentazioa *errepresentazio-atributuei* dagokie.
- **Maila erlazionala:** maila hau litzateke gorago hainbatetan aipatu dugun lexikoaren ikuspegi erlazionala emango lukeena. Errepresentazio maila honen gainean ezarri dugu sistemaren portera deduktiboa. Maila hori *atributu erlazionalari* dagokie, zeintzuek, errepresentazio-atributuek ez bezala, ezagutza deduzitua edukitzeko ahalmena baitaukat.



**III.5 irudia.-** Atributuen sailkapena STRUCTURES azpi-EBan<sup>3</sup>.

*Informazio orokorreko atributuak*, INFO-GENERALE klasearen barneko, orotariko informazioa gordetzeko sortuak dira (HDBtik edo beste inondik etorritako informazioa).

<sup>1</sup> Adibidez, ENTITES klasean PARTIE-DE atributua definitu baldin badugu 'partie de' erlazioa klase horretako elementuen ezaugarri izan daitekeelako da. Atributu horrek, bere CLASSE-ATTRIBUT fazetan METALANGAGIERS2 balioa dauka, atributua klase horretakoa, eta beraz baita RELATIONNELSeko ere (METALANGAGIERS2 RELATIONNELSen azpiklasea da), dela adieraziz. Erlazio meronimiko hori areago definitzeko PARTIE-DE atributuari loturiko fazetaz erabiliko dira: adibidez, erlazioaren alderantzizkoa PARTIE-DE+INV dela fazeta berezi batean erazagutuko da.

<sup>2</sup> Hobeto esanda, hiru mailaz hitz egin behar genuke, HEBan "ohizko" adierazpidea ere badagoela kontuan harturik. HDBan dagoen informazioaren parte handi bat HEBra ere badator, bere errepresentaziorako *informazio orokorreko atributuak* erabiliz (ohizko moduan): definizioa testu eran, sarreraren kategoria sintaktikoa, adibideak, definizioaren analisiaren emaitzak, eta abar.

<sup>3</sup> ATTRIBUTS OBJETS-BASIQUESen azpiklasea da (ikus III.2 irudia).

### III. kapitulua

Horren adibide HDBan analizatutako sarrerei dagokien informazioa: informazio hori gordetzeko CATEGORIE-SYNTAXIQUE, TEXTE-DEFINITION, TEXTE-EXEMPLES, etab. bezalako informazio orokorreko atributuak definitu dira, HDBtiko informazio horiek, era "konbentzionaletan" izan arren, HEBak ere eduki ditzan sortuak. Atributu horiek DEFINITIONS klasean definitu dira eta, horrela, klase horren elementuek heredatuko dituzte.

*Errepresentazio-atributuen* artean hiru mota bereizi behar dira: atributu definitorioak, sintagmatikoak eta metalengoaiatzkoak, REPRESENTATIONNELSen azpiklase diren DEFINITOIRES, SYNTAGMATIQUES eta METALANGAGIERS1 klaseetan bilduak, hurrenez hurren (ikus III.6 irudia).

*Atributu definitorioak* ere —DEFINITOIRES azpiklasekoak— DEFINITIONS eta bere azpiklaseetan kokaturik daude, eta LPPLren azterketan aurkitutako definizio-era desberdinak errepresentatzeko sortuak dira. Atributu horien bitartez —besteak beste— errepresentatuko dira hasiera batean HDBtik HEBra ekarritako definizioak. Atributu definitorio hauen deskribapenean aurkituko dugun ezaugarri inportante bat bakoitzari dagokion erlazionalaren zehaztapena izango da ("kideko erlazionalaz" hitz egingo dugu engoitik); zehaztapan hori atributu hauek duten *RELATIONNELS-CORRESPONDANTS* izeneko fazetan egingo da: adibidez, DEF-QUIAD atributu definitorioaren kidekoa QUI-A-DE atributu erlazionala da. Erreferentzia zehatz honi esker, eta atributu definitorioari loturiko iratxo baten bitartez, lortuko dugu DEF-QUIAD atributuaren balioa QUI-A-DE atributura automatikoki pasaraztea (atributu hori beterik duen THESAURUSeko unitate baten barruan).

*Atributu sintagmatikoak* —SYNTAGMATIQUES azpiklasekoak— THESAURUSeko unitateen arteko erlazio sintaktikoak errepresentatzeko erabiltzen ditugu; horien artean ditugu, adibidez, preposizioei dagozkien atributuak, THESAURUSeko unitateak, (balizko) desanbiguatze semantiko bat egiten ez den bitartean, erlazionatzen dituztenak. Honetan ere baliook atributu sintagmatikoetatik dagozkien erlazionaletara *goratzeko* modu berbera erabiltzen da: atributuari lotutako iratxo batek —atributua sortutako lekuan— egingo du balio-intsuldatzea automatikoki. Kasu honetan baina, gauza ez da hain erraza, zeren AVEC bezalako atributu sintagmatiko baten azpian hainbat erlazional egon baitaiteke, sakonean egon daitekeen erlazio semantiko adina hain zuzen: INSTRUMENTAL, COMITATIF, MANIERE, etc. Beraz, intsuldaketa egin aurretik preposizioon desanbiguatze semantikoa egin beharko litzateke. Desanbiguatze horretarako —HEBaren aberasketa-fasean egingo litzatekeena— *fazeta erlazionalak* sortu ditugu; fazeta erlazional horiek atributu sintagmatikoetan definitzen dira, atributu erlazionalen erreferentzia diren deiturak dituzte<sup>1</sup> eta MYCINen

---

<sup>1</sup> Fazeta horiek dagozkien atributu erlazionalen izen berekoak ditugu: *INSTRUMENTAL*, *COMMITATIF*, *MANIERE*, etc.



### III. kapitulua

gure sisteman —errepresentazioaren maila definitorioan— nola adierazi diren ulertzeko garrantzi handia duen zenbait. Izan ere, aipatu aspektu horiek maiz samar bazterrean geratu izan baitira sare tipoko errepresentazioetan: esate baterako, kuantifikazioa litzateke hizkuntz aspektu horietako bat. Gure ereduak bide bat eskaintzen du, maila definitorioan, horrelako informazioa jasotzeko, horretarako sortu eta definitutako fazeta berezietan. HIZTSUAN dedukzio-prozesuak egin behar direlarik, maila definitorioan jasotako informazio hori hor egongo da beti erabilgarri, nahiz eta maila erlazionalan ez esplizitua izan.

Definizioetako izen-sintagmen ezaugarri diratekeen determinazioa, generoa eta numeroa, edo aditz-sintagmen erreflexibotasuna, modua, denbora eta pertsona, maila definitorioko atributuetan zehazteko *fazeta morfosintaktikoak* erabiltzen ditugu. Fazeta horiek atributuek har ditzaketen balioen arabera —atributuen balio-eremua— agertuko dira ala ez atributuon deskribapenean: izenezko balioak har ditzakeen atributuak —NOMS edo NOMINALES klaseetako elementuen erreferentzia, edo oro har, ENTITESeko azpiklase— determinazio, genero eta numeroari dagozkion fazetak izango dituen bitartean, ACTIONS/EVENEMENTSen eremuko balioak har ditzakeen atributuak aditz-modu, -denbora, -aspektu eta pertsonari buruzko fazetak edukiko ditu<sup>1</sup>.

*Atributu erlazionalak* azkenik —RELATIONNELS klasekoak—, THESAURUSeko kontzeptuen arteko erlazio lexiko-semanticoen inplementazio dira.

Atributu hauek CONCEPTS-TYPE ete bere azpiklase diren ENTITES, QUALITES, ACTIONS/EVENEMENTS eta ETATSen daude definituta. Gogorarazi behar da hemen konfigurazioak, definizioz mota-kontzeptu baten azpiklase direnez, CONCEPTS-TYPEko azpiklase ere badirela eta, beraz, dagozkien atributu erlazionalak ere heredatuko dituztela, beren erlazioen ustiaketa mota-kontzeptuak bailiren egiteko bidea emanaz. Atributu erlazional hauei lotuta zenbait iratxo egongo da, errepresentatzen dituzten erlazio lexiko-semanticoen propietateak, erlazio alderantzizkoak, eta abar inplementatzeko bitarteko gisa. Gorago ikusitako atributuekiko alderik nabarmenena erlazionalotan ezagutza deduzitua gordetzeko aukera da, hau da, HDBan esplizitua ez izanik HEBan martxan jarritako prozesu deduktiboei esker lortutako ezagutza bertan edukitzeko aukera. Halaber, atributu hauetan jasoko dira, adibidez, atributu sintagmatikoetako erreferentziak desanbiguatu ondorengo emaitzak.

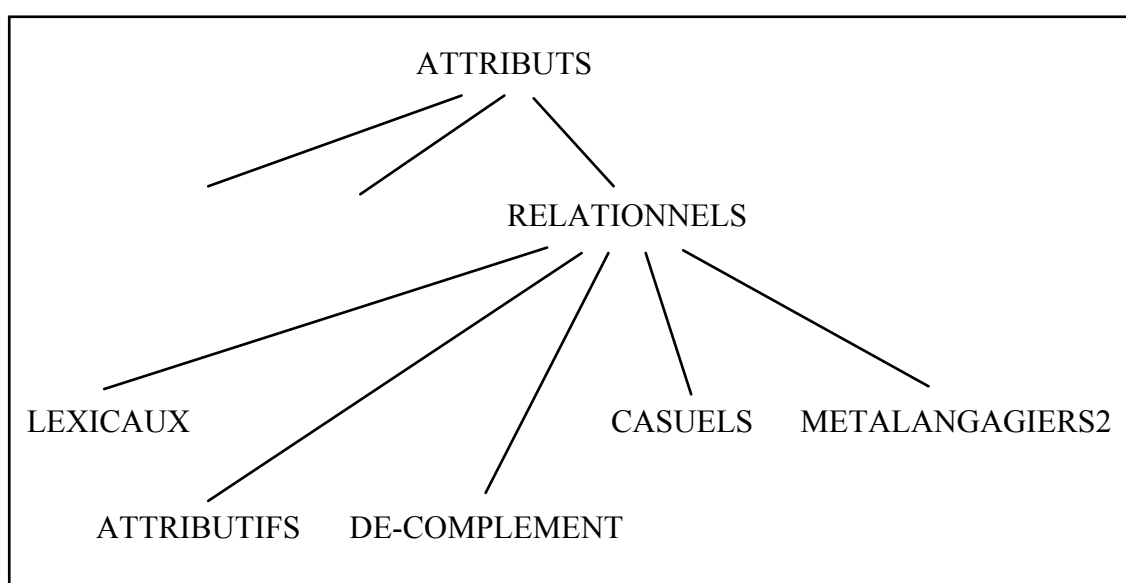
Horrexek zilegizatzen du, guk uste, proiektu honetan hautatu *errepresentazio-ereduaren erredundantzia*. Izan ere, HDBtik ekarritako eta definizioen analisiaren emaitza den informazioa HEBan jasotzen da, maila definitorio esan diogun horretan, errepresentazio-

---

<sup>1</sup> Izenezko balioak eduki ditzaketen atributuei erantsitako fazetak: *DETERMINATION, NOMBRE, GENRE*.  
Aditezko balioak eduki ditzaketen atributuei erantsitako fazetak: *REFLEXIVITE, MODE, ASPECT, TEMPS, PERSONNE*.

atributuen bitartez. Baina informazio hori gordetzen da, orobat, maila erlazionalean, THESAURUSEko ikuspegi erlazonala eratuz. Maila definitorioan beti gordeko da "jatorrizko informazioa", maila erlazionaleko atributuek nola edo hala deduzitutako informazioa eduki dezaketelarik. Ez da ahaztu behar hala ere, inplementazioari dagokion aldetik, grabatzen dena —nahiz eta erredundanteki izan— THESAURUSEko kontzeptu-unitateen erakusleak besterik ez direla.

Atributu erlazonalak RELATIONNELS klasearen azpiko bost azpiklaseotan sailkatu ditugu: LEXICAUX, CASUELS, ATTRIBUTIFS, DE-COMPLEMENT eta METALANGAGIERS2 (ikus III.7 irudia).



**III.7 irudia.-** Atributu erlazonalen sailkapena STRUCTURESen.

LEXICAUX azpiklasekoak erlazio lexikalen errepresentaziorako sortuak dira eta CONCEPTS-TYPE zein bere azpiklaseetakoren batean egon daitezke definiturik. Hona hemen atributu lexikalon zerrenda: ANCETRE, ANTONYMS, DEFINI-PAR, DEFINITION-DE, DESCENDANT, GRADUATION, HYPERCYCLE, HYPERONYME, HYPOCYCLE, HYPONYME, QUALITE eta SYNONYMS.

CASUELS azpiklasekoak berriz, ACTIONS/EVENEMENTS klaseko unitateetan daude definiturik eta egintza horien adierazle diren aditzen kasu-erlazioen balio tipikoak hartzen dituzte. Ohizko kasu-erlazioak ditugu: AGENT, BENEFACTIF, COMITATIF, DATIF, DIRECTION, INSTRUMENTAL, LOCATIF, MANIERE, OBJECTIF, RESULTAT, SOURCE, TEMPS eta THEME.

ATTRIBUTIFS eta DE-COMPLEMENT klaseetako atributuak ENTITES klaseko unitateetan daude definiturik eta izen-adierazledun kontzeptuon berezko ezaugarriak dituzte baliotzat (ikus "qualia" direlakoei buruzkoa § III.1.2.1 azpiatalean). ATTRIBUTIFS

### III. kapitulua

azpiklasean *CARACTERISTIQUE*, *ETAT* eta *PREDICATION* izeneko atributuak aurkituko ditugu; *DE-COMPLEMENT*en berriz, *INSTRUMENTAL*, *LOCATIF*, *OBJECTIF*, *ORIGINE*, *POSSESSEUR* eta *POSSESSION*.

*METALANGAGIERS*<sup>2</sup> azpiklaseak LPPLko metalengoaia lexikografikotik erauzitako erlazio bereziak errepresentatzeko sortu atributuak biltzen ditu. Klase honetakoak dira *ENTITES*en definituriko *ACTION-DE*, *ART-DE*, *CARACTERE-DE-CQE*, *CARENCE*, *CE-QUI*, *CE-QUON*, *DE-CEQUI*, *ENSEMBLE-DE*, *ETAT-DE-CQE*, *FACULTE-DE*, *MANIERE-DE*, *MANQUE-DE*, *MEMBRE-DE*, *PARTIE-DE*, *PIECE-DE*, *QUALITE-DE-CQE*, *QUI* eta *SORTE-DE*; *ACTIONS/EVENEMENTS*eko *CESSER-DE*, *COMMENCER-A*, *DE-NOUVEAU*, *DEVENIR*, *ETRE*, *ETRE-EN*, *FAIRE*, *METTRE-EN* eta *RENDRE*, eta *QUALITES*eko *CONFORME-A*, *CONTRAIRE-A*, *DONT*, *PEU*, *QUI*, *QUI-A-DE*, *QUI-NA-DE*, *QUI-NE*, *QUON-NE-PEUT*, *RELATIF-A*, *SANS* eta *TRES*.

Atributu erlazionalak kontzeptuarterko erlazio lexiko-semantikoen adierazpide izanik, atributuon fazeten bitartezko zehaztapenak sistemaren portaera deduktiboaren alderdi bat erakusten du. Aipatu dugu jadanik *RELATION-INVERSE* iratxoa. Iratxo hori, dagokion atributuaren balioa eguneratzera doan bakoitzean exekutaraziko da, alderantzizko erlazioa ezartzearen *INVERSES-CORRESPONDANTS* fazetatik jasoz atributuaren alderantzizko erlazioa zein den (horrela erlazioa bi zentzuetan automatikoki mantenduz).

Aipatu nahi genukeen atributu erlazionalon beste fazeta bat, nahiz eta geroago zabalago azalduko dugun (§ III.3.1), erlazio lexiko-semantikoen konposaketa bideratu asmoz sortutako *COMPOSITION-LEXICALE* izeneko dugu. Fazeta horren bitartez deskriba daitezke atributu batentzat, oso era sinplean, atributuak denotaturiko erlazioa emaitza gisa ematen duten konposaketa lexikaleko bideak.

Bukatzeko, beste aspektu interesgarri bat atributu erlazional hauen eta *THESAURUS*eko kontzeptuen arteko lotura litzateke: adibidez, *CARACTERISTIQUE* atributua ongi lot dakioke |propriété I 3| kontzeptuari, edo *OBJECTIF* |objectif I 4|<sup>1</sup> kontzeptuari, erlazio-kontzeptu lotura horiek kasuan kasuko atributuaren *CONCEPT-TYPE-CORRESPONDANT* izeneko fazetan azalduz. Honek emango liguke aukera atomiko edo primitibotzat hartutako atributuen esangura "hobeki zehazteko", horretarako atributuok hiztegiko kontzeptu bati edo batzuei itsatsiz. Zenbait kasutan erraza gertatzen bada ere, beste batzuetan kontua konplexuago bihurtzen da (ikus § III.3.2n). Alderdi hau ez dago oso garatua egindako prototipoan; alabaina, garapen hori interesgarria ez ezik, *HIZTSU*Aren zenbait funtzionalitatearen eskakizun diren dedukzio-lanak egin ahal izateko premiazkoa ere bilaka daitekeelakoan gaude.

---

<sup>1</sup> *propriété I 3*: caractère propre (LPPL, 445. orr.)  
*objectif I 4*: but (LPPL, 380. orr.).



Txostenaren D eranskinean du irakurleak orain arte sortu eta definitutako atributu guztion deskribapen osoa: deitura, domeinua, balio-eremua, balio lehenetsia, herentzi rola, deskribapenaren osagai diren fazetak, eta abar.

#### III.1.3.2.3 Iratxoak: VALEURS-ACTIVES.

STRUCTURES azpi-EBko VALEURS-ACTIVES klasearen azpian biltzen dira iratxo edo balio aktiboak. Iratxo horiek atributuei lotzen zaizkie, atributuok sortzen diren lekuan, eta atributu horien balioak atzitzean —irakurketan nahiz eguneratzerakoan— exekutarazten dira automatikoki.

KEEren eskakizun bat balio aktibo edo iratxo horiek sistemaren ezagutza-base bateko unitate izatea da. Horretara, STRUCTURESen definitutako iratxoak azpi-EB honetako VALEURS-ACTIVES klaseko elementu izateaz gain sistemaren ezagutza-base den ACTIVEVALUESEko kide izango dira.

HEBaren hasierako eraikuntzari dagokionean, atributu definitorio, sintagmatiko eta metalengoaiazkoen (METALANGAGIERS1 klasekoak) balioak beren kideko erlazioaletara goratzeko iratxoak definitu eta inplementatu dira —DEFINITOIRE-RELATIONNEL eta SYNTAGMATIQUE-RELATIONNEL<sup>1</sup> izenekoak—; halaber, inplementatu dira konfigurazioetako atributu erlazioaletako balioak konfigurazio horiek errepresentatutako kontzeptuei pasarazten dizkiena —DEFINIENS-DEFINIENDUM—, erlazioak bi zentzuetan mantentzen dituen iratxoak —RELATION-INVERSE—, eta THESAURUSen definizio klasikoei esker eratzen den kontzeptuarteko egitura hierarkikoa (hiperonimi eta hiponimi erlazioak) esplizitu bihurtzen duena —CONSTRUCTION-TAXONOMIQUE—.

#### III.1.3.2.4 Erregelen sailkapena: REGLES.

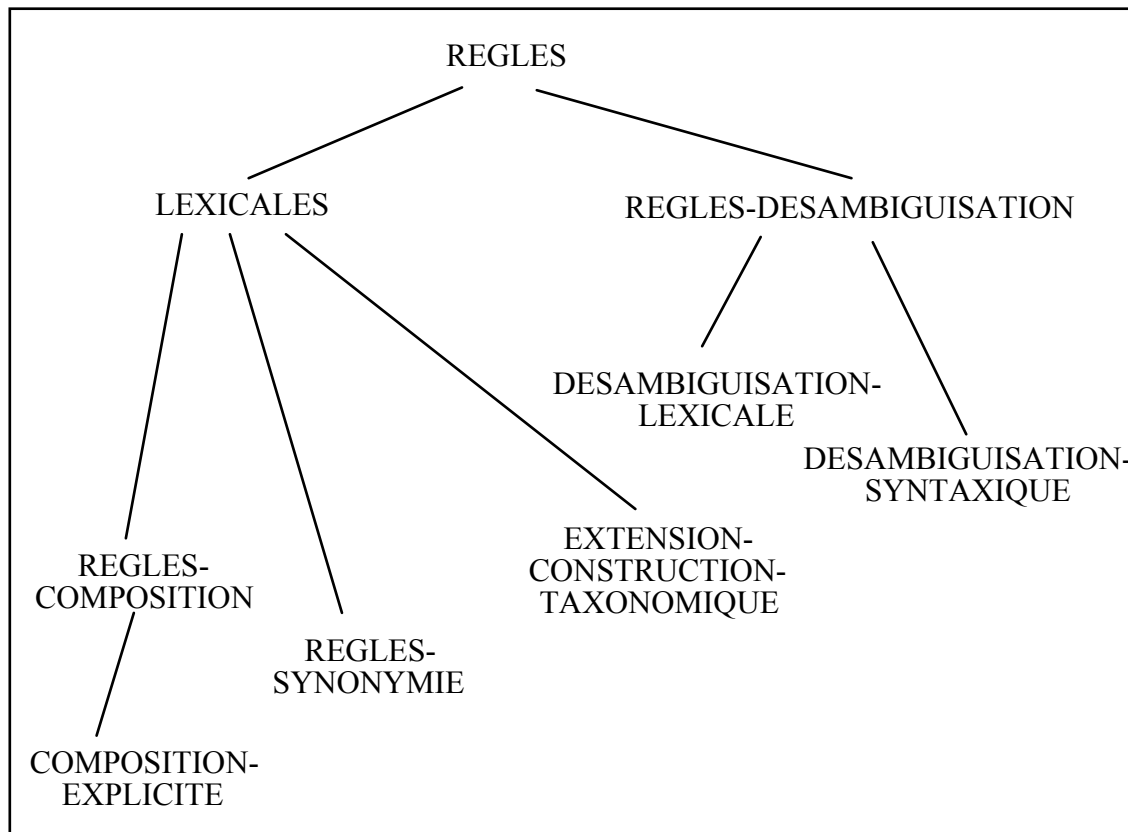
REGLES klasearen barnean sailkatu ditugu prozesu deduktibo desberdinetan erabili beharreko erregelak. Erregelak ere objektu gisa tratatzen ditugu eta zenbait azpiklasetan —REGLES klase orokorraren azpian— sailkatzen, horrela ezagutza deduktibo hori hobeki egituratuz (ikus III.8 irudia). Azpiklaseon artean bada LEXICALES izeneko erregela-klasea, bertan erlazio desberdinen propietateak eta beren arteko konbinazioak ustiatzen dituzten erregelak daudelarik, eta desanbiguaketa-erregelak biltzen dituen REGLES-DESAMBIGUISATION klasea (DESAMBIGUISATION-LEXICALE eta DESAMBIGUISATION-SYNTAXIQUE azpiklaseetan banatua). LEXICALES klaseak,

---

<sup>1</sup> SYNTAGMATIQUE-RELATIONNEL iratxo berbera lotzen zaie, balioen goratzeari dagokionerako, atributu sintagmatikoei zein METALANGAGIERS1 klasekoei.

### III. kapitulua

bere aldetik, ondoko azpiklaseak ditu: EXTENSION-CONSTRUCTION-TAXONOMIQUE eta REGLES-SYNONYMIE —HEBaren aberasketa-fasean erabiltzen direnak—, eta REGLES-COMPOSITION; azken klase honetako erregelen erabilera dinamikoa izango da, gehienbat erabiltzailearen eskaripean usatuko direlarik. Erregela-klase guzti hauez hitz egingo dugu zabalago alderdi deduktiboaz ihardutean.



**III.8 irudia.-** Erregelen sailkapena STRUCTURESen.

#### III.1.3.3 DICTIONNAIRE azpi-EBa.

DICTIONNAIRE azpi-EBaren zeregina THESAURUSeko kontzeptuak hitzaren bitartez atzitu ahal izatea da: hiztegi-sarrera eta bere adieren arteko erlazioa islatzen du.

Thesaurus edo sareko korapiluneek hiztegiko hitzen adiera edo zentzu desberdinak errepresentatzen dituzte. Horrela bada, THESAURUSeko |plante I 1| eta |plante I 2| unitateak bi kontzeptu desberdinen adierazpide ditugu. Baina, bada |plante| unitate bat DICTIONNAIRE azpi-EBan, STRUCTURESeko ENTREES klaseko instantzia, eta unitate horren SENS atributuaren bitartez *plante* hitza bere bi adierekin —*plante I 1* eta *plante I 2*— erlazonaturik dago:

|plante|

SENS: |plante I 1|, |plante I 2|

### III.1.3.4 Sarea: THESAURUS.

HIZTSUAren kontzeptu-sarea HEBko THESAURUS azpi-EBan dago errepresentaturik. Errepresentazio honek honako ezaugarriok ditu:

- Kontzeptu-thesaurusa errerepresentatzeko *frame-sare* bat erabiltzen da.
- *Nodo* edo korapilune bakoitzean kontzeptu-unitate baten adierazpide den frame bat dago. Funtsean bi *kontzeptu-unitate* mota daude: mota-kontzeptuak eta konfigurazioak. Mota-kontzeptuen artean era desberdinetako unitateak daude: berezko definizioa duten kontzeptuak, HEBaren hasierako eraikuntzan<sup>1</sup> erreferentzia "hutsak" bezala sortutako kontzeptuak, desanbiguatu gabeko erreferentzia gisa sortutako kontzeptu anbiguoak, eta abar.
- *Arkuak* kontzeptu-unitateen arteko erlazio lexiko-semantikoen errepresentazio dira eta erreferentzi baliodun atributuen bitartez inplementatu dira.
- Kontzeptuen arteko hiperonimi eta hiponimi erlazioak esplizitu egin dira, *kontzeptu-taxonomia* bat eratuz. Hierarkia horren ezarpenerako inplementazio-inguruneak eskainitako erlazio hierarkikoaz baliatu gara; horrela bada, HEBaren inplementazio honetan KEEren SUBCLASS.OF lokailuak '*hyponyme de*' erlazio lexikala errepresentatzen du. Sistemaren erlazio hierarkiko horren portaeraz balia gaitzke honela kontzeptuen beren hiponimoetarako herentzia ustiatzerakoan.

W. Lenders-ek (1990), bere lexikoaren erduan, lexiko-sarrera erlazio-multzo bezala definitzen du, sarrera hori lexiko bereko beste sarrera batzuekin eta banako baten ezagutzaren beste atal batzuekin lotzen duten erlazioen multzo bezala alegia. Autore horren iritziz, hiztegia *sare erlazional dinamikotzat* hartu behar genuke; testuinguru honetan dinamikotasunak esan nahi du hiztegia ez dela sistema egonkor bat, aldaezina, baizik eta ezagutza berriak eskura ditzakeen sistema bat, bere edukia era iraunkor batez aldatzeraino. Hiztegia sare erlazional dinamiko bezala ikusteak duen beste alderdi bat da ez dela erredundantea —hitzak toki batean bakarrik daude definiturik, eta beste askotan erreferentziaturik—, eta ezagutza implizitua deduzitzeko gai diren prozedura "txertatuak" (built-in) eduki ditzakeela. Hona Lenders-ek (1990:49) dioena hiztegiaren sare izate horretaz:

---

<sup>1</sup> "HEBaren hasierako eraikuntza" diogunean, DICTIONNAIRE eta THESAURUS azpi-EBen eraikuntzaz ari gara noski; izan ere, STRUCTURESek alde zuzenetik eskuz eratua izan behar baitu.

### III. kapitulua

Understanding a dictionary as a "network" means that a lexical entry is represented by a lexical node which is specified by sets of lexical relations.

Gure errepresentazioan ere ageri da lexikoaren eredu horrekin zerikusi handirik duen hainbat ideia: oinarritzko sare-egitura, ezagutzaren dinamikotasuna —aberasketa, konposaketa lexikala—, hiztegiko "objektuei" erantsitako prozedurak, etab.

Thesauruserako proposatzen dugun errepresentazio-ereduaren lehen iturria M. Quillian-en (1968) "Semantic Memory" izeneko artikulua klasikoa dela esan dezakegu.

"Semantic Memory" horretan, Quillian-en helburua giza burmuineko memoria semantikoa errepresentatzea da. Giza memoria semantikoaren bere irudiak sare baten itxura du: "type node" eta "token node" direlakoan multzoa, sareko nodoen bitartez errepresentatuak eta erlazio desberdinen bidez elkarlotuak. Errepresentagai duen memoria semantiko horri eduki bat eman nahian hiztegi elebakar bat —definizioarena— erabiltzen du, hiztegiko sarrerak sareko mota-kontzeptutzat hartuz ("type node") eta "token"ak mota-kontzeptu horien besteren definizioetako agerpenak errepresentatzeko erabiliz.

Quillian-ek definizio baten osagaien artean bi erlazio mota ezartzen ditu: plano barneko loturak eta planoartekoak —"intraplane links" eta "interplane links"—. Plano barneko loturak definizioaren osagaiak dauden mailan ezarri erlazioak dira, definiendunaren —"type node"— eta definizio-testuko elementuen artekoak; halaber, horien artean sartzen dira definizioetako osagaien arteko erlazioak ere. Planoarteko erlazioak aitzitik, "type node" eta definizioetako hitz-agerpenen adierazpide diren "token" desberdinen arteko erlazioak dira, "token"ok "type"en instantziak errepresentatzen dituztelarik.

THESAURUS mota-kontzeptu eta konfigurazioen multzo bat da (CONCEPTS-TYPE eta CONFIGURATIONS dira STRUCTURESen bi kontzeptu-unitate mota horiek biltzen dituzten klaseak), Quillian-en "type node" eta "token"en kideko edo lirakeenak.

REFERENCES klaseko unitateen sorkuntza dela medio —unitate horiek beste kontzeptu batzuen definizioan agertzeagatik sortzen direla gogoratu— THESAURUS azpi-EBa *multzo itxi* bezala ikus dezakegu, non kontzeptu-erreferentzia guzti-guztiak THESAURUSen esistitzen diren unitateen erreferentzia baitira, nahiz eta batzuek ez izan, hasiera batean behintzat, berezko ezagutza semantikorik.

CONFIGURATIONS klaseko elementu diren THESAURUSeko unitateak, *"etiketarik gabeko kontzeptuak"* (Litkowsky, 78)ren zentzuan, kontzeptu bat beste baten definizioan agertutakoan eratutako egitura sintagmatikoen errepresentazio dira. Horrela errepresentatutako sintagmak beren hun-kontzeptuarekin zuzenean lotzen dira; bertako unitateek osatzen dute sintagmaren adierazpide semantikoa, beti kontzeptu bat baino gehiagori dagokiona. Konfigurazioek beren kideko hurbila dute Quillian-en "token"etan.

Bien arteko alderik nabarmenena da, guk uste, gurean konfigurazioa dagokion mota-kontzeptuarekin lotzeko klase/azpiklase lokailua erabiltzen dugula, klase/instantzia lokailu baten orde. Beraz, Quillian-en "token"ak gurean ez dira mota-kontzeptuen instantzia ezpada azpiklase, horrek dituen abantaila guztiakin, propietateen herentziaz balia baikaitezke mota-kontzeptuen nahiz konfigurazioen kasuan. Errepresentazio honi esker badaukagu, noski, mota-kontzeptu bakoitza definizioetako bere agerpen guztiakin zuzenean loturik.

Ikus dezagun orain, adibide pare baten bitartez, zer-nolako itxura duen THESAURUSEko errepresentazioak. Aurreneko adibidea definizio aristoteliko bati dagokio, LPPLtik aterea:

***géranium I 1: une plante d'ornement***

Definizio honek sortarazten ditu, HEBaren hasierako eraikuntza egiten denean, gutxienez bi kontzeptu-unitate berri THESAURUSEn: definiendumari dagokiona batetik, eta definienseko izen-sintagmaren adierazpide izango den konfigurazioa bestetik. Konfigurazio hori sortu ahal izateko, beharrezkoa da aldeztu aurretik bere hun-osagaia den hitza —*plante*— errepresentatuko duen unitaterik egotea THESAURUSEn; eta ez badago, oraintxe sortuko da. Bestalde, beste unitate bat ere sortu beharko da lehenagotik ez bada existitzen: *ornement* errepresentatzen duena, hain zuzen. Eman dezagun bai *plante* eta bai *ornement* behar den bezala desanbiguaturik iristen zaizkigula HDBtik —*plante I 1* eta *ornement I 1* adierei dagozkiela—; alabaina, suposatuko dugu |*plante I 1*| kontzeptua THESAURUSEn sortua dagoen arren *ornement* hitzari dagokionik ez dela bertan existitzen. Beraz, hiru unitate berri sortuko lirateke: |*géranium I 1*|, |*plante I 1*#3|<sup>1</sup> et |*ornement I 1*|.

Ikus dezagun unitateon edukia zein den —adibidearen aurkezpena arintzearen, pittin bat soildua— errepresentazio maila definitorioan (atributu definitorioak, sintagmatikoak eta informazio orokorrekoak):

**|*géranium I 1*|**

MEMBER.OF: NOMS

GROUPE-CATEGORIEL: NOM

CLASSE-ATTRIBUT: INFO-GENERALE

TEXTE-DEFINITION: "une plante d'ornement"

CLASSE-ATTRIBUT: INFO-GENERALE

---

<sup>1</sup> |*plante I 1*#3| konfigurazioak gunetzat *plante I 1* kontzeptua duen sintagma bat errepresentatzen du. Adibidean erabilitako 3 zenbakiak berriz, konfigurazio hori |*plante I 1*| kontzeptuaren peko 3. konfigurazioa dela adieraziko luke. HEBaren hasierako eraikuntza burutzen duen programak esleitzen ditu (1etik hasi eta jarraian) konfigurazio-zenbaki horiek, konfigurazioak sortu ahala.

### III. kapitulua

DEF-CLASSIQUE: |plante I 1#3|  
CLASSE-ATTRIBUT: DEFINITOIRES  
DETERMINATION: UN  
GENRE: F  
RELATIONNELS-CORRESPONDANTS: DEFINI-PAR

#### |plante I 1#3|

SUBCLASS.OF: |plante I 1|  
MEMBER.OF: NOMINALES  
TEXTE: "plante d'ornement"  
CLASSE-ATTRIBUT: INFO-GENERALE  
DE: |ornement I 1|  
CLASSE-ATTRIBUT: SYNTAGMATIQUES  
RELATIONNELS-CORRESPONDANTS: ORIGINE, POSSESSEUR, MATIERE, OBJECTIF  
OBJECTIF: 0.9

#### |ornement I 1|

MEMBER.OF: REFERENCES

MEMBER.OF atributuaren bitartez adierazten dugu unitateak zein klaseren barnekoak diren: |géranium I 1| NOMS klaseko (DEFINITIONSen azpiklase) elementu edo instantzia da, HDBtik bere definizioarekin etorritako kontzeptua dela adieraziz, |plante I 1#3| NOMINALESekoa da (CONFIGURATIONSen azpiklase) baina baita |plante I 1| kontzeptuaren azpiklase edo hiponimo ere, bere hun-kontzeptua zein den adieraziz, eta azkenik, |ornement I 1| REFERENCES klaseko elementua da.

Klase-barnekotza horiei esker, unitateok klase horietan definitutako atributuak —berauen fazetekin— heredatuko dituzte. Adibidez, |géranium I 1| eta |plante I 1#3| unitateek informazio orokorreko atributuak heredatu dituzte, beren errepresentazioko zenbait aspektu jasotzeko egokiak direnak. Izen definituaren adierazpide den framean, GROUPE-CATEGORIEL eta TEXTE-DEFINITION atributuak ageri dira: atributuok, DEFINITIONSen sortuak izanik, beronen azpiklasea den NOMS dela bide heredatu ditu |géranium I 1| unitate horrek. |plante I 1#3| konfigurazio-unitateak heredatutako TEXTE atributua berriz, CONFIGURATIONS klasean definiturik dago, konfigurazio bakoitzak errepresentatzen duen sintagmaren testua gordetzeko pentsatua. Ikus daitekeenez, CLASSE-ATTRIBUT fazeta atributu guztietan ageri da, atributua zein klasetakoa den adieraziz.

|géranium I 1| unitatean gainera, bada atributu definitorio bat balio bezala THESAURUSEko beste unitate baten erreferentzia esplizitua daukana. Atributu hori DEF-CLASSIQUE da —hitza era klasikoan definitua izan dela, "genus et differentia" eran alegia, adieraziz— eta balioa definizioaren errepresentazioa den |plante I 1#3| unitatearen erreferentzia. DEF-CLASSIQUE atributua DEFINITIONS klasetik heredatua da eta bi fazeta aipagarri ditu kasu honetan: batetik *DETERMINATION*, jatorrizko definizioako artikulua

presentzia gordetzeko erabilia, eta bestetik *RELATIONNELS-CORRESPONDANTS*, atributu definitorioari dagokion erlazionala zein den zehazten duena (atributu definitorioetatik erlazionalerako balio-goratzeari du helburu, berehala ikusiko dugunez).

Era berean, |plante I 1#3| unitatean atributu sintagmatiko bat ageri da, DE, |ornement I 1| unitatearen erreferentzia daukana. Atributu sintagmatiko hori NOMINALESetik heredatua da eta bi kontzeptu-unitateon arteko erlazio sintaktikoa errepresentatzen du (testuan *de* preposizioak denotaturiko erlazioa). Kasu honetan, *RELATIONNELS-CORRESPONDANTS* fazetak erlazio hori maila sintaktiko soiletik sakonago batera pasa dadin, semantikora edo, dauden aukera desberdinak zehazten ditu: erlazio sintaktiko hori "izan" baitaiteke, sakonean, jatorria, jabea, gaia, helburua, eta abar. Halaber, atributu berean dagoen beste fazeta bat *OBJECTIF* delako erlazionala da, 0.9ko ziurtasun-balioa daukalarik. Fazeta horrek adierazten du DE erlazio sintagmatikoa *OBJECTIF* bihurtzeko bide dagoen ziurtasuna, erlazioa maila semantikora goratzeko dagoen ziurtasun aski handia alegia<sup>1</sup>.

Ikus dezagun orain THESAURUSEk duen "itxura erlazionala", unitateok sortu eta gero. III.9 irudian aurkezten duguna da |géranium I 1| kontzeptuaren inguruko THESAURUSEn parte, non irudi bidez "sare ikuspegia" (nodoak eta arkuak) azpimarratu nahi genukeen. Bista hori hasierako eraikuntzaren ondoren hartua dela esan dezakegu, zenbait prozedura deduktibok bere lana egin eta gero beraz. Irudirako egintzat emango ditugu, bereziki, atributu definitorioetatik erlazionalerako balio-goratzeari, sintagmatikoetatik erlazionalerakoak (esan bezala, adibide honetan suposatu dugu gai izan garela *de* preposizio hori *OBJECTIF* bezala desanbiguatzeko), alderantzizko loturen sorkuntza, konfigurazioetatik beren baliokide diren kontzeptuetarako balio-intsuldatzeari eta eraikuntza taxonomikoa (hiperonimi eta hiponimi erlazioen araberakoa).

III.9 irudian bada, ikus daiteke |géranium I 1| kontzeptuaren inguruko bista erlazional hori. Azpimarratu nola egitura taxonomikoa esplizitu eginga den bertan (1) —eta (1')— lokailuen bitartez, sistemaren SUBCLASS.OF erlazioaren bitartez inplementatutako hiperonimia eta hiponimiaren adierazpide delarik. Irudian ikus daitekeen beste erlazio taxonomiko bat |plante I 1| eta bere hiperonimoa den |végétal I 2|ren artekoa da, *plante I 1* en definiziotik eratorria agi danean. Bi muturreko geziak bi zentzuetan ezarritako erlazioak adierazten dituzte.

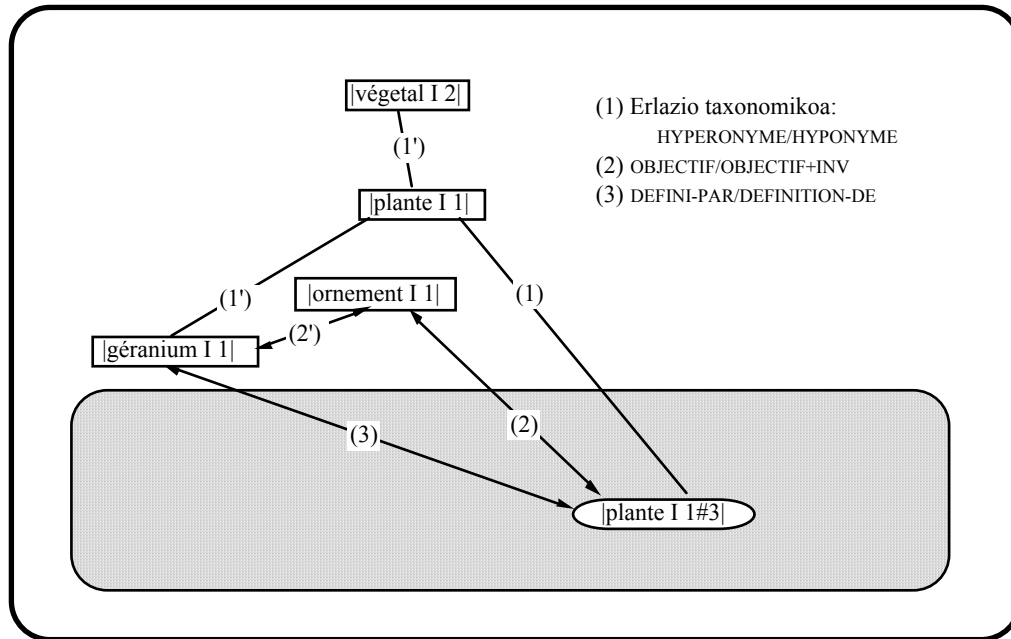
Irudian ez dira marraztu THESAURUSEko unitateen eta STRUCTURESEko azpiklaseen artean dauden erlazioak, irudia bera gehiegi ez lardaskatzearen: gorago aipatu erlazioak

---

<sup>1</sup> Kontua ez da orain fazeta horren baliora nola iritsi garen azaltzea, errepresentazio-formalismoa deskribatzea baizik. Bistakoa da, kasu honetan, jatorrizko definizio-formulak —*nom "de" nom*— ez duela, analisi sintaktikoaren mailan behinik behin, *de* preposizioak denotatzen duen erlazio sintaktikoa hobeto zehazteko aukerarik ematen. Alabaina, esan behar da zenbait kasutan posible izan dela horrelako zehaztasunak definizio-analisiko patroiz batzuen ESEEtan egitea.

### III. kapitulua

alde batera utziz —NOMS, REFERENCES eta NOMINALES klaseekiko barnekotza-erlazioak—, irudian ageri den mota-kontzeptu bakoitza (|géranium I 1|, |plante I 1|, |végétal I 2| eta |ornement I 1|) STRUCTURESeko ENTITES klasearen azpiklase da (dena dela, era honetako loturak ikusteko, hor beheko III.10 irudira jo daiteke, bertan ageri dira eta).



#### III.9 irudia.- |géranium I 1| kontzeptuaren inguruko ikuspegi erlazonala THESAURUSen<sup>1</sup>.

Azpimarratzekoa da hemen bi unitateren arteko *baliokidetzakontzeptuala* edo esaten diogun erlazioa: irudian (3) zenbakia duen DEFINI-PAR/DEFINITION-DE erlazio-bikoteak ezarritakoa litzateke. Izan ere, lokailu horren bitartez estekaturiko unitateek, |géranium I 1| eta |plante I 1#3|k alegia, kontzeptu berbera errepresentatzen baitute, bata bestearen definizioaren adierazpide denez.

Jakina da erlazio guzti hauen inplementazioa frameetako atributu erlazionalek dituzten erreferentziatzko balioetan oinarritzen dela. Ikus dezagun, adibide gisa, zein litzatekeen |géranium I 1| kontzeptuaren framea, atributu erlazionalek —kasu honetan ENTITES klasetik heredatuak— bete eta gero (errepara izartxoei):

<sup>1</sup> Hiperonimia/hiponimiako erlazio taxonomikoak marra lodixegoz marratu dira irudian, gezi-muturrik gabe. (1') eta (2') zenbakia duten erlazioak deduzituak dira. Kapitulu honetan aurrerago dauden irudi guztietan ' (lehena) ikurra duten loturek deduzitutako erlazioak adierazten dituzte.



**|géranium I 1|**

SUBCLASS.OF: ENTITES, |plante I 1| (\*)

MEMBER.OF: NOMS

GROUPE-CATEGORIEL: NOM

CLASSE-ATTRIBUT: INFO-GENERALE

TEXTE-DEFINITION: "une plante d'ornement"

CLASSE-ATTRIBUT: INFO-GENERALE

DEF-CLASSIQUE: |plante I 1#3|

CLASSE-ATTRIBUT: DEFINITOIRES

DETERMINATION: UN

GENRE: F

RELATIONNELS-CORRESPONDANTS: DEFINI-PAR

DEFINI-PAR: |plante I 1#3| (\*)

CLASSE-ATTRIBUT: RELATIONNELS

INVERSES-CORRESPONDANTS: DEFINITION-DE

OBJECTIF: |ornement I 1| (\*)

CLASSE-ATTRIBUT: RELATIONNELS

INVERSES-CORRESPONDANTS: OBJECTIF+INV

Bukatzeko, III.9 irudiko bista erlazionalera itzuliz, azpimarratu nahi genuke nola THESAURUSeko kontzeptu-unitateak bi "geruzatan" bezala dauden kokaturik (Quillian-en bi planoak gogora ekarriz): mota-kontzeptuak gainekoan eta konfigurazioak azpikoan. Konfigurazio oro izango da honela, errepresentatzen duen egitura sintagmatikoaren hunkontzeptuaren hiponimo gisa adierazten denez, mota-kontzepturen baten azpiklase.

Eta gatozen orain bigarren adibidea ikustera. Kasu honetan metalengoaiazkotzat hartzen ditugun bi definizio ditugu aztergai, hau da, "formula bereziz" egindako definizioen multzokoak. Aditz-definizio anitz, LPPLn, *rendre* aditz-infinitiboan oinarritutako formula erabiliz egina dago; halaber, hainbat izen definitzen da *qui* hitzarekin hasitako formula batez<sup>1</sup>. Aukeratutako definizioak **publier I 1** eta **ajusteur I 1** sarrerei dagozkie. Bi definizio hauek errepresentatzeko, maila definitorioan bereziki sortutako DEF-RENDRE eta DEF-QUI metalengoaiazko atributuak erabiltzen dira:

**publier I 1:** *rendre public*

**ajusteur I 1:** *qui ajuste des pièces de métal*

Lehenengoaren adierazpide ondoko frame hau dugu:

---

<sup>1</sup> Izenen kasuan 'ce qui' formulaz egindako definizioak maizago suertatzen badira ere.

**|publier I 1|**

MEMBER.OF: VERBES

GROUPE-CATEGORIEL: VERBE

*CLASSE-ATTRIBUT: INFO-GENERALE*

TEXTE-DEFINITION: "rendre public"

*CLASSE-ATTRIBUT: INFO-GENERALE*

DEF-RENDRE: |public I 1|

*CLASSE-ATTRIBUT: DEFINITOIRES*

*RELATIONNELS-CORRESPONDANTS: RENDRE*

non ikusten den kasu honetan ez dela inongo konfiguraziorik sortzen, DEF-RENDRE atributu definitorioaren balioak zuzenean |public I 1|<sup>1</sup> kontzeptura garamatzalarik. Bigarren definizioak aldiz, *ajusteur I 1* kontzeptuari dagokionak, bi konfigurazio sortaraziko ditu (aparte utzita REFERENCES klasean sort ditzakeen unitateak). |ajusteur I 1| kontzeptu-unitateak, bere DEF-QUI atributu definitorioan |ajuster I 1#1| konfigurazio-unitatearen erreferentzia du baliotzat, konfigurazio horrek '*ajuster des pièces de métal*' sintagma errepresentatzen duelarik. Aditz-konfigurazio horrek, bere aldetik, OBJET atributu sintagmatikoan izen-konfigurazio baten erreferentzia dauka, '*pièce de métal*' sintagmaren errepresentazio denarena alegia.

**|ajusteur I 1|**

MEMBER.OF: NOMS

GROUPE-CATEGORIEL: NOM

*CLASSE-ATTRIBUT: INFO-GENERALE*

TEXTE-DEFINITION: "qui ajuste des pièces de métal"

*CLASSE-ATTRIBUT: INFO-GENERALE*

DEF-QUI: |ajuster I 1#1|

*CLASSE-ATTRIBUT: DEFINITOIRES*

*MODE: IND*

*ASPECT: NT*

*TEMPS: PRES*

*PERSONNE: 3*

*RELATIONNELS-CORRESPONDANTS: QUI*

**|ajuster I 1#1|**

SUBCLASS.OF: |ajuster I 1|

MEMBER.OF: VERBALES

TEXTE: "ajuster des pièces de métal"

*CLASSE-ATTRIBUT: INFO-GENERALE*

---

<sup>1</sup> Baliteke ordea, metalengoaiazko definizio hauetan ere atributu definitorioaren balioa betetzen duena konfigurazio bat izatea: esaterako, *adoucir I 1* kontzeptuaren definizioa —*rendre plus doux*— errepresentatzerakoan, '*plus doux*'en errepresentazio den konfigurazioa sortu beharko da, eta *adoucir I 1* kontzeptuari dagokion unitateko DEF-RENDRE atributuaren baliotzat bera ipini.

OBJET: |pièce I 1#2|  
CLASSE-ATTRIBUT: SYNTAGMATIQUES  
DETERMINATION: UN  
NOMBRE: PL  
RELATIONNELS-CORRESPONDANTS: THEME

**|pièce I 1#2|**

SUBCLASS.OF: |pièce I 1|  
MEMBER.OF: NOMINALES  
TEXTE: "pièce de métal"  
CLASSE-ATTRIBUT: INFO-GENERALE  
DE: |métal I 1|  
CLASSE-ATTRIBUT: SYNTAGMATIQUES  
RELATIONNELS-CORRESPONDANTS: ORIGINE, POSSESSEUR, MATIERE, OBJECTIF  
MATIERE: 0.9

Adibide honetan garbi ikusten da nola, sarri, konfigurazioak "etiketarik gabeko" kontzeptuen adierazpide diren, hizkuntzan ez baita —frantsesez bederen— '*ajuster des pièces de métal*' esangura duen aditzik, edota '*pièce de métal*' adierazteko izenik. Hori ez da gertatzen ordea, mota-kontzeptuekin, DEFINI-PAR/DEFINITION-DE erlazioen bitartez lotutako konfigurazioen kasuan, izan ere kasu horietan konfigurazioa definitutako kontzeptuaren errepresentazio baliokidea baita (ikus hemen goian *géranium I I* en definizioaren adibidea). Guk proposatutako errepresentazioan, konfigurazioek maiz itzulinguruz adierazitako kontzeptuak denotatzen dituzte, hiztegian askotan sarrerarik ez duten kontzeptuak<sup>1</sup> (ez derrigorrez behintzat).

Konfigurazio-unitateok horrela errepresentatzeak duen interes bat, kontzeptua eta kontzeptu horren besteren definizioetako agerpen guztien artean zuzeneko loturak mantentzearena da. Hori dela medio, kontzeptuaren erabilera-adibideen alegiazko multzo bat daukagula esan dezakegu, parafraseaketa-funtzioetarako eta, esate baterako, baliagarri gerta dakigukeena.

Kontzeptua eta bere definizioaren adierazpide den konfigurazioa errepresentazio aldetik banatzeak —definizio klasikoen kasuaz ari gara—, errepresentazio problema ezagun baten irtenbide bat idarokitzen du. Eman dezagun K kontzeptuaren definizioa '*grande fourmi*' dela; esan ote dezakegu K-ren ezaugarri bat '*grande*' izatea dela? Hortxe dago arazoa. Gure errepresentazioan, bi unitate desberdin egongo lirateke, K-ri dagokiona eta '*grande fourmi*' errepresentatuko lukeen konfigurazioa. Eta azkenak bakarrik edukiko luke, printzipioz,

---

<sup>1</sup> Hau interesgarria da, gure ustez, ingurune eleanitzean: posible baita, beste hizkuntzaren batean adibidez, frantsesez |pièce I 1#2| konfigurazioa sortuz errepresentatu behar izan den kontzeptuaren baliokideak hizkuntz adierazlerik izatea, hau da, bera denotatzen duen hitzik. Kasu horretan, konfigurazioetan oinarritu errepresentazio hau baliagarria gerta daiteke bi kontzeptuon arteko baliokidetza esplizitu egiteko.

### III. kapitulua

'grande' kontzeptuarekiko lotura; horrela adieraziko litzateke kalitate horren "erlatibotasuna" K kontzeptuarekiko: esan baitaiteke K "handia" dela "inurria" den neurrian, baina ez, inola ere, era absolutu batez<sup>1</sup>.

THESAURUSen esplizitu egin diren hiperonimia/hiponimiako erlazio taxonomikoak direla eta, bi gauza aipatu nahi genituzke: (1) bigiztei dagokiena, eta (2) herentziari buruzko guztia. Baina horiek, alderdi inferentzialari dagozkiolarik bereziki, hurrengo atalerako utziko ditugu.

Eta bukatzeko, ikus dezagun oraindik beste adibide bat. Adibide horren helburua ondokoa da: **pansement I I** kontzeptuaren definizioa dela medio —*action de panser une plaie*— sortutako era guztietako loturak irudi bidez azaltzea (ikus III.10 irudia); irudi horretan ikus daitezke, hiru azpi-EBez gain, (a) hiru azpi-EBotako unitateen arteko loturak, eta (b) definizio hori dela eta, HEBaren hasierako eraikuntzan DICTIONNAIRE eta THESAURUSen sortu edo erreferentziatutako<sup>2</sup> unitateen arteko loturak.

III.10 irudiko bista ere ez da osoa, definizio horri dagozkion lokailuak besterik ez baitira marratu (eta ez denak). Aipatu definizio horrek, formula finkatuzkoa kasu honetan ere ('*action de*'), |pansement I I| eta |panser I ?#1| unitateak sortarazten ditu, elkarren artean kontzeptu definituaren errepresentazio den unitateko DEF-ACTIOND atributu definitorioaren bitartez lotuz ((2) lokailua irudian). Halaber, DICTIONNAIREko |pansement| unitatea THESAURUSeko |pansement I I|ekin lotzen da, SENS atributuaren bitartez ((4) lokailua irudian). Definitutako kontzeptuak bere lekua hartzen du taxonomian |action I 3| kontzeptuaren azpian ((1) lokailua irudian), definizio-formula honen bidez definitutako kontzeptu guztiek hartzen baitute kontzeptu hori hiperonimotzat (halaxe zehaztu delako definizioarekin parekatutako patroari egokitu zaion ESEEn)<sup>3</sup>.

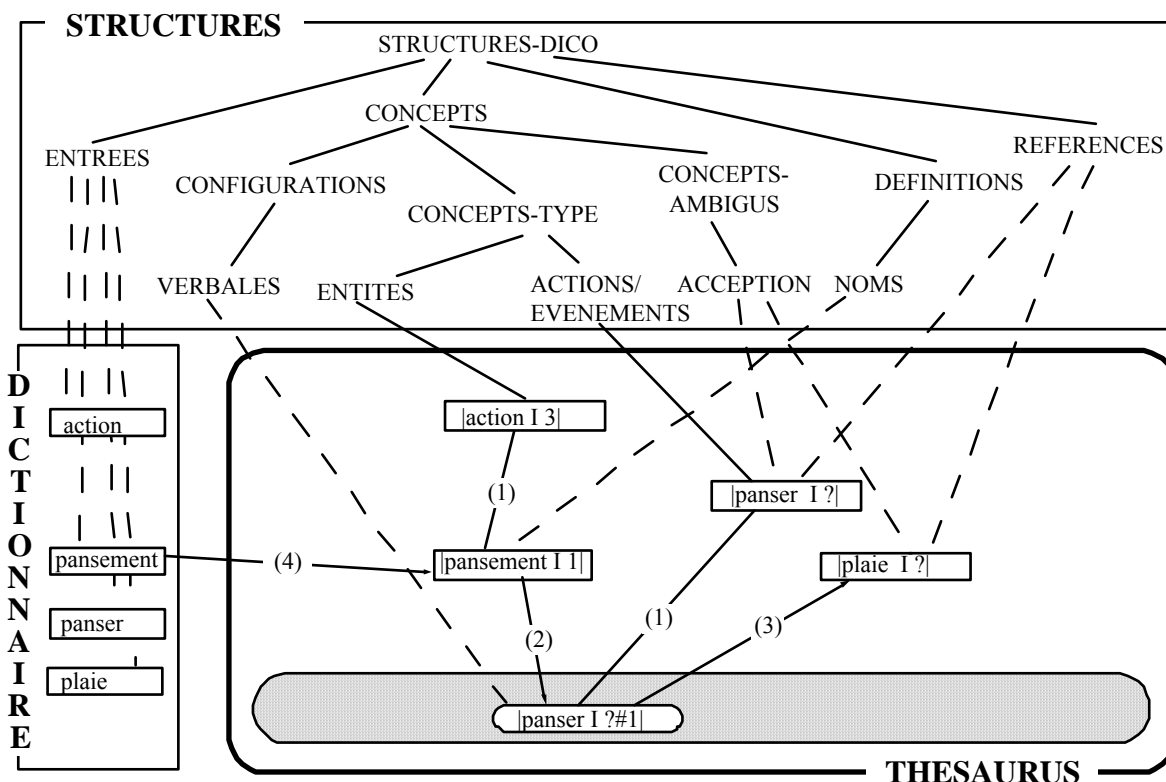
|panser I ?#1| konfigurazioa, '*panser une plaie*' sintagmaren adierazpide, bere hunkontzeptua den |panser I ?|en hiponimo bezala sortzen da, zein, bere aldetik, REFERENCESEko eta CONCEPTS-AMBIGUSen ACCEPTION azpiklaseko elementu den. Era berean, |plaie I ?| ere bi klase horien barneko elementua da, eta |panser I ?#1| konfigurazioarekin beronen OBJET atributuaren bidez dago loturik ((3) lokailua irudian)

---

<sup>1</sup> Dena dela, arazoa ebazteke geratzen da; orain kontua izango da nola eta noiz pasa konfigurazio-unitateetako atributuen balioak konfigurazio horiek errepresentatu unitateetara.

<sup>2</sup> Unitateen sorkuntza aldeztetik erreferentziatu gabekoen kasuan soilik gertatuko da. Ordea, unitateak dagoeneko THESAURUSen esistitzen baldin badira, ez dago atzera sortzen ibili beharrik, eta nahikoa izango da lotura egokiak ezartzea.

<sup>3</sup> THESAURUS azpi-EBari dagokionean, III.10 irudiko bistan maila definitorioko loturak besterik ez dira marratu ((1) lokailuen kasuan ezik, horiek maila erlazionalekoak baitira), irudia gehiegi ez lardaskatzearren.



**III.10 irudia.-** HEBko azpi-EB desberdinetako unitateen arteko erlazioak (*pansement I 1* kontzeptuaren definizioari dagokionean)<sup>1</sup>.

Azkenik, definitutako kontzeptua NOMS klasearen barnekoa ere bada, eta |panser I ?#1| konfigurazioa VERBALESkoa. Gainera, DICTONNAIREko unitate guztiak STRUCTURESeko ENTREEES klaseko elementu dira<sup>2</sup>.

Txosten honen E eranskinera jo dezake irakurleak THESAURUS azpi-EBan eratutako taxonomia batzuen itxura ikusi nahi izanez gero.

<sup>1</sup> Irudiko lokailuek ondoko erlazioak adierazten dituzte:  
 (1) Erlazio taxonomikoa: HYPERONYME/HYPONYME  
 (2) DEF-ACTIOND (atributu definatorioa, hasierako eraikuntzan bere kideko erlazionalera (ACTION-DE) goratua)  
 (3) OBJET (atributu sintagmatikoa, kasu konkretu honetan THEME erlazionalera gora daitekeena)  
 (4) SENS  
 — CLASS/SUBCLASS lotura  
 ..... MEMBER.OF lotura

<sup>2</sup> STRUCTURESen aurredefinitutako "moten" azpiko egitura taxonomikoari dagokionez, irudian ENTITES eta |action I 3|ren artean eta ACTIONS/EVENEMENTS eta |panser I ?|en artean dauden loturak besterik ez dira marratu. |plaie I ?| unitatea ere, jakina, ENTITESekin loturik egongo da, |panserment I 1| eta |panser I ?#1| ENTITES eta ACTIONS/EVENEMENTSekin, hurrenez hurren, egongo diren legez (kasu honetan zeharka, beren hiperonimoen bitartez).

### III.1.3.5 Asmatutako errepresentazio-ereduari buruzko zenbait gogoeta. Beste errepresentazio batzuekiko alderaketa.

Azpiatal honetan zenbait gogoeta egin nahi genuke, proiektuan asmatu eta arestian azaldu dugun errepresentazio-ereduari buruz. Bestalde, bereziki aukeratutako adibideak erabiliz, gurean eta hurbil diren beste eredu batzuetan definizioak errepresentatzeko modu desberdinak alderatuko ditugu. Zehatz esateko, gure eredu "sare semantiko" klasikoekin erkatuko dugu, (Jensen & Binot, 88)tik ateratako adibide bat dela medio, Arangoren tesian (1983) sortutako ereduarekin, Sowa-ren (1984) grafo kontzeptualekin eta Chouraqui eta Godbert-ek (1989) proposatzen dutenarekin.

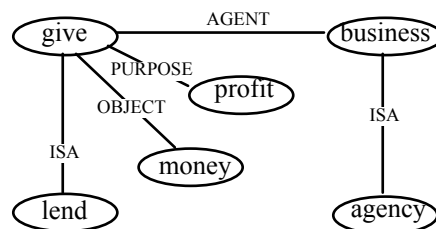
Hiztegi-ezagutza errepresentatzeko gure ereduak lexikoaren ikuspegi erlazionalean hartzen du oin, hiztegia elkarrekin erlazionaturiko kontzeptuen multzotzat hartuz. Kontzeptuarteko erlazioak errepresentazio-eredu honen adierazgai diren definizioetatik erauziak dira.

Hiztegi-ezagutzaren errerepresentaziorako gure ereduak honako ezaugarriok ditu:

- Adierazle edo hitzen eta kontzeptu- edo adiera-multzoaren arteko bereizkuntza garbia —DICTIONNAIRE azpi-EBa batetik, eta THESAURUS edo sarea bestetik—.
- Azpi-EB horietan errerepresentatutako objektu-klaseei buruzko metaezagutza daukan STRUCTURES azpi-EBa, lexikoko kontzeptu-unitateen errerepresentazioarekin zuzeneko zerikusirik ez duen zenbait aspektu aparte maneiatzeko aukera ematen duena.
- Definizioen bi errepresentazio maila bereizten dira argi eta garbi:
  - 1) Maila definitorioa, atributu definitorio, metalengoiarazko eta sintagmatikoetan oinarritua, non maila lexiko-sintaktikoko informazioa adierazten den: informazio "ziurra", jatorrizkoa (iturburu-hiztegitik ekarria). Errepresentazio maila hau lengoaia naturaletik hurbil geratzen da (erlazio parafraseagarriak).
  - 2) Maila erlazonala, atributu erlazonalen bitartez eratua, non ezagutza deduzitu, ez-ziur, aberasgarria dagoen, maila semantikokoa batik bat.
- Fazeta erlazonalen bitartez atributuak zehazteko aukera, atributuok sortzen diren lekuan (STRUCTURES azpi-EBan alegia): atributu sintagmatikoetatik erlazonaletarako balio-goratzekak egiteko dagoen ziurtasun maila zehaztu, erlazioaren portaera —beste batzuekiko konposaketa egitean— definitu, eta abar.

- Konfigurazio-unitateen sorkuntza: unitateok, errepresentazioari halako erredundantzia maila bat ematen badiote ere, hizkuntzan adierazlerik derrigor ez daukaten kontzeptuak denotatzeko erabili izen, aditz, adjetibo eta adberbiozko sintagmentzat adierazpide argi eta dotorea direlakoan gaude. Konfiguraziook kontzeptu-hierarkiaren parte izateak berriz, mota-kontzeptuen eta "etiketarik gabeko" kontzeptuon errepresentazioa bateratu egiten du.
- THESAURUSen kontzeptuen ikuspuntu desberdinak dauzkagu, STRUCTURESekoa klase bateko edo gehiagotako elementu izatearen ondorioz:
  - a) Mota-hierarkia lañoa, zeinen azpian mota-kontzeptu eta konfigurazio guztiak biltzen baitira, hiperonimia/hiponimia erlazio-bikotearen arabera.
  - b) Kontzeptu anbiguoak —definizioetan dagoen anbiguetate lexikala dela kausa sortuak— bereziki sortutako klaseetan daude bildurik eta errerepresentaturik.
  - c) Iturburuan definizioa duten kontzeptuak eta kontzeptu "hutsen" (erreferentziak, berezko definiziorik ez izan arren besteren definizioetan agertu izanagatik sortuak THESAURUSen) arteko bereizkuntza.
- Errepresentazioa bertan aplikatuko diren dedukzio-mekanismo desberdinen funtzioan diseinatua da: iratxoak, metodoak, erregelak, etab. Objektuei zuzendutako programazioaren ikuspegitik, arestian aipatu kontzeptuen ikuspuntu desberdin horiek izatea interesgarria da oso. Ikus, ondorengo ataletan, sistemaren alderdi inferentzialaz dioguna.

Eta gogoeta hauek eginda, ekin diezaiogun bada gure erredua beste batzuekin erkatzeari. Hasterako esan, errerepresentatutako definizioetan parte hartzen duten kontzeptuen alderdi erlazionala besterik ez dugula kontuan hartuko, eta beraz, THESAURUS azpi-EBan dagoenari buruz soilik arituko garela.



**III.11 irudia.-** *lend* eta *agency*ren arteko bide semantikoa LDOCE hiztegiaren (Jensen & Binot, 88:156).

Har dezagun lehenbizi (Jensen & Binot, 88:156)tiko adibide bat, sare semantiko klasikoaren errepresentazio moduari dagokiona. Adibidea erlatibozko perpaus bat bere

### III. kapitulua

aitzinkariari lotzeko dagoen anbiguetate bati buruzkoa dugu; definizio-informazioa errepresentatzeko ohizko sare semantikoa erabiltzen da. Hemen interesatzen zaiguna errepresentazioa bera da, autoreek diotenez, LDOCE hiztegiko definiziotik erabat automatikoki aterea dena. III.11 irudian dago autoreek "bide semantikoa" esaten diotenaren errepresentazio grafikoa, LDOCEn *lend* eta *agency* hitzak elkarlotzen dituen bidea alegia.

Errepresentazio partzial hau, LDOCEk, bi hitz horientzat (bakoitzaren adiera bat, hobe esan) ematen dituen definizioei dagokie:

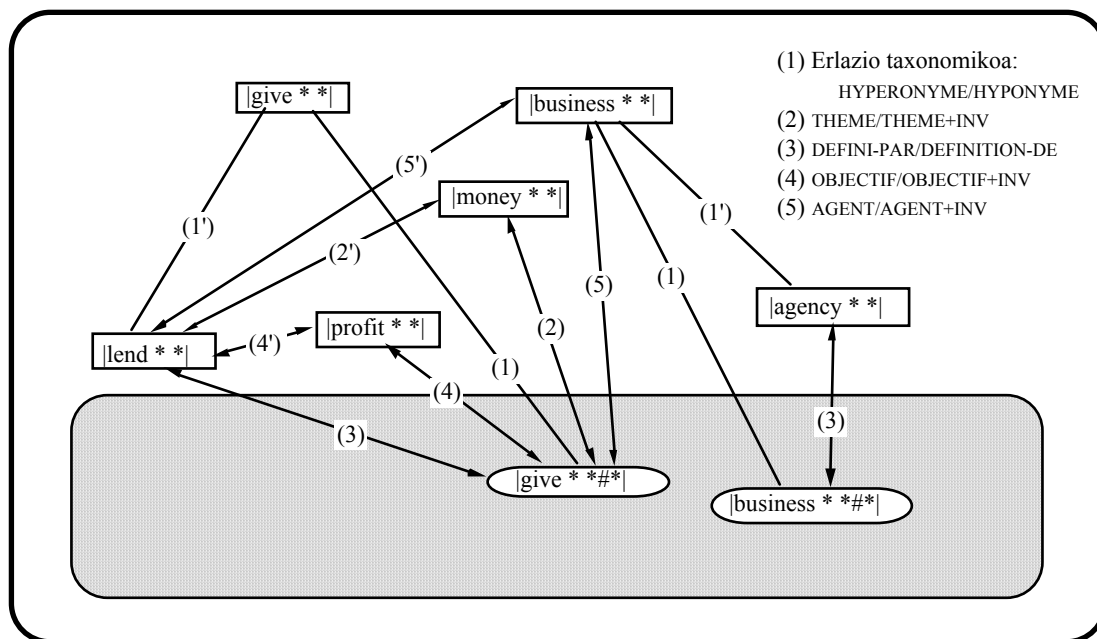
*lend*: to give out (money) for profit, esp. as a business.

*agency*: a business that makes its money esp. . . .

Galdera asko sortzen digu irudiko errepresentazio horrek. Nola jakin, esate baterako, irudiko *give* etiketadun nodoa mota-kontzeptuari baldin badagokio behintzat, zeintzuk diren kontzeptu horren definizio-ezaugarriak? Esan ote daiteke *give* horren helburua *profit* eskuratzea dela, iruditik ondoriozta bide daitekeenez, edota bere objektua *money* dela? Sare puska horretan adierazitakoa *give* kontzeptuaren instantzia jakin bat litzateke areago, *lend* kontzeptuaren definizioan agertu den *give* horrena hain zuzen, nahiz eta hori, irudian, ez esplizituki errerepresentatua egon: nola jakin *giveren* inguruko lotura horiek *lend* kontzeptuari dagozkiola eta ez direla, berez, *giverenak*? Nola errepresenta litezke *give* genustzat hartzen duten beste aditz batzuen definizioak? Argi dago *givek money, profit* eta *business* kontzeptuekin dituen loturak heredatuko direla, ISA lokailua dela bide, hiperonimotik hiponimora, hau da, lotura horiek *lend* kontzeptuak ere hartuko dituela, baina hori ez da aski, gure ustez, errepresentazio egoki batentzat. Dena dela, egileen alde esan behar dugu errepresentazio hori era dinamiko batez erabiltzen dutela, ezagutzaren errerepresentaziorako lengoaia egokitzen lengoaia naturala bera jotzen dutenez. Eta horregatik seguruenik, sareok ezagutza modu iraunkor batez errerepresentatzeko erabiltzen ez dituztenez, beren eredia baliagarri zaie, nahiz eta guretzat aipatu ditugun arazo horiek izan.

III.12 irudi horretan ikus daiteke aztergai ditugun definizio horiei gure eredian legokiekeen adierazpidea. Ohar bat egin nahi genuke errepresentazioaren azalpenari ekin aurretik: definizio horiek interpretatzeko moduz ez gara orain kezkatuko, interesatzen zaiguna errepresentazioa bera eta errepresentazio biak alderatzea baita; beraz, ez dugu aztertuko, adibidez, nola jakin daitekeen *business* kontzeptuak agente kasu-rola betetzen duela *lenden* definizioan.





**III.12 irudia.-** *lend* eta *agency* kontzeptuei legokiekeen errepresentazioa<sup>1</sup>.

Gorago aipatu arazoak ederki ebazten dira, guk uste, *lend* kontzeptuaren definizioa errepresentatzen duen `|give * *##|` konfigurazioarekin (ikus III.12 irudia). Gure errepresentazioan, aditz-konfigurazio hori —*giveren* adierazpide den mota-kontzeptuaren hiponimo eta '*give money for profit, esp. as a business*' aditz-sintagmaren errepresentazio— OBJECT, PURPOSE eta AGENT erlazioen baliokideen bidez lotuko litzateke *money*, *profit* eta *business* kontzeptuekin, hurrenez hurren. Konfigurazioak mota-kontzeptu horiekiko dituen erlazioak, jakina, automatikoki igaroko dira konfigurazioak errepresentatzen duen definizioari dagokion kontzeptura —*lend*—, izan ere kontzeptu hori eta konfigurazioa baliokide kontzeptualtzat jotzen baititugu. Bestalde, hori ez da *lend* kontzeptuak *give* kontzeptutik hartuko duen erlazio herentzia. *lend* kontzeptuak heredatuko lukeena, herentzia guk ulertzen dugun moduan behintzat, *giveren* erlazioen balio prototipikoen multzoa bailitzateke —*giveren* esanguraren maila intentsional hutsean ulerturik—: adibidez, *giver* baldin bada *give* egintzaren agente tipikoa, *giver* hori hereda lezakete agente gisa *giveren* hiponimoek, nahiz eta *lenden* kasu konkretuan, berezko agentea duenez, *business* geratu agente rol hori betetzen, *giver* berezitu bat den neurrian (emanik AGENT atributuaren herentzi rola "override" erakoa dela, hau da, balio lokalak daudelarik heredatuak aintzakotzat hartzen ez dituen).

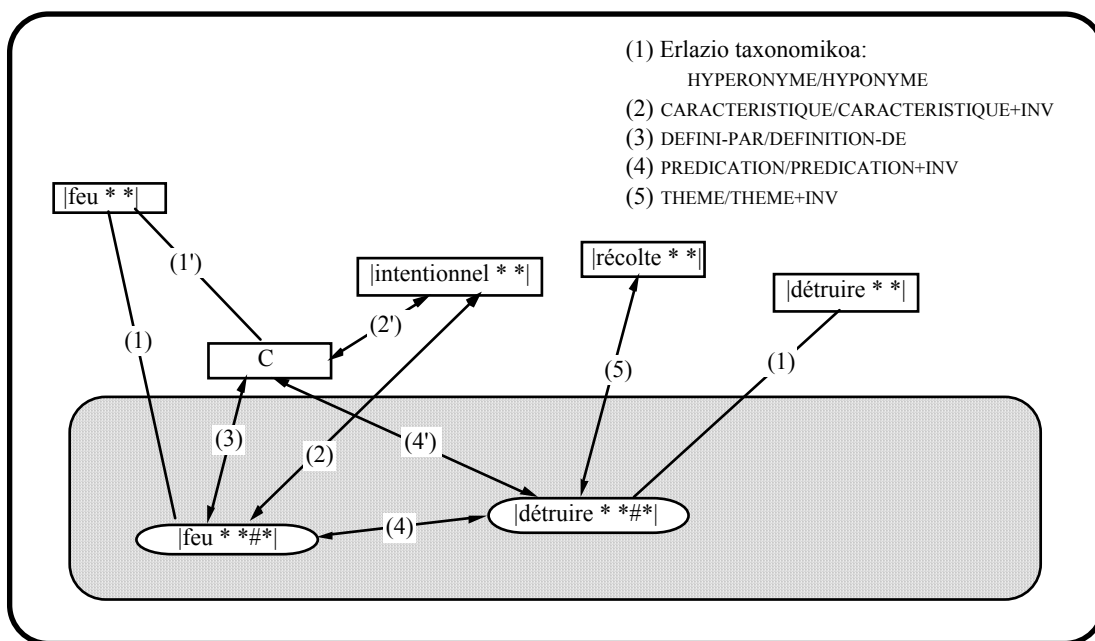
<sup>1</sup> Izartxoek adierazten dituzte homografo-, adiera- eta konfigurazio-zenbakiak (beren kokagunearen arabera). Irudiko konfigurazioek errepresentatu testuak:  
`|give * *##|`: *to give out (money) for profit, esp. as a business*  
`|business * *##|`: *a business that...* (agency kontzeptuaren definizio-testua)

### III. kapitulua

Eta goazen bigarren adibidea ikustera. Kasu honetan G. Arango-k (1983:73) bere tesian ematen duen definizio bat hartuko dugu errepresentazioen alderaketa egiteko gaitzat: eman dezagun, dio berak, '*feu intentionnel qui détruit les récoltes*' esaldia hiztegiko kontzeptu baten definizioa dela (kontzeptu horri C esango diogu). Definizio horren errepresentaziotzat honako hau ematen digu:

**C:** < CLE: feu  
 PROPRIETES: ( \* intentionnel  
 ACT: (CLE: détruire CAS: (OBJET: récolte))) >

Frameetan oinarritu errepresentazioa da, gurean hartutakoaren antzekoa. Erabilitako atributuak desberdinak izanik ere oso antzekoak ditugu: berak erabiltzen duen CLE horrek kontzeptuaren genusa adierazten du, eta PROPRIETES izeneko framea erabiltzen du kontzeptuari dagozkion ezaugarriak edo biltzeko. PROPRIETES horren barruan, ACT izeneko frame berezia erabiltzen du propietatearen balioa egintza bat denean; orobat, CAS framea erabiltzen du aditzak bere kasu-atributuak betetzen dituzten balioekin dituen erlazioak errepresentatzeko.



**III.13 irudia.-** '*feu intentionnel qui détruit les récoltes*' definizioari legokiokeen errepresentazioa<sup>1</sup>.

Gure kasuan berriz (ikus III.13 irudia), bi atributu desberdin —CARACTERISTIQUE eta PREDICATION— erabiliko genituzke definizioa errepresentatzeko, behin maila erlazionalean.

<sup>1</sup> Irudiko konfigurazioek errepresentatu testuak:  
 |feu \* \*#\*|: *feu intentionnel qui détruit les récoltes*  
 |détruire \* \*#\*|: *détruire la récolte*

Hala ere Arangorenarekiko alderik nabarmenena ondokoa litzateke, alegia guk bi konfigurazio-unitate sortuko genituzkeela sarean, fisikoki, C-ren framean bi erreferentzia egongo liratekeelarik (berriro gogoratu erakusle soilak direla, informazioa berez ez baita C-ren framean egongo): bata C-ren baliokidea den konfigurazioa, eta bestea, eta horixe da interesatzen zaiguna orain, *'détruire la récolte'* sintagmaren adierazpide litzatekeena (|détruire \* \*##\*| unitatea, irudian). Bi konfigurazio horien sorkuntzak bereizten ditu, guk uste, bi ikuspegiak. Konfigurazio horiek, gainera, beste inon ere erreferentzia litezke; konparazio batera, balizko C' kontzeptu baten errepresentazioan, bere definizioa *'feu qui détruit les récoltes'* delarik, C kontzeptuaren errepresentaziorako sortutako konfigurazioaren erreferentzia egiteko aukera legoke<sup>1</sup>.

Sowaren (1984) grafo kontzeptualen eredua, § III.1.1en ikusi dugunez, ezagutzaren errepresentaziorako lengoia oso bat da, sare semantikoen ildokoa. Hemen erakutsi nahi genukeena formalismo horretako grafo mota jakin bat da —mota-kontzeptuen definizioei dagokiena ("type definitions")—, eta hitz-zentzuen errepresentazioan formalismoaren aplikazioari buruz (Pazienza & Velardi, 87)n egindako oharpen batzuk komentatu.

Esan gabe doa grafo kontzeptualen eredu honek, helburu orokorrekoa izaki, adierazteko ahalmen handiagoak dituzkeela gureak baino, gurea, azken finean, definizio-testuen errepresentaziorako soilik asmatua baita. Aipatzekoak lirateke, bereziki, kuantifikazioari edota ukapenari dagozkionak, kontzeptu baten instantzia edo banakoen errepresentazioa, notazio linealean korreferentziak adierazteko erabil daitezkeen aldagaiak, eta abar, Sowarena orotariko ezagutzaren errepresentaziorako sistema oso bat egiten dutenak.

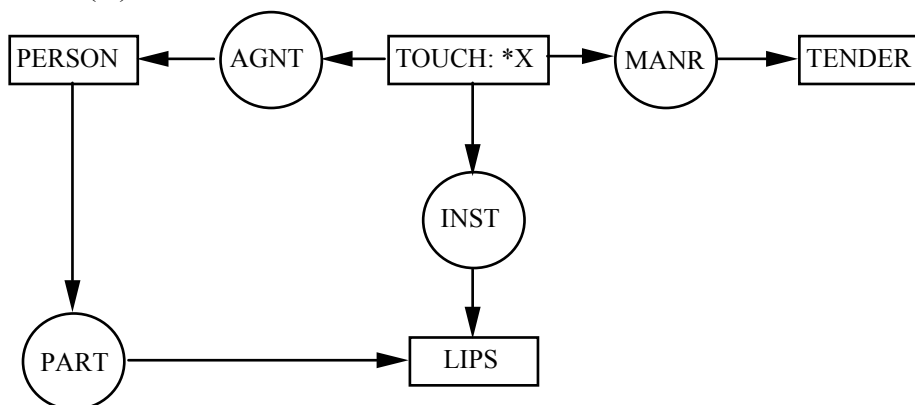
Grafo kontzeptualon ereduan, "type definition" delako grafoa mota-kontzeptu berri bat errepresentatzeko erabitzen den grafo mota berezi bat da; errepresentazio hori "differentia" izeneko grafo baten funtzioan egiten da. Sowaren liburuan (1984) esaten denez, grafo mota honen helburua "genus et differentia" eran definitutako kontzeptuak errepresentatzea da. Ikus dezagun bada (III.14 irudia) Sowaren liburu horretatik ateratako adibide bat, *(to) kiss* kontzeptuari dagokiona (definizioa testuan esplizituki ageri ez bada ere, errepresentaziotik ondoriozta daiteke gutxi gorabehera *'to touch with the lips (of a person) in a tender manner'* dela). Sowaren "type definition" horiek LISPezko funtzio-definizio gisa ikus daitezke:

---

<sup>1</sup> Hala ere, oraingoz ez da egiten, eta bereziki ez da egin HEBaren prototipoa sortzeko erabili programan, sortu beharreko konfigurazioen kopurua era honetan murriztuko lukeen egiaztapenik. Gainera, eragiketa hori ez litzateke kontuan hartu behariko sinplifikazio kasu horietan bakarrik: izan ere, benetan behar dena konfigurazioon sailkapen-mekanismo orokor bat baita. Sailkapen-mekanismo horrek, arrazoiketa terminologikoko sistemetako sailkapenaren moduan (ikus § III.1.1), gauza izan behar luke sailkatzeko, adibidez, *'feu qui détruit les récoltes'* errepresentatzen duen konfigurazio bat *'feu intentionnel qui détruit les récoltes'* errepresentatzen duenaren gainean (hiperonimo bezala, alegia), biak hun-kontzeptu berberaren hiponimo —*'feu'* mota-kontzeptua— egiteaz gainera. Horrela lortuko genuke kontzeptuarteko taxonomia konfigurazioetara ere benetan hedatzea.

abstrakzio bati etiketa bat esleitzeko metodo bat (hortik, hain zuzen, X parametro formalaren erabilera, grafoan genusa zein den adierazteko):

**type KISS (X) is**



**III.14 irudia.-** (to) kiss-en mota-definizioa ("type definition") (Sowa, 84:106).

Adierazpide horrek badu bere antza gure eremuan definizio honen errepresentazioari legokiokeen |touch \* ##\*| konfigurazioarekin (ikus III.15 irudia), III.14 irudi horretako nodo kontzeptualak —laukizuzen tankerakoak— beren etiketei dagozkien mota-definizioen erreferentziatzen hartzen baditugu behintzat. Bien arteko alde ohargarri bat ordea, grafo kontzeptualak berrorientagarriak izatea da, hau da, grafo berak balio dezakeela errepresentazio desberdinetarako, grafoaren "burua" zein nodo den finkatzen ez delako<sup>1</sup>. III.15 irudian ikus daiteke zein litzatekeen gure ebazpidea definizio hori errepresentatzeko, kontzeptuarteko erlazio semantikoak zeintzuk diren badakigula suposatuz<sup>2</sup>.

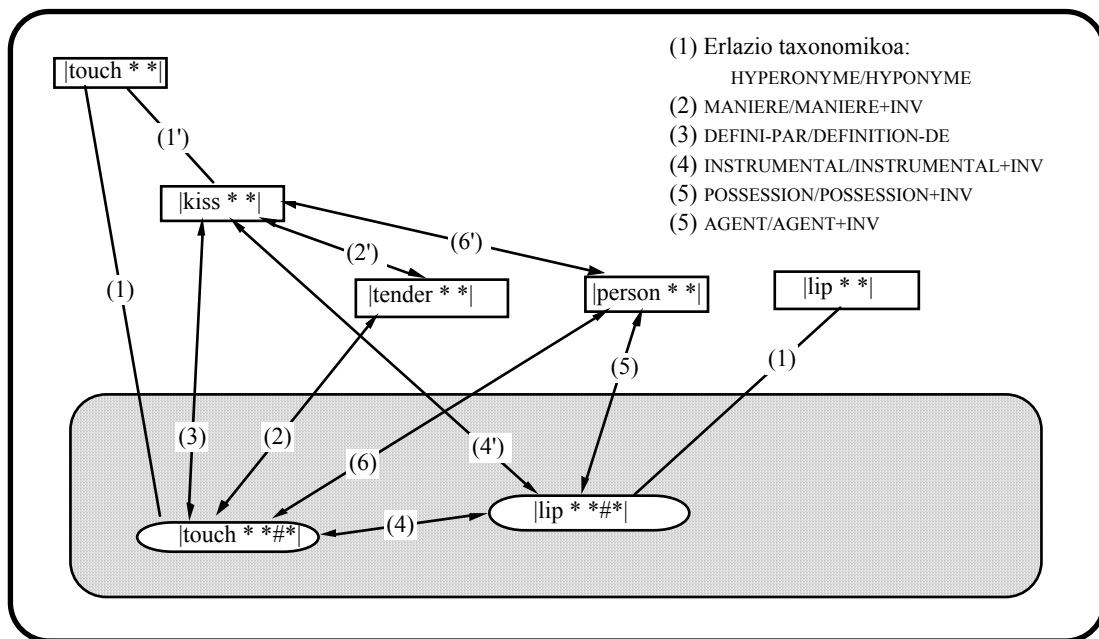
Grafo mota hau —"type definitions"— (Pazienza & Velardi, 87)n erabiltzen dute mota edo klase abstraktuen hierarkia bat eratzeko, hitz-adierak errepresentatzeko oinarrizko grafoak erabiltzen dituztelarik (20 grafo adierako, batez-beste). Beraz, bi kontzeptu-maila bereizten dituzte: klase abstraktuak batetik eta hitz-zentzuak bestetik; horrela, hitz-zentzuak, hierarki eratzaile baino klase abstraktuen hierarkiako "hostoak" lirateke. Adibidez, *house* kontzeptuaren definizioak —*building for residence*— badu, bere CLASSIFICATION deskriptorean ondoko mota-definizioa ("type definition"), oraingoan grafoaren baliokide den forma linealean adierazia (Pazienza & Velardi, 87:253):

<sup>1</sup> Izan ere, "type definition" hauetan grafo-burua parametro formala daramana da. Sowa-ren liburuan (1984:106-107) ikus daiteke '*circus-elephant*' kontzeptuaren errepresentazioa eta '*elephant-circus*'-enaren arteko alde bakarra zera dela, alegia lehenengoan parametro formala daraman mota '*elephant*' den bitartean —'*circus-elephant*' hitz elkartuaren genusa denez— bigarrean '*circus*' mota dela. Grafoon berrorientagarritasun horrek implementazioaren aldetik eduki ditzakeen gorabeheraz, ordea, ez diosku ezer.

<sup>2</sup> Berrito diogu kontua orain ez dela erlazio horiek lengoai naturalerako definizio batetik nola erauzi, baizik eta azken errepresentazioa, atributu erlazional egokiak erabiliz, nolakoa litzatekeen ikustea.

[BUILDING] <— (MEANS) <— [RESIDENCE]

non BUILDINGek genusa edo gainmota<sup>1</sup> errepresentatuko lukeen eta gainerakoak, (MEANS) <— [RESIDENCE] espresioak alegia, differentia. Kasu batzuetan grafoek errepresentatzen dituzte hitz-zentzu biren arteko erlazioak, baina jeneralean goiko adibide horretan bezala gertatzen da, eta erlazioak klase abstraktuekin ezartzen dira. Kontzeptu-definizioek bi deskriptore izaten dituzte, ikusi dugun hori —CLASSIFICATION— eta DEFINITION izenekoa, non kontzeptuaren egitura eta funtzioak adierazten diren. Azken deskriptore horrek hiru eremu izaten ditu, erlazio kontzeptualen hiru kategoriei dagozkienak: ROLE, COMPLEMENT eta LINK<sup>2</sup>. Gurean bezala, proiektu honetan ere alde zehatzak zehaztu dute, eta gutxi edo gehiago finkatu, erlazio —primitibo— multzo bat: multzo horren ezarpena iristeko, italiarren kasu-gramatikazko analisia egin dute ("kasu-gramatika" (Fillmore, 68)ren zentzuan hartuta) eta sakonki aztertu dituzte aplikazio-eremuan aurkitutako adibideak.



III.15 irudia.- (to) kiss-ek denotatutako egintzaren errepresentazioa<sup>3</sup>.

<sup>1</sup> Gurean ez bezala, hemen BUILDING ez da hitz-adiera baten erreferentzia baizik eta hierarkiako klase abstraktu batena. Parentesi arteko espresioak —(MEANS)— erlazio motako nodo bat adierazten du.

<sup>2</sup> ROLEN egoten dira AGENT, OBJECT, MEANS, PURPOSE, etab. bezalako erlazioak. COMPLEMENTeko grafoek entitate baten egitura edota egintza baten gertaera deskribatzen dute, STATUS, TIME, MANNER, QUALITY, PART-OF, eta gisako erlazioez. LINK en azkenik, POSSESSION, INCLUSION, OPPOSITE, PRECONDITION, etab. bezalako erlazioak egoten dira.

<sup>3</sup> Irudiko konfigurazioek errepresentatu testuak:  
|touch \* ##\*|: *to touch with the lips (of a person) in a tender manner*  
|lip \* ##\*|: *lip of a person*

### III. kapitulua

Alderaketa sail honi bukaera emateko berriz, Chouraqui eta Godbert-ek (1989) proposatzen duten ezagutza-errepresentaziorako formalismoa hartuko dugu aztergai; autore hauek, formalismoa zehaztu eta gero, hiztegi-definizioen semantikari aplikatu diote probatzearren.

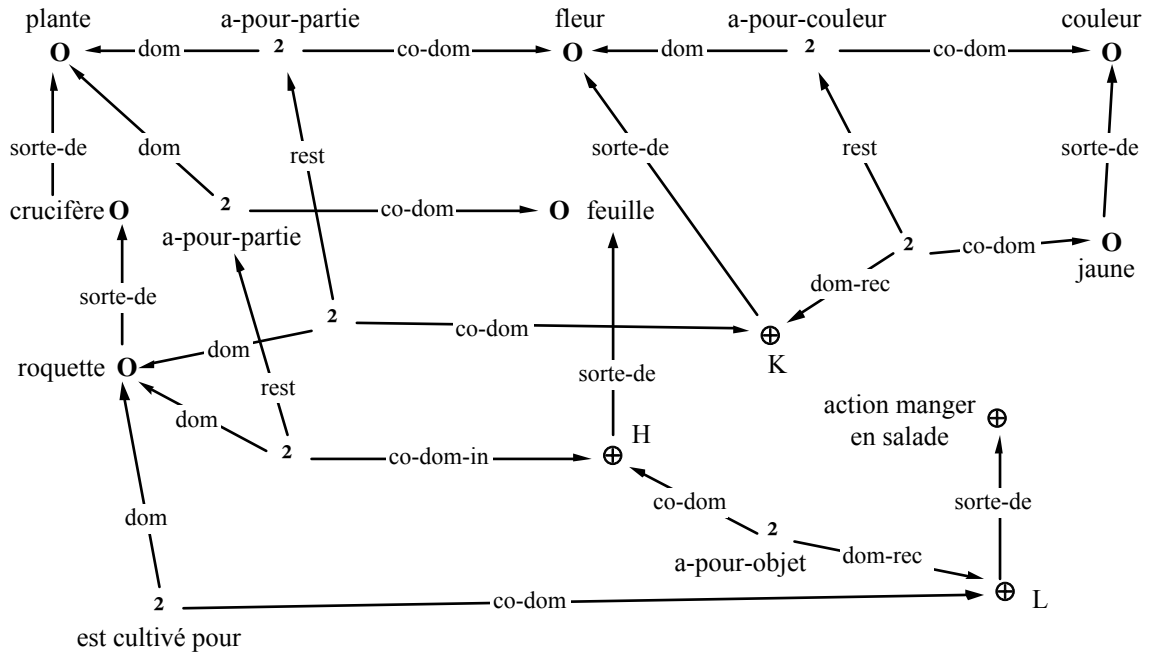
(Chouraqui & Godbert, 89:866)en definitzen dira “. . . *les fondements d'un réseau de type SI-Net, dont la structure et la sémantique, définies à partir d'un ensemble de types de nœuds et de liens et fondées sur les notions classiques de concepts, individus et relations descriptives, permettent d'exprimer à la fois l'interprétation intensionnelle et l'interprétation extensionnelle des objets à représenter.*” Beren ikuspegia interesgarria da berez, baina batez ere egin duten formalizatze saioagatik. Esan bezala, definitutako formalismoa probatzeko hiztegi-definizioen errepresentazioari aplikatzen diote; hala ere, errepresentazio hori lengoaiia naturalezko definizio-testuetatik nola atera ez dute erakusten. Eta azken puntu hori inportantea da, batik bat kontuan hartzen badugu beraiek erabilitako "erlazio" kontzeptua nolakoa den: aldez aurretik erlazio "primitiborik" definitu gabe, erlazio motako nodoak sortarazten dituzte sarean, hala nola '*a-pour-partie*', '*a-pour-couleur*' edo '*est-cultivé-pour*' bezalakoak.

III.16 irudian ikus daiteke ondoko definizio honi dagokion errepresentazioa:

***roquette***: *crucifère à fleurs jaunes, cultivée pour ses feuilles, que l'on mange en salade*

Irudiko errepresentazioa dela eta, ondokoak azpimarratu nahi genituzke:

- grafoko geziak "lokailuen" adierazpide dira, eta "erlazioak" erlazio motako nodoen bitartez ( $\Delta$ ) adierazten dira;
- espezializazioa ez dagokie soilik kontzeptu-nodoei baizik eta baita erlazio-nodoei ere (formalismoa maila "epistemologikoan" kokatzen da, (Brachman, 79)ren zentzuan);
- erlazioak ongi zehaztuak dira egiturazko lokailuen bidez: domeinua, kodomeinua, murriztapena (erlazio mailako espezializazioa), eta abar.
- autoreek diotenez, K litzateke '*fleur jaune*' kontzeptua, H '*feuille de roquette*', eta L '*manger des feuilles de roquette en salade*' egintza, zeinentzat '*action manger en salade*' kontzeptu definitua erabiltzen baita (*manger* eta *salade* kontzeptuekin eratutako kontzeptua).



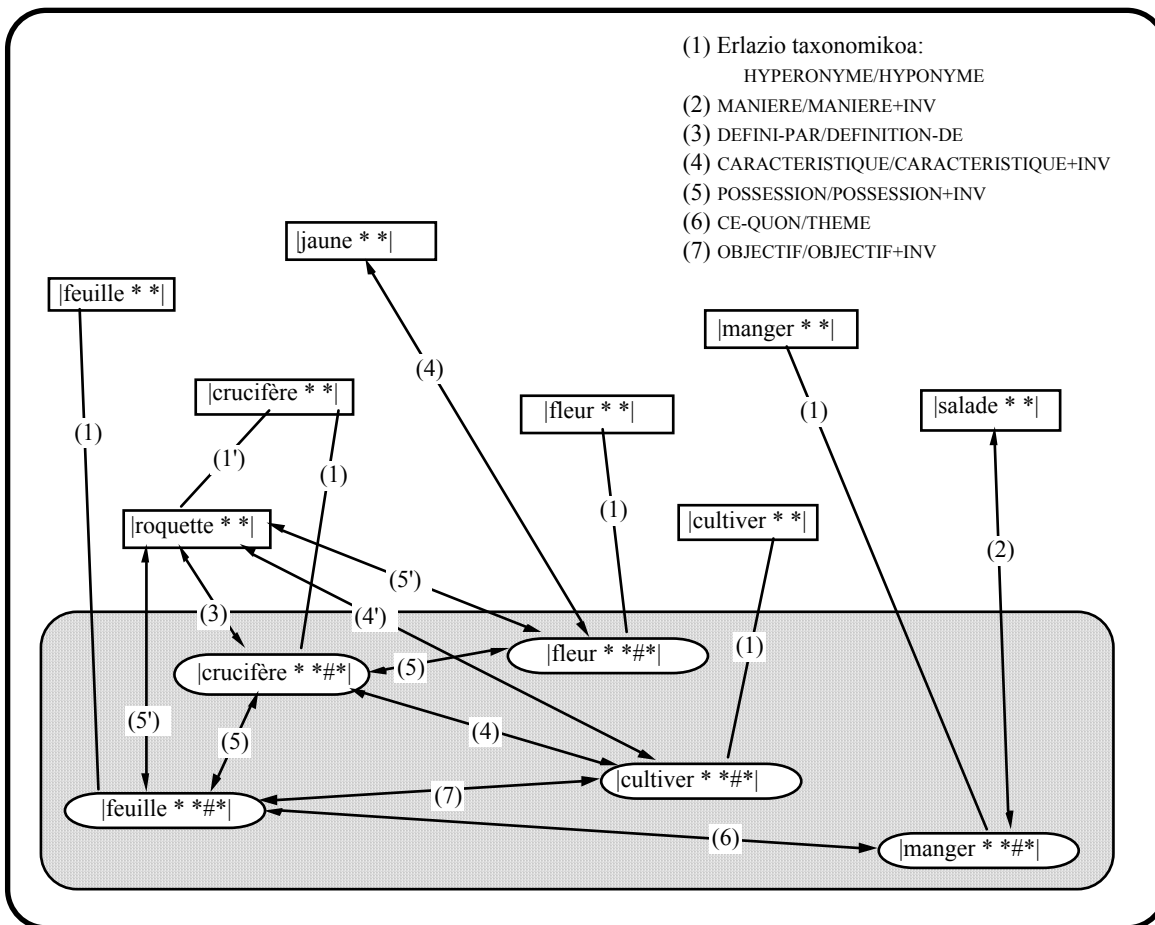
**III.16 irudia.-** *roquetteren* definizioaren errepresentazioa (Chouraqui & Godbert, 89:866).

Esan behar da III.16 irudian badela emandako definiziotik atera ez den parterik: esate baterako, *crucifère* eta *planteren* arteko lotura, *plante* eta *fleur* a-pour-partie erlazioaren bitartez lotuak egotea, eta beste, ez baitago inola ere hori ateratzerik emandako definiziotik soilik.

III.17 irudian ikus daiteke nola errepresentatuko litzatekeen definizio hori bera gure ereduaren arabera. Hemen ere esan behar da definizio honek baduela hainbat arazo sintaktiko-semanticoko, bere analisi automatikoa benetan zaila bihurtuko zuketena. Beraz, orain ere suposatzen beharko dugu gai izan garela *ses* posesiboaren erreferentziaren arazoa askatzeko, eta orobat azken erlatibozko perpausarena (autoreek emandako errepresentazioaren arabera, '*que l'on mange en salade*' perpausaren aitzinkaria *feuilles* litzateke).

Chouraqui eta Godbert-ek sekundariotzat hartzen dituzten kontzeptuak —H, K eta L irudian, esate baterako— gure konfigurazioen hurreko dirateke. Izan ere |feuille \* ##\*| konfigurazioa dagokio H-ri eta K litzateke gure |fleur \* ##\*|. Halaber, beraiek L nodoaren bitartez adierazia, '*manger des feuilles de roquette en salade*', gure |manger \* ##\*| aditz-konfigurazioari legokioke<sup>1</sup>.

<sup>1</sup> Kasu honetan, definizioko erreferentzi arazoaren ebazpenak |feuille \* ##\*| konfigurazioa |manger \* ##\*|en errepresentazioan berrerrabiltzeko aukera emango luke, bien arteko CE-QUON erlazioaren bitartez lotuz.



III.17 irudia.- roquetteren definizioari legokioken errepresentazioa<sup>1</sup>.

Gure ustez bada, bi ikusmoldeon arteko alderik nabarmenenak —aztertu errepresentazio horiei dagokienez— hauek lirateke: batetik erlazio primitiboen aukera, Chouraqi eta Godbert-enean ezinbesteko ez dirudiena; bigarren, gurean ez dagoen erlazioen espezializazioa; eta azkenik, beraiek lokailu eta erlazioen artean eginiko bereizketa. Bestalde ezin da ahaztu (1) autore horien formalismoaren helburua ezagutzaren errepresentazioa dela, oro har, inondik ere ez dena gure ereduarena, eta (2) gurean abiapuntua benetako hiztegi bat izan dela —nahiz definizioak errepresentatzeko ezarri ditugun atributu erlazionalen hautapenerako nahiz HEBaren eraikuntzarako—, Chouraqi eta Godbert-ek grafoak eskuz egiten dituzten bitartean, beren errepresentazio-ereduaren baliagarritasuna egiaztatzearen.

<sup>1</sup> Irudiko konfigurazioek errepresentatu testuak:  
|crucifère \* \*#\*|: *crucifère à fleurs jaunes, cultivée pour ses feuilles, que l'on mange en salade*  
|feuille \* \*#\*|: *feuille de roquette* (konfigurazio hau *ses* posesiboaren interpretazioa dela medio sortua dateke)  
|fleur \* \*#\*|: *fleur jaune*  
|cultiver \* \*#\*|: *cultivé pour les feuilles (de la roquette)*  
|manger \* \*#\*|: *manger (les feuilles de la roquette) en salade*



### **III.1.4 HEBaren prototipoaren ikuspegi orokorra: zenbait datu.**

Atal honetan HEBaren prototipoari buruzko ikuspegi orokor bat eman nahi genuke, batik bat berorren tamainari buruzko zenbait datu.

STRUCTURES azpi-EBari dagokionean aipatua dugu behin baino gehiagotan txosten honen D eranskina, bertan aurki baitaiteke azpi-EB horren egungo egoerari dagokion listatu osoa eta baita bertan ezarritako klase-hierarkiaren zenbait bista ere.

DICTIONNAIRE azpi-EBak bere aldetik, 2400 unitate dauzka: bakoitzak hiztegiko hitz bat errepresentatzen du eta estekatua dago —SENS atributuaren bitartez— THESAURUSen dagozkion kontzeptuekin, bere adiera desberdinekin alegia. DICTIONNAIREko unitate hauek STRUCTURESEko ENTREES klaseko elementu edo instantzia dira.

THESAURUSen berriz 6130 kontzeptu desberdin dauzkagu —hasierako eraikuntzan sortutako konfigurazio eta erreferentzia edo "kontzeptu hutsak" barne—:

- 2607 DEFINITIONS klaseko elementu<sup>1</sup>
- 1785 REFERENCES klaseko elementu
- 1738 CONFIGURATIONS klaseko elementu

DEFINITIONS klasean 1461 izen, 844 aditz et 302 adjetibo daude.

Hasierako eraikuntza burutu eta gero, erreferentzien (REFERENCES) %70ak aurkezten du *ambiguetate mailaren bat*, gainerantzekoak erreferentzia ez-ambiguoak direlarik. Erreferentzia ambiguen artean gehienek adiera mailako ambiguetatea dute. Hona hemen kopuru batzuk:

- 1125 elementu ACCEPTION klasean (adiera mailako ambiguetatea)
- 115 elementu HOMOGAPHE klasean (homografo eta adiera mailako ambiguetatea)
- 15 elementu COMPLEXES klasean (erreferentzia konplexuak, adiera bati baino gehiagori dagozkionak)

Mota-kontzeptuen hierarkiari dagokionean berriz, CONCEPTS-TYPEren azpiklaseak honela ageri dira (konfigurazio-unitateak ere barne):

- 3548 azpiklase ENTITESen

---

<sup>1</sup> FFtik prototiporako iturburutzat hartutako 2607 adierei dagozkienak, hain zuzen ere.

### III. kapitulua

- 1948 azpiklase ACTIONS/EVENEMENTSen
- 771 azpiklase QUALITESen
- 11 azpiklase ETATSen<sup>1</sup>

Eman nahi genukeen beste ikuspegi bat, THESAURUS azpi-EBa sare semantiko gisa aurkeztu dugunez, bertako kontzeptuen artean ezarritako "arkuen" kopurua da.

Behin hasierako eraikuntzaren urratsa burututakoan, eraikuntza taxonomikoa eginda eta zuzeneko zein alderantzizko erlazio batzuk ezarri eta gero beraz, THESAURUSen 19691 kontzeptuarteko arku kontatu ditugu —jakina, atributu erlazionalei dagozkienak besterik ez ditugu hemen kontatzen—. THESAURUSeko kontzeptu bakoitza bada, batez-beste 3.21 kontzepturekin estekatua dago hasierako eraikuntza horren ondoren.

Eta, oraingoan konfigurazio-unitateak kontatu gabe, ENTITESen azpiko unitateek (REFERENCES zein DEFINITIONSeko elementu izan) batez-beste 3.7 lotura dituzte THESAURUSeko beste unitateekiko; kopuru hori 3.2koa da ACTIONS/EVENEMENTS eta QUALITESen kasuan eta 1.89koa ETATSen azpiklaseentzat.

Konfigurazioek, berriz, 2.7 lotura dauzkate batez-beste.

Ohargarria da, inondik ere, REFERENCES klaseko elementuen lotura kopurua aski handia dela, kontuan izanik batez ere hasiera batean "hutsik" sortutako kontzeptu-unitateak direla horiek<sup>2</sup>: 3.47 lotura batez-beste.

Eta azkenik, azpimarratu bakarrik hasierako eraikuntzan ezarritako 19691 kontzeptuarteko erlazio horiek zer-nolakoak diren: 11538 dira lotura esplizituak —zuzeneko zein alderantzizko—, 7003 lotura hierarkikoak (ANCETRE<sup>3</sup>, DESCENDANT, HYPERONYME, HYPONYME), eta gainerakoak —1150 erlazio— herentziaz lortutako balioei dagozkie.

## III.2 HEBaren aberasketa dinamiko erdiautomatikoa.

HEBaren gainean egindako aberaste-prozesuak —ezagutza-basea eraikitzean nahiz ondorenean— deskribatzea da atal honen helburua. Hurrengo atalean (§ III.3) azalduko dira

---

<sup>1</sup> Kontuan izan behar da THESAURUSeko unitate guzti-guztiak daudela CONCEPTS-TYPE klasearen azpian, eta, batzuetan, beronen azpiklase batean baino gehiagotan (partizipioek denotatu kontzeptu-unitateen kasua litzateke, adibidez).

<sup>2</sup> Dena den, erreferentzia bat sortzen bada beste kontzepturen batetik erlazioen bat "iristen" zaiolako sortzen da. Orduan, alderantzizko erlazioa kudeatzen duen iratxoak erreferentziatik "abiatzen" den arku bat sortzen du, eta hartara, erreferentzia hori ez da gehiago "kontzeptu hutsa" izango.

<sup>3</sup> Ikus § III.3.1en.

HIZTSUAren *azpiegitura deduktiboa* osatuko luketen dedukzio-mekanismoak, eta batik bat prototipoan ezarritakoak.

Prozesu eta mekanismo horiek oro, HIZTSUAren ahalmen deduktiboaren parte diren neurrian, sistemaren Dedukzio-Moduluari dagozkio (ikus § I.3.1.1).

HEBaren *aberasketa dinamiko erdiautomatiko* esaten diogu bertako ezagutza hedatzearen prozesuari. Prozesu hori HEBan ezarritako prozedura multzo baten bitartez burutu da, bi momentu jakinetan:

A) HEBaren hasierako eraikuntza burutu ahala (§ III.2.1)

B) HEBaren prototipoaren lehen bertsioa sortu eta gero (§ III.2.2)

Zehatzago esateko, "aberaste" diogunean inferentziarekin edo egitate deduzituekin zerikusia duen guztia esan nahi dugu. Gure helburua, hainbat aldiz esan dugu, HEBko ezagutzaren ustiapena baita: nola bihurtu esplizitu hiztegiko ezagutza implizitua. Kontua beraz, hiztegiko ezagutzen gainean arrazoiketa-mekanismo sail baten ezarreran datza.

Esan gabe doa arrazoitze horrek bere mugak izango dituela. Muga horiek informazio iturritzat izan dugun hiztegi txikiak berak jartzen dizkigu lehen-lehenik, baina prototiporako bidean jarraitu metodoak ere areagotu egin dituzke: analisi sintaktiko-semanticoko partziala, lengoia naturalazko edozein testutan dagoen anbiguotasuna, eta, dudarik gabe, eraikuntza-prozesuan egindako hutsegiteak.

Litekeena da, bestalde, sistemaren portaera inferentziala ezartzean ere huts egitea. Hala eta guztiz ere, biziki interesgarria dirudi ezagutza partzial, betegingabe, anbiguoarekin lan egiteak, hortik zer lor daitekeen jakiteko erronkari erantzun bat eman nahi baldin bazaio behintzat. Gure interesa beraz, horrela jokatzek dituzkeen arazoak nondik norakoak diren aztertzea, eta jokamolde horrek, erlazio lexiko-kontzeptualen inguruko arrazoiketa deduktiboa gaitzat duen ikerketarako, zer ekar lezakeen ikustea izan da.

### **III.2.1 HEBaren hasierako eraikuntzan lortutako aberasketa.**

THESAURUS azpi-EBa, izaera semantikodun atributu erlazionalen bitartez elkarlotutako kontzeptuen sare bezala bada, jadanik, errepresentazio maila definitorioarekiko aberatsago den zerbait. Atal honetan deskribatuko dugu zertan datzan HEBaren hasierako eraikuntza burutzean izandako aberastea, ondoko alderdiok kontuan hartuz:

- errepresentazio maila definitivotik erlazionalerako balioen goratzea;
- definiensaren adierazpidea den unitatetik dagokion definiendumarenerako balio-intsuldaketa;

### III. kapitulua

- erlazioak bi zentzuetan mantentzea:
  - kontzeptutik hitzerako atzibidea;
  - erreferentzia "hutsak";
- kontzeptuarteko egitura taxonomikoaren eraketa (esplizitu bihurtzea):
  - propietateen herentzia

§ III.1.3.2.3n aipatu dugu jada, HEBaren hasierako eraikuntzari ekin aurretik *exekuzio automatikoko prozedura* batzuk definituak genituela STRUCTURES azpi-EBan. Prozedura edo iratxo horiek lanean ihardun dute HEBa eraiki ahala —aberasketaren lehen fase edo urratsean—, berehala azalduko ditugun ekintzak burutuz.

Halaxe, bost iratxo desberdin aritu dira lanean THESAURUSeko kontzeptu-unitateak sortu ahala (ESEEtan zehaztutako ekintzen ondorioz —unitate-sorkuntza eta unitate-atributuen balio-esleipenak—, definizioetan agertzen ziren kontzeptuetan eraginez)<sup>1</sup>:

- DEFINITOIRE-RELATIONNEL eta SYNTAGMATIQUE-RELATIONNEL izenekoak: iratxo horien zeregina *atributu definitorio, sintagmatiko eta metalengoaiazkoetatik<sup>2</sup> dagozkien erlazioaletarako balio-goratzek* egitea da. DEFINITOIRE-RELATIONNEL iratxoa atributu definitorio guztietan kokatu da; SYNTAGMATIQUE-RELATIONNEL berriz, sintagmatiko eta metalengoaiazkoetan. Balio-goratzek egiteko, iratxo hauek automatikoki exekutaraziko dira atributu definitorio, sintagmatiko edo metalengoaiazko baten balioa eguneratzera doan bakoitzean: *RELATIONNELS-CORRESPONDANTS* fazetaren balioari erreparatu, fazeta horrek adierazten baitu zein edo zeintzuk atributu erlazioaletara pasa behar den balioa, eta kopia egingo du.

Atributu definitorioetatik erlazioaletarako goratzea berehalakoa baldin bada ere<sup>3</sup>, sintagmatikoen kasuan gauza ez da horren erraza, iratxoak erreparatu egin behar du eta —kopiaketari ekin aurretik— ziurtasun-baliodun fazeta erlazioalen zerrenda (§ III.1.3.2.2), balioaren goratzea posible den ala ez jakitearren: hasierako

---

<sup>1</sup> Iratxo guzti hauek eguneratzeak egindakoan exekututzen direnetakoak dira, hau da, bere atributuan balio berri bat esleitzerakoan edota balio bat ezabatzerakoan. LISPEz idatzita daude, beren adierazpidea den unitateko AVPUT izeneko slotaren balio gisa.

<sup>2</sup> METALANGAGIERS1 atributu-klasekoak.

<sup>3</sup> Gogoratu atributu definitorioen funtzioa HDBtik datorren informazio analizatua errepresentatzea dela, erlazioaletan ezagutza deduziturik eduki daitekeen bitartean (jatorrizkoaz gain, jakina). Beste adibide bat ematearren, esan dezagun DEF-PARTIED definizio-informazioa gordetzeko atributu bat dela, hau da iturburu-hiztegiaren lexikografoak halako formula espezifiko batez definitutakoak jasotzeko asmatua. PARTIE-DE aldiz, meronimi erlazio lexikala errepresentatzen duen atributu erlazioala da, kontzeptu bat beste batzuekin lotzen duenean hura hauen parte bat dela adierazten duelarik. Azken horien artean izango da, dudarik gabe, DEF-PARTIED atributuaren balioa ere, baina ez derrigor bera bakarrik.

eraikuntzaren fasean, balio-goratzeko bat bidezko zatitzaile jokoaren fazeta erlazionalen balioa 0.5 baino handiago izatea eskatu dugu<sup>1</sup>.

- DEFINIENS-DEFINIENDUM: iratxo honen zeregina *konfigurazioen atributu erlazioetatik konfigurazio horiek definitzen dituzten kontzeptuen atributu izenkideetara balio-intsuldaketa* burutzea da. Hots, iratxo honek —atributu erlazional guztietan kokatua— kontzeptu baten definizioaren errepresentazio diren konfigurazioetan egingo du bere lana (kontzeptua eta konfigurazioaren artean —definienduma eta definiensa— DEFINI-PAR/DEFINITION-DE erlazioak daudenean). Horrela, iratxoak egiaztatu egin behar du balio-eguneratzea definiens erako konfigurazio batean doala gertatzera eta, hala bada, balio hori kopiatu definiendumaren unitatean ere.

Iratxo hau non kokatu eta non ez inportantea izango da, zeren horrek zehazten baitu zeintzuk diren definiensetik definiendumera automatikoki "pasaraz" daitezkeen erlazioak eta zeintzuk ez (ikus adjetibo batzuen erlatibotasunari buruzkoa § III.1.3.4en). Prototipoa eraikitzerakoan hala ere, atributu erlazional guztien balioak intsuldatzea erabaki dugu.

- RELATION-INVERSE: iratxo honi zor zaio *erlazioak bi zentzuetan mantentzea*; horretarako,  $C_1$  unitatean  $R$  atributu erlazionala eguneratzen denean  $C_2$  balioa hartuz, iratxoak  $R$ -ren *INVERSES-CORRESPONDANTS* fazetara jotzen du atributu horren alderantzizkoa zein den jakiteko, eta  $C_2$  unitateko atributu horretan —alderantzizko erlazioa adierazten duen horretan, hain zuzen— gordetzen du balio bezala  $C_1$ . Iratxoak berdin funtzionatzen du balio-ezabaketak egitean<sup>2</sup>.

Azpimarratzekoa da alderantzizko erlazio horiek, batzuetan, erlazio lexikalen propietateak islatzen dituztela: adibidez, kontzeptu baten sinonimoak dauzkan SYNONYMES atributuaren alderantzizkoa SYNONYMES bera da, horrela sinonimiaren simetria alderantzizko erlazioaren bitartez inplementatzen da eta.

---

<sup>1</sup> Gorago esan dugun legez, ziurtasun-balio horietako batzuen zehaztapena ESEEtan egin ahal izan da.

<sup>2</sup> Aukeratu dugun programazio-inguruneak ez ditu alderantzizko erlazioak automatikoki kudeatzen. Eraginkortasunak eta sistemaren konsistentziak eskatzen dute, gure ustez, alderantzizko erlazio horiek HEBan esplizitu egitea. Alderantzizko erlazioak iratxoaren bitartez inplementatzea aski egokia bada ere, nahiago izango genuen KEEren egiazen mantenimendurako sistema (TMS) erabiliz egitea. TMSri esker, egitate "deduzituak" HEBan justifikazio ("assumption") batzuei lotuak geratzen dira, eta sistema bera da gero justifikazio horien kudeaketa egiten duena. Hartara,  $F'$  egitatea sortarazi duen  $F$  desagertuko balitz ( $F'$  hori justifikazio batzuei loturik egongo da),  $F'$  ere desagertuko litzateke. TMS erabiliz bada, alderantzizko erlazioak "zuzenekoei loturiko" egitate deduzituak lirateke. Kontzeptuak elkarrekin lotzen dituzten atributu erlazionalak klaseko sloten bitartez ("member slots") inplementatu dira, hiponimoetarako balio-herentzia posible izan dadin. Tamalez, erabilitako KEEren bertsioak ez du aukerarik ematen justifikazioak klaseko slot horien bitartez adierazitako egitateei esleitzeko, baizik eta soilik berezko sloten ("own slots") bitartez adierazitakoei. Hori dela eta, ezinezkoa izan zaigu alderantzizko erlazioak TMS mekanismo horren bidez inplementatzea.

### III. kapitulua

RELATION-INVERSE iratxo honi esker lortutako beste aberasketa bat, DICTIONNAIRE eta THESAURUS azpi-EBetan egindako ezagutzen egituratzetik datorrela esan dezakegu: DICTIONNAIREko sarrera-hitzak beren adierekin (THESAURUSen) loturik daude SENS atributuaren bitartez —ENTREES klasean definitua— eta, ordainez, THESAURUSeko kontzeptuak dagozkien sarrerarekin lotuak daude MOT-ENTREE delako atributuaren bitartez (CONCEPTS-TYPEn definitua, SENSEn alderantzizkoa). Hortaz, HIZTSUAN posible da *kontzeptutik hitzerako zuzeneko bidea*, hitzetik kontzepturako ohizkoaz gain.

Erlazioak bi zentzuetan mantentzeak duen albo-ondorio bat, oso inportantea baina, REFERENCES klaseko elementuak "berezko" informazioaz janztea da: izan ere, REFERENCESeko  $R_C$  unitatea sortzen delarik, C beste kontzeptu baten definizioan erreferentziatua izan delako alegia, erlazioak bi zentzuetan ezartzeak  $R_C$ -rantz doan erlazioaren alderantzizkoa  $R_C$ -tik C-rantz ezartzea esan nahi du. Kontzeptu bat zenbat eta gehiago aipatua izan besteren definizioetan orduan eta "osatuago" geratuko da, alderantzizko erlazio horiek direla medio<sup>1</sup>.

- CONSTRUCTION-TAXONOMIQUE: iratxo honek *egiten du esplizitu THESAURUSen kontzeptuarteko egitura taxonomikoa* (hiperonimia/hiponimia erlazioak), definizio klasikoen bitartez ezarria. Iratxo hau DEFINI-PAR atributu erlazionalean dago kokatuta —atributu horren bitartez adierazten dugu definienduma eta definiensaren arteko lotura, "genus et differentia" erako definizio klasikoetan—. Bere zeregina ondokoa da: atributuari dagokion kontzeptua taxonomian kokatu, kontzeptuaren definiensa errepresentatzen duen konfigurazioaren gainklasearen azpian hain zuzen. Horrela, definienduma ere definiensaren gainklasearen hiponimo bihurtzen da (izan ere, gure errepresentazioan, definienduma eta definiensa errepresentatzen duten unitateak kontzeptu baliokidetzat hartzen ditugu).

*Bigiztei* dagokienean berriz —egitura taxonomiko hau eratzerakoan sort daitezke eta—, esan behar da ez dela posible  $C_2$  kontzeptua  $C_1$ -en azpiklase egitea  $C_1$   $C_2$ -ren azpiklasea bada. Bestalde, jakina da hiztegian beti aurkitu ohi dela horrelako bigiztarik, hiztegiaren berezko ezaugarri bat baita zirkulartasuna (Vossen & Serail, 90): hitzak erabiltzen dira hitzak definitzeko. CONSTRUCTION-TAXONOMIQUE iratxoak horrelako bigiztak daudenentz egiaztatzen du eta, egonez gero, era berezi

---

<sup>1</sup> HEBan kontsistentzia eta osotasuna mantentzea inportantea da oso. Hasierako eraikuntzan erreferentzia edo "kontzeptu huts" horiek sortzeko hartutako erabakia osotasunerantz emandako lehen urrastzat har daiteke: HEBan ez dago, THESAURUSen errepresentaziorik —kontzeptu-unitate, nodorik— ez daukan kontzeptu-erreferentzia bat bera ere. Atributu erlazionaletan kokatu iratxoek berriz —RELATION-INVERSEK, adibidez—, kontsistentziaren aldetik beren garrantzia dute, eta kasu honetan, adibidez, atributuak errepresentatutako erlazioaren alderantzizkoa automatikoki kudeatuko dela segurtatzen digute.

batean errepresentatzen ditu, inolako informaziorik ez galtzarren.  $C_1$  et  $C_2$ -ren artean SUBCLASS.OF lokailua ezarri baino lehen, iratxo horrek erreparatzen du ea  $C_2$ -ren gainklaseen artean —hiperonimoak—  $C_1$  dagoen. Hala izanez gero, CONSTRUCTION-TAXONOMIQUEk  $C_2$   $C_1$ -en HYPERCYCLE slot berezian gordeko du balio bezala. Horretara, slot horrek edukiko lituzke hierarkian bigiztarik sortu zuketuen kontzeptuaren hiperonimoak<sup>1</sup>. Bistakoa denez, jokabide honek badu arazo bat: propietateen herentzia ez da automatikoa izango bigiztak dauden kasuetan. Oraingoz ez da ezer egin herentzi arazo hori konpontzeko, baina ziurtatzen da, behintzat, iturburu-hiztegitiko informaziorik ez dela bidean galtzen.

Hasierako fase honetan egin den aberasketaren aldetik ezer azpimarratzekotan, gure ustez, ezarritako *herentzi mekanismoek* eskainitako aukerak aipatu beharko lirateke. Herentzia hori ezarri ahal izan bada, arestion aipatu dugun eraikuntza taxonomiko horri esker izan da.

Herentziaren kudeaketarako era egoki bat eskaintzen dugula esan dezakegu: kontua da erlazio lexiko-semantiko bakoitzari dagokion herentzi rola ongi zehaztea. Herentzi rolaren zehazte hori errepresentazio maila erlazionalean egin behar da: maila definitorioko erlazio eta propietateek ez dute parterik herentzian<sup>2</sup>. Erabilitako softwareak hainbat herentzi rol desberdin eskaintzen du —*INHERITANCE* fazetan zehaztu behar da rol hori, atributuak sortzean—, eta gainera, aukera ematen du rol berriak definitu eta inplementatzeko. Interesgarrien artean, hona batzuk:

- *OVERRIDE.VALUES* herentzi rol tipikoa, bertako balioei lehentasuna ematen diena heredatuekiko (unitate "ume" batek atributu batean balio lokalik edo bertakorik ez badu orduan bere "gurasoena" heredatuko du); hori litzateke, adibidez, ACTIONS/EVENEMENTSen definitutako OBJECTIF bezalako kasu-erlazio baten herentzi rola. Rol honen bitartez besterik ezeko herentzi era bat modeliza daiteke, balio lokalek, esistituz gero, heredatuekiko lehentasuna hartzen dutelarik.
- Herentzia galarazteko aukera ematen duena: *UNIQUE.VALUES*. Rol hori duten erlazioetan gurasoetatik umeetarako balioen herentzia galarazi egiten da. Horixe litzateke kontzeptu baten sinonimoak dauzkan atributu erlazionalaren herentzi rol egokia (SYNONYMES).

---

<sup>1</sup> HEBaren eraikuntza-programaren errore bat galaraziz honela, ez da eta zilegi horrelako lotura bat ezartzean bigiztarik sortzea.

<sup>2</sup> Atributuok STRUCTURESen sortu dira *UNIQUE.VALUES* herentzi rolaekin: herentzi rol horrek balioen azpiklaseetarako herentzia galarazten du.

### III. kapitulua

- Balio lokalak eta heredatuak bateratzeko aukera ematen duena: *UNION*. Rol hori dagoenean atributuaren balio eratorria<sup>1</sup> bertako balioen eta balio heredatuaren bildura da. Horixe litzateke rol egokia entitate baten ezaugarriak biltzen dituen *CARACTERISTIQUE* bezalako atributu batentzat.

Horrela bada, herentziaren gestioa *STRUCTURE*Seko kontzeptu-klase desberdinetan definitutako atributu erlazionaletan herentzi modu edo rol horiek egokiro zehaztean datza.

Sortutako prototipoan erlazioen herentzia zehazteko inguruneak eskaintako rolez baliatu gara (ikus txostenaren D eranskina). Hala ere, aitortu behar da oraindik ez ditugula ebaluatu atributu desberdinetan herentzi rol horiek definitu izanaren ondorioak.

#### III.2.2 Bigarren urratsa *HEB*aren aberasketan.

Azpiatal honetan berriz, *HEB*aren prototipoa sortu eta gero egindako aberaste-prozesuak azaldu nahi genituzke. Batch eran egin diren —eta urrats honetan egiteko interesgarritzat jotako— prozesu horien xedea, *HEB*an esplizituki errepresentatutako ezagutza hedatzea izan da.

*HEB*ko ezagutza aberastu nahi horretan, beti ere galdera bera datorkigu burura: ezagutza berri hori a priori eskuratzea komeni ote da, hots, prozesu deduktibo bat martxan jarri eta lortutako emaitzak *HEB*an gordez, ala aitzitik, ez ote da hobe arrazoiketa-mekanismo egokiak ezarri eta behar denean bakarrik jartzea martxan, erabiltzailearen eskaripean esate baterako? Galdera horri emandako erantzunaren menpe dago, hain zuzen ere, arrazoitze horiek era desberdinetan inplementatu ahal izatea: inferentzi erregelak, unitateei lotutako metodoak, iratxoak, eta beste.

Erabakia nolabait "salomonikoa" izan dela esan behar: prozesu batzuk alde aurretik exekututzea erabaki dugun bitartean, beste batzuk premien arabera martxan jarri ahal izateko moduan ezarri ditugu. Azpiatal honetan aurrenekoei buruz besterik ez gara mintzatuko, hau da, *HEB*ak hasierako eraikuntzaren ondoren zeukan ezagutza aberasteari begira batch-ean egindako prozesuez alegia. Ondoko alderdiok ukitu dituzte, bereziki, prozesu horiek:

- Sinonimiaren propietateen ustiapena.
- Sinonimian oinarrituriko egitura taxonomikoaren hedakuntza.
- Desanbiguatze lexikala.

---

<sup>1</sup> Balio eratorria esaten denean, atributu baten balioei gagozkiela, balio guztien multzoa esan nahi da: balio multzo horretan, bertako balioez gain, atributuaren herentzi rolaren arabera heredatuak ere —heredaturik baldin badago, noski— egongo dira.



- Desanbiguate sintaktikoa.

Erlazio lexiko-semanticoen propietateak ustiatzea —erlazio taxonomiko eta sinonimikoenak, bereziki— interesgarri izan bide liteke HEBaren aberaste dinamiko bati begira. Meronimiari dagozkion propietateak, edota berorren harremanak erlazio taxonomikoekiko —konposaketa lexikala—, besteak beste, oraingoz alde batera utzi ditugu erabiltzailearen eskaripean egingo den arrazoiketan edo ustiatzekotan.

### **III.2.2.1 Sinonimiaren propietateen ustiapena.**

Erlazio lexikal bereziki inportantetzat jotzen dugu *sinonimia*, testuinguru desberdinetan ustiagarri gerta daitekeenez: dedukzio-prozesuak, erabiltzailearekiko interfazea (Marchionini, 89), eta abar. (Wilks *et al.*, 89:209)n, K. Sparck-Jones-en lanari buruz (1986)<sup>1</sup> hau diote: “*Of twelve possible semantic relations, synonymy was chosen as the fundamental feature of natural language.*”. Kapitulu honetako III.1.2 pasartean ere aipatu dugu hitzen zentzuak errepresentatzeko sinonimian oinarritzen den lanik (Miller, 85; Miller *et al.*, 88).

Prototiporako egin duguna, sinonimi erlazioa sistematikoki ustiatzea izan da, simetria eta trantsitibotasun propietateak dituela suposatuz<sup>2</sup>. Horrela bada, kontzeptu orok duen SYNONYMES atributuak kontzeptu horren sinonimo guztiak edukiko ditu, iturburu-hiztegian esplizituki emandakoak zein sinonimiaren propietateen ustiatze honen ondorioz deduzitutakoak (hala ere, DEF-SYNONYME atributu definitorioan izango ditugu beti definizio sinonimikoetan esplizituki emandako sinonimoak).

Propietate horien inplementazioa dela eta, lehen ere esana dugu simetria hasierako eraikuntza burutzean ustiatua zela, alderantzizko erlazioen ezarpen automatikoari esker: izan ere, SYNONYMESen alderantzizkoa SYNONYMES bera da. Trantsitibotasunari dagokionean berriz, erregela bidezko arrazoiketa erabili dugu. Erregela bat berriaz diseinatu eta martxan jarri dugu aurreranzko kateamenduan, THESAURUSeko kontzeptu guztien SYNONYMES atributua aztertu eta, hasierako eraikuntzan esplizitu egindako egitateetan oin hartuz, propietate horren bitartez deduzi daitezkeen lotura guztiak ezarri dituelarik<sup>3</sup>. Azpimarratu

---

<sup>1</sup> Berorren tesi-txostenaren (1964) berrargitalpena.

<sup>2</sup> Sinonimiaren propietateok, eta batez ere trantsitibotasuna, horrela ustiatzeak ekar litzakeen arazoez ederki ohartzera (Warnesson & Marcotorchino, 82; Calzolari, 83; Warnesson, 84; Sparck-Jones, 86; Chodorow *et al.*, 88). Hala eta guztiz ere, oraingoz horrela jokatzeko erabakia hartu dugu, problemaren benetako nondik norakoak geroago aztertuko ditugulakoan.

<sup>3</sup> Gorago esan dugun bezalaxe —alderantzizko erlazioen inplementazioaren gainean—, propietate hauek inplementatzeko ere nahiago izango genuen dedukzio-erregelak erabili izana (erregela arrunten ordeztuz), deduzitutako egitateak sistemak automatikoki kudeaturiko justifikazio batzuei loturik gordez. Tamalez, horrelako erregelak klase-sloten bitartez adierazitako egitateekin erabiltzea ezinezkoa da—KEEren oraingo bertsioan, behintzat—, eta ezin izan dugu horrela egin; hala ere, bi erregela mota horien sintaxia hain da antzekoa non balizko aldaketa bat ez bailitzateke batere kostako.

### III. kapitulua

nahi genuke nola, "gehegizko dedukzioak" galaraztearren, "kateamendua eteteko" erabakia hartu dugun anbiguetateren bat zuten kontzeptuak aurkitutakoan (CONCEPTS-AMBIGUSeko azpiklaseren batekoak), izan ere kasu horietan ezinezkoa baita jakitea zein adiera konkretuekin jarraitu. Erregela hau REGLES-CONSTRUCTION-SYNONYMIQUE klaseko SYN-TRANS da (ikus D eranskina).

Erregela bidezko arrazoiketaren aspektu interesgarri bat KEEn, arrazoiketa aurrera joan ahala erabiltzailearekin elkarreragiteko erraztasuna litzateke; elkarrekintza hori horretarako bereziki duen mekanismo bati esker gauzatzen da. Mekanismo hori dela medio posible da, adibidez, erabiltzailearen parte hartze zuzena premisak ebaluatu edota ondorioak baieztatukoan, erabiltzaileak arrazoitze-prozesua aldarazteko benetako aukera duelarik. Adibide gisa, esan dezagun ASK.USER izeneko funtzioa simetriaren trantsitibotasuna adierazten duen erregelaren ondorioan ipiniz gero erabiltzaile adituak —lexikografoa dugu gogoan, inondik ere— aukera izango lukeela, interaktiboki eta oso modu erosoan, lotura sinonimiko egiazki okerrak ezarri baino lehen galarazteko.

#### III.2.2.2 Egitura taxonomikoaren hedakuntza.

Aberaste-fase honetako bigarren aspektu bat, arestian azaldutakoarekin zerikusi zuzena duena gainera, *sinonimian oinarritutako egitura taxonomikoaren hedakuntzari* dagokiona da. Kontua oso sinplea da, ondoko bi egitateok esplizitu egitea bezain sinplea: (1) C kontzeptu baten sinonimoen genusa C-ren genusa ere bada, eta (2) C kontzeptuaren genusaren sinonimoak ere C-ren genusak dira. Horrela, sinonimiak emandako informazioari esker (deduzitua nahiz jatorrizkoa), hasierako eraikuntzan lortutako egitura taxonomikoa hedatu egiten da.

Egitura taxonomikoaren aberasketa hau egiteko bi erregela erabili ditugu, aurreranzko kateamenduan martxan jarritz. Bi erregela horiek EXTENSION-CONSTRUCTION-TAXONOMIQUE erregela-klaseko elementu dira, STRUCTURES azpi-EBan (ikus D eranskina).

#### III.2.2.3 Desanbiguaketa HEBan.

Baina HEBko ezagutza aberastearen aldetik garrantzirik handieneko aspektua, agian, hiztegi-ezagutza horrela errepresentatzeak berak eragin lezakeen desanbiguaketa litzateke. Izan ere, HEBan egin daitezkeen dedukzio-lanetarik nabarienak *bere anbiguetate maila txikiagotzera zuzendutakoak* lirateke. Fase honetan erakutsi nahi genuke nola balia gaitzkeen HEBaren prototipoan esplizitu eginiko ezagutza lexiko-semantikoaz

desanbiguatze horretarako. Bestalde, azpimarratzekoa da anbiguetate horren iraupena benetako arazoa dela HIZTSUAREN dedukzio-mekanismoen funtzionamendurako.

Anbigutasun hori, hiztegi-definizioetatik eta beroriekin egindako analitiko datorrena, bi motakoa da funtsean: anbigutasun lexikala eta anbigutasun sintaktikoa. Hortaz, bi desanbiguatze-prozesu sail ditugu gogoan: desanbiguatze lexikalekoak eta desanbiguatze sintaktikokoak.

Hurrengo azpiataletan, desanbiguketa lexikal eta sintaktikoari buruzko literaturari ikustalditxo bat eman ondoren, labur-labur azalduko dugu arazo honi buruz zer azterbide hartu ditugun gurean eta zer-nolako prozesuak ezarri asmo ditugun HEBan.

### III.2.2.3.1 Desanbiguketa lexikala.

Testuinguru jakin batean hitz lexikalki anbiguo baten erabilerari zein adiera dagokion zehazki jakitean datza desanbiguketa lexikala; hots, hitzak dituen zentzu desberdinen artean, hitzaren erabilera jakin horri dagokiona hautatzean. Ohizko hiztegi elebarkarretan, anbiguetate lexikala hitzen definizio-testuetan ageri da. Hiztegiko sarrerak homografo, adiera edo zentzu, zentzu-ñabardura eta abarretan sailkatuak egon arren, lexikografoak ez du horrelako zentzu-erreferentzia zehatzik erabiltzen adiera bati dagokion definizioa idaztean, ezpada hitzak erabiltzen ditu. Hitzok polisemikoak izanik, anbigutasun lexikala nabarmena da hiztegian. Lexikografoa ahalegintzen da, batzuetan, definizioan erabili duen hitzaren zentzua zehazten, baina behin ere ez du egiten era esplizitu batez —Mel'cûk (1984) eta bere taldekoen hiztegi esplikatiboak salbuespena lirateke honetan—, hitzari dagokion adieraztenbakia emanez esate baterako, baizik eta, "informatikoki" ikusita behintzat, askoz ere ilunagoa den bideren bat erabiliz, hala nola sinonimo baten erreferentzia eginez, modu perifrastiko batez, eta abar.

Hori, bistan da, arazo bat da THESAURUS bezalako sare bat eratu nahi izanez gero, non oinarritzko unitatea zentzua den eta ez hitza. Eta arazoa hortxe dago definizioetan erreferentziaturiko hitzei dagozkien unitateak sortzerakoan, eta baita unitateok elkarrekin edota definitutako hitzarekin erlazionatu behar direnean, STRUCTURESEKO klase desberdinetan zehaztutako atributuen bitartez. Gure kasuan, HDBan egindako desanbiguatze lexikaleko prozesu guztiak gorabehera (ikus datu eta oharra txosten honetako § II.4.2 eta II.7.2n), kontzeptu-unitateen %30ak oraindik anbiguo dira THESAURUSen (konfigurazioak kontatzeke) hasierako eraikuntza burutu eta gero. Kontzeptu horiek erreferentzia anbiguoak dira, REFERENCES klaseko (klase horretako %70a inguru) eta CONCEPTS-AMBIGUSEKO azpiklaseren bateko elementu.

Desanbiguketa lexikalaren arazo honi buruzko ideiarik funtsezkoena da testuinguruak desanbiguatzen laguntzen duela (Sinclair, 87), hau da, adiera anitzeko hitz (polisemiko)

### III. kapitulua

baten aurrean gaudelarik jakin dezakegula, hitzaren testuingurua aztertuz, zein den zuzena. Definizio-hiztegi baten kasuan, jakin behar dena da ea testuinguruak —definizio-testuak, alegia— aski zabalak diren desanbiguatze horietarako. B.M. Slator, bere lanean (1989), testuinguruari esker desanbigua daitekeelako ideia indartzera dator, desanbiguaketa lexikalaren problema bere lekuan jarritz:

The lexical ambiguity problem is not so acute in traditional natural language systems with lexicons of a few hundred words; but the difficulties rise dramatically in the face of the huge lexicons proposed by the latest generation of computational lexicographers. These large systems, with multiple senses for a large portion of the words, need a motivated strategy for word sense preference —preferably one that is context dependent and knowledge based.

K. Dahlgren-ek (1988) desanbiguatze lexikalerako hiru informazio mota biltzen dituen metodoa proposatzen du: esamolde ihartu eta usuak —agerkidetza lexikala—, informazio sintaktikoa eta zentzu oneko arrazoiketa. Lehenengo informazio mota desanbiguakizun den testuaren analisisian oinarritzen den bitartean, eta ez da hori oraingoz HEBrako nahi genukeen desanbiguatze mota<sup>1</sup>, informazio sintaktikoaren bidezko desanbiguaketak edo zentzu oneko arrazoiketan oinarritutakoak interesgarri dirudite. Azken informazio mota hori dela eta, autoreak proposatzen du hitz anbiguo bi edo gehiagoren arteko "antzekotasun ontologikozko" testak (taxonomiako nodo beraren azpian egotea) egitea.

(Byrd *et al.*, 87)n Lesk-en lana (1986) aipatzen dute, honek hiperonimoen desanbiguaketa egiteko proposatzen duen metodoa differentiako hitzak hiperonimoaren adiera desberdinen definizioetakoekin konparatzea delarik: differentiarekin gehien parekatu definizioari dagokion adiera izango da hiperonimoaren adiera zuzentzat aukeratutakoa.

(Hirst, 87)n berriz, R. Amsler-en lanei buruz (1980; 81) diharduela, autoreak dio definizio-testuen egitura erabil daitekeela desanbiguatze lexikalerako, ISA eta PART-OF erlazioei jarraituz.

P.S. Jacobs-ek (1989) aldarrikatzen du hitz-zentzuen determinazioa “. . . *a knowledge-intensive problem*” dela, eta lan horretan erabil litezkeen informazio-iturri desberdinen lista ematen du. Horien artean morfologia aipatzen du —eratorpenari dagokiona, bereziki—, hitzen maiztasuna, berbaldiaren mintzagaia, hiztegian hitzek dituzten adierak, agerkidetza, ebakidurak —koherentzia lexikala—, lehenetasun semantikoak eta informazio sintaktikoa.

(Wilks *et al.*, 89:198)n ondoko hau irakur daiteke:

. . . any attempt to extract semantic information from a machine dictionary must acknowledge that the words used in dictionary entries of words and their senses may themselves be lexically

---

<sup>1</sup> Informazio mota hori interesgarriago litzateke HDBan gordetako definizio-testuetan desanbiguaketa burutzeko.

ambiguous and must be disambiguated. When human readers see ambiguous words in the definitions of real dictionaries, they appear to recognise those words as used in a particular sense, understand the intended senses of the words, and hence disambiguate the words in the dictionary definitions.

Ondoren, autore horiek problema honetarako hiru ebazpide desberdin proposatzen dituzte:

- Giza jokamoldea imitatuz hiztegiko definizioak desanbiguatuko dituen programa bat jarri martxan.
- Hiztegiko definizioetako anbiguetate lexikal oro kendu, zentzuez besterik ez osaturiko definizioak lortuz, (Amsler, 80)ren antzera.
- Ez egin kasurik arazoari harik eta bere tratamendua benetan premiazkoa bilakatu arte.

Beren lanetan, Y. Wilks-ek eta (1989; 90) desanbiguaketa lexikala ikaste-prozesutzat hartzen dute —gurean aberasteaz hitz egiten dugun bezala, apika—; ikaste horretan baina, kontuz ibili behar da superjeneralizazio eta super-espezifikazio gehiegizkoak ez egiteko. Desanbiguaketa lanetarako, autore horiek erabiltzen dituzte agerkidetza lexikalari buruzko neurri estatistikoak eta erabilera edo hitzen eremuari buruzko kodeen (LDOCETikoak) bitartez eratutako hierarkia pragmatikoak.

Krovetz-ek (1989) berriz, hitz baten zentzu desberdinei buruzko hiztegi-informazioa —kategoria sintaktikoa, morfologia, murrizapen semantikoak, etc.— hitz horren agerpen jakin bati zein zentzu dagokion zehazterakoan erabili beharreko informazio-iturri aniztzat hartzea aldarrikatzen du.

(Papegaaij *et al.*, 86)n, autoreek "ezagutza lexikaleko banku" bat darabilte —taxonomia dimentsioanitza, informazio logiko, ontologiko eta testuinguruzko menpekotasunari buruzkoa daukana—, desanbiguaketa lexikala helburu duen itzulpen automatikoko sistema baten lexiko gisa.

(Byrd *et al.*, 87) lexikologia konputazionalerako tresna eta metodoei buruzko lanean ere, egileek heltzen diote hiztegi bateko hitz-zentzuen bereizketaren arazoari: proposatutako metodoa, oraingoan, sinonimian oinarritua da, sinonimia hitz-zentzuen arteko erlazio simetriko bezala hartuz. Beren metodoa, hiztegi-sarrera batzuen definiziotzat emandako sinonimo-zerrendetan dauden asimetriak bilatzean datza, eta zerrenda horien arteko ebakidurak kalkulatzeko. (Chodorow *et al.*, 88)n zehatzago azaltzen dituzte metodo hauek, eta (Byrd, 90)ek COLLINS thesaurusean desanbiguaketa lexikala burutzeko egindako aplikazioei buruzko hainbat datu ematen du. Simetriren metodo horri esker mota horretako erreferentzia gurutzatuen %37a desanbiguatzea lortu dute, sinonimo-zerrenden arteko

### III. kapitulua

ebakiduren metodoaz erreferentzien %80a iritsiz (horietan barne lehen metodoaz lortutako desanbiguatze gehienak).

A. Copestake-k (1990) dioenez, desanbiguatze lexikala burutzea ezinbestekoa da taxonomia koherenteak eratu nahi baldin badira. Desanbiguaketa hori egin ahal izateko, hitzen hiperonimoei dagokienean, informazio desberdinetan oinarritutako zenbait heuristiko aipatzen du: LDOCEko kode semantikoen bateragarritasuna, hitzen erabilera maila (argoteko hitzak ez bide lirateke hiperonimo), adiera-zenbakia (txikiena izaki hiperonimo izateko probabilitaterik handienekoa), eta definizioetako hitzen forma flexionatuen arteko "overlap" mailaren determinazioa.

Vossen eta Serail-ek (1990) "desanbiguatze partzialaz" dihardute beren txostenean — gurean CONCEPTS-AMBIGUSen azpiko ACCEPTION klaseko instantzia diren erreferentziek dutena gutxi gorabehera, hots, beren homografo-adierazlea ezaguna da baina ez dakigu zein adierari dagozkion zehazki—. Halaber, guk COMPLEXES klaseko elementu egin ditugun kontzeptuei buruz ere hitz egiten dute, hau da, adiera-zenbaki bi edo gehiago dituzten kontzeptuak, beren desanbiguaketa oraindik osatzeke dagoela edo adieraziz<sup>1</sup>. Desanbiguaketa lexikala beharrezkotzat jotzen badute ere, autore hauek idarokitzen dute arazoa maiz erabili eta oso orokor diren hitz gutxi batzuen baina ez dela, eta hitz gutxi horiek eskuz desanbigua daitezkeela. Beren estrategiak beraz, maizeneko hitz horien eskuzko desanbiguaketa du funtsa, ondoren gainerantzekoak "automatikoki" desanbiguatu ahal izateko; azken desanbiguaketa hori anbiguo geratutako erreferentziei lehen adiera-zenbakia esleitzea besterik ez da (gainera, erreferentzia horien gehiengoak anbiguetaterik ez du izango, monosemikoak izango direnez).

HEBaren aberasketa-fase honetan desanbiguaketa lexikalaren aldetik egitea pentsatu duguna, aurreko pasarteetan azaldutakoen moduko heuristikoak ezartzea izan da. Heuristiko horiek aurreranzko kateamenduan funtzionatuko duten erregelen bidez inplementatu eta batch-ean exekutaraziko dira. Sinonimiaren propietateen ustiatzea dela eta lexikografoarekiko elkarrekintzari buruz esan dugunaren antzera, hemen ere halako elkarrekintza —dedukzio-prozesua aurrera joan ahala— eskaintzeko asmoa genuke, emaitzen fidagarritasuna segurtatzearen.

HEBko THESAURUSen sortutako kontzeptu anbiguoek behin-behineko status bat dutela esan dezakegu. Izan ere, hor egongo dira harik eta desanbiguaketa-prozesuren batek

---

<sup>1</sup> Askotan ordea, adiera anitzen erreferentziak dituzten unitate konplexu hauen desanbiguaketa ez da posible izango eta bere horretan hartu beharko dira dedukzio-prozesuetan. Unitate konplexu hauek (Vossen & Serail, 90)eko "zentzu anitzeko adierazpenei" dagozkie.

"ezabatzea" lortu arte<sup>1</sup>. CONCEPTS-AMBIGUSeko elementuentzat definitutako metodoak —objetuei orientatutako programazioaren zentzuan— izango dira unitate horien ezabaketa burutuko dutenak beren *erabateko desanbiguaketa* lortzen denean, desanbiguatze hori unitate anbiguoak THESAURUSeko beste unitateekin dituen lotura guztien desagertze bezala ulerturik. Alabaina, desanbiguaketa lexikaleko prozesu desberdinek unitate anbiguoon *emendiozko desanbiguatzeak* burutuko dituzte, zeren eta, desanbiguaketa bat posible den bakoitzean, kontzeptu anbiguoranzko lotura bat dagokion adierarantz birzuzenduko baita —desanbiguaketaren ondorioz, hain zuzen—, horrela unitate anbiguoak lokailu bat (alderantzizkoa) galduz. Unitateak lokailu guztiak galdu dituelarik ez du zentzurik THESAURUSen mantentzeak.

Konkretuki, desanbiguaketa lexikaleko ondoko erregelak inplementatu ditugu, DESAMBIGUISATION-LEXICALE erregela-klaseko elementu eginez (ikus D eranskina):

- Definizio sinonimikoetan agertutako erreferentzien desanbiguaketa: sinonimiaren trantsitibotasunak "sinonimo klaseak" eraikitzen ditu, non, erreferentzia anbiguo bat hitz beraren erreferentzia desanbiguatu batekin batera gertatzeak, lehenengoa ziurtasun handiz desanbiguatzeko aukera ematen baitu (ikus RDL1 erregela D eranskinean).
- Ezagutza taxonomikoan oinarritutako erreferentzi desanbiguaketa: erlazioek THESAURUSeko kontzeptuekiko dituzten loturez baliatuz (*CONCEPT-TYPE-CORRESPONDANT* fazeta), erlazio horien balio diren erreferentzia anbiguoak desanbigu daitezke (ikus RDL2 erregela D eranskinean).

### III.2.2.3.2 Desanbiguaketa sintaktikoa.

Desanbiguatze sintaktikoa, guretzat, atributu sintagmatikoetan geratutako balioak dagozkien atributu erlazonaletara goratzean datza, balio horiek ezin baititugu, printzipioz behintzat, zeregin deduktiboetan erabili (atributu horien azpian dagoen "erlazio semantiko sakonekoa" ez dugu eta ezagutzen).

Ikus dezagun, adibidez, erlazio semantiko anitzen isla izan daitezkeen preposizioen kasua: *à* preposizioa, frantsesez, xedezko, datibozko, edutezko, etc. kasuen azaleko gauzatzea izan daiteke. Beraz, LPPLko definizio-testuetako preposizio horren hainbat

---

<sup>1</sup> THESAURUSen |CI?| kontzeptua egoteak, C kontzeptuaren adiera guztien multzoa (lehen homografoa, noski) adierazten du, nolabait: unitate honetara iristen diren lokailuak C-ren adiera desberdinetara zuzendu beharko liriateke erabateko desanbiguaketa lexikala burututakoan. Hori esan nahi dugu, hain zuzen, *unitateon ezabaketa* aipatzen dugunean, ezabaketa hori desanbiguaketa burututakoan egingo bailitzateke, hau da, unitateak dituen lotura guztiak desagertutakoan. COMPLEXESeko elementuen kasua ere antzekoa da: |CI 3/4| kontzeptua egoteak esan nahi du bertara iristen diren lokailuak birzuzendu egin beharko direla, desanbiguaketa lexikala egitean, |CI 3| kontzeptura edo |CI 4|era.

### III. kapitulua

agerpen errepresentatzeko A atributu sintagmatikoaz baliatu behar izan dugu, non eta bere desanbiguatzerako adierazpen zehatzik ez zegoen ESEEn —eta, hartara, balioaren goratzea dagokion kasu-atributura egina zen HEBaren eraikuntza egin denean—. Desanbiguaketa bada, goratze hori posible egitean datza, maila sintaktikoan dirauten erlazioak sakoneko erlazio semantikoak bihurtzarren.

Eta guri interesatzen zaigu, batik bat, ikustea ea THESAURUSen esplizitu egindako erlazio taxonomiko horiek (erlazio sinonimikoez ere baliatuz) erabilgarri diren erlazio sintaktiko horietako zenbait desanbiguatzeko.

Desanbiguaketa sintaktikoa dela eta, K.C. Litkowsky-k (1978, 80) definizioetan oinarritutako preposizioen desanbiguaketa du aztergai. 1978ko lanean, autore honek definizio-patroiak —"recognition rules"— erabiltzea proposatzen du, horiei esker "with", "by" eta "into" bezalako preposizioak, aldeztatik determinatu aditz batzuekin agertzen direnean, desanbiguatu eta sakonean gordetzen duten erlazio semantikoa (tresna, bitartekoa, etc.) zein den jakin daitekeela erakutsiz. 1980koan berriz, errepresentazio semantikoan slot gisa azalduko diren kasuak ezagutzearren preposizioen hiztegi-definizioak erabiltzea proposatzen du. Halaber, definizioez balia gaitzkeela slot horien betegarri zer izan daitekeen jakiteko: hautapen-murritzapenak.

J.L. Binot eta K. Jensen-ek (1987; Jensen & Binot, 87; 88) hiztegi-definizioak desanbiguaketa sintaktikorako baliagarri izan daitezkeela erakusten dute; dena den, kasu horietan helburua ingelesezko perpausetan preposizio-sintagmak zein sintagmari dagozkion —ziurtasun osoz ez bada ahalik eta ziurren— determinatzea da. Horretarako, autoreek arrazoiketa modu heuristikoa bat darabilte, esleipen-posibilitate desberdinen probabilitateen konparazioan oinarritua (esaldiak eta esaldiko osagaien hiztegi-definizioak analizatzaile berberarekin analizatu eta gero). Arrazoiketaren funtsa, hiztegien giza erabilera nolabait islatu nahi luketen erregela multzo batean legoke. Oinarrizko estrategiaren xedea, desanbiguakizun diren preposizioek elkarloturiko terminoen artean "konexio semantikoak" bilatzea da. Arrazoitze honen emaitzak adierazteko esleipen posible bakoitzari egokitutako ziurtasun-balioak erabiltzen dira, guk geurean fazeta erlazionalen bitartez egiten dugunaren antzera.

(Calzolari & Picchi, 88)n autoreek proposatzen dituzten desanbiguaketa sintaktikoko erregelak hiperonimo batzuen eta erlazio batzuen arteko agerkidetzak bilatzen saiatzen dira; hori dela medio zehatz daiteke, erlazio horiek zein testuingurutan gerta daitezkeen probabilitate handiago batekin, edota elementu anbiguo batzuk (preposizioak, adibidez) non desanbigua daitezkeen sakoneko erlazio bakar baten gauzatze bezala. Erregela multzo horiek unitate jakin batzuei egokitzen zaizkie —hiperonimo, erlazio jakin batzuk, etc.—, beren



aldameneko testuingurua desanbigua dezaten; orobat, zenbait unitateri lot dakizkioke beren testuinguruetan interpretazioak ad-hoc egiteko gai izango diren prozedura bereziak.

(Wilks *et al.*, 90)en azpimarratzen dute hiztegi-definizioen izaera intentsionala adierazgarri izan daitekeela desanbiguatze-lanetarako. Horren argibidetzat ematen duten kasuan (145. orr.) azaltzen dute nola, "ammeter"en definizioa "an instrument for measuring ... electric current" izanik, hitz horren definizioa errepresentatzen duen framean PURPOSE slota beteko litzatekeen "measuring" balioarekin. Informazio hori berehala erabil daiteke kasu-rolen "betegarrien" lehentasunak kalkulatzeko. Konparazio batera, "The technician measures alternating current with an ammeter" esaldiko "with" preposizioa tresnazko kasu-sartzaile gisa desanbiguatzea hobetsiko da beste aukera posibleekiko —komitatibo edo edutezkoa—, izan ere "ammeter" kontzeptuaren intentsioaren errepresentazioko PURPOSE slota "measuring" balioarekin egoteak esaldiko "with" horri tresnazko zentzua ematera bultzatzen baitu.

Desanbiguatze sintaktikoko prozesu horiek burutzeko gure proiektuan egin asmo ditugunak, aldezturik lortutako ezagutza taxonomikoaren ustiapenean oinarrituko liriteke eta aurreranzko erregelen bitartez inplementatuko genituzke.

Kontua da atributu sintagmatikoen balioak aztertu eta ea goratzeko modukoak diren ikustea<sup>1</sup>. Atributu erlazional jakin baterako goratze hori zilegiztatzeko, balio bakoitzaren hiperonimoak aztertu beharko dira eta ikusi ea atributu erlazionalarekiko bateragarriak diren: esate baterako, tresna baten izena izan daiteke INSTRUMENTAL atributuaren balio, gai batena MATIERE atributuarena eta egintza baten izenak bete dezake OBJECTIF bezalako atributu bat. Balioen goragarritasunari buruzko ezagutza hori adierazteko bi bide eskainiko ditugu:

- Batetik *CONCEPT-TYPE-CORRESPONDANT* fazeta (ikus § III.1.3.2.2), erlazioen eta mota-kontzeptuen arteko loturak ezartzeko erabilia (adibidez, INSTRUMENTAL atributua |instrument I 3| kontzeptuarekin eta ETAT |état I 1|ekin). Kasu hauetan, desanbiguaketa egiten duen erregelak, atributu erlazionalekin lotutako kontzeptuen eta goragarri den balioaren hiperonimoen arteko bateragarritasuna egiaztatu beharko du.
- Esplizituki desanbiguaketa-erregelan bertan.

HEBaren aberaste-fase honetan desanbiguaketa sintaktikorako erregelak jarriko dira martxan (ikus DESAMBIGUISATION-SYNTAXIQUE erregela-klasea, D eranskinean) mota horretako desanbiguaketak egitearren. Aztertutako balioa atributu erlazional batera

---

<sup>1</sup> Gogoratu atributu sintagmatiko orok duela, bere *RELATIONNELS-CORRESPONDANTS* fazetan, balioa zein atributu erlazionaletara gora daitekeen jakiteko informaziorik.

### III. kapitulua

goratua izan daitekeela egiaztatu eta gero<sup>1</sup>, erregelak egin behar duen gauza bakarria atributuaren izenkide den fazeta erlazionalean ziurtasun-balio handiago bat ipintzea da. Fazeta-balio hori behar adinakoa baldin bada, 0.5 baino handiagoa adibidez, balioaren goratzea atributuari lotutako iratxoaren (SYNTAGMATIQUE-RELATIONNEL) zeregina izango da. Esan gabe doa balioa atributu erlazional batean gordetzen den momentuan, atributu horren erlazioak definiendumari pasarazten dizkion iratxoa —DEFINIENS-DEFINIENDUM, definiens baten adierazpide den konfigurazioaren kasuan— edota alderantzizko erlazioen mantenimendua bermatzen duena automatikoki exekutaraziko direla erlazio berriak ezarritz.

### III.3 Alderdi inferentzialaren ikuspegia.

HEBrako proposatzen dugun tresneria inferentzialaren ikuspegi orokorra eman nahi genuke atal honetan. Hemen azalduko ditugu asmatutako dedukzio-mekanismoak, eta batik bat, prototipoan dagoeneko inplementatu ditugunak. Mekanismo guzti horiek, HIZTSUAren ahalmen deduktiboaren gauzatze diren neurrian, sistemaren Dedukzio-Moduluaren parte dira logikoki (ikus sistemaren arkitektura, § I.4.1.1en), nahiz eta inplementazioari begira batzuk STRUCTURES azpi-EBan ezarri ditugun.

Orain deskribatu nahi duguna da erabiltzailearen eskari bati erantzuterakoan martxan jartzen diren dedukzio-mekanismoen multzoa; mekanismo horien helburua, oro har, HEBan esplizituki errepresentaturik ez dagoen ezagutza eskuratzea da. Sistemak erabiltzailearen galdera bati erantzun behar dionean, dedukzio-prozedura sail bat jartzen da martxan beharrezko ezagutza hori lortzeko helburuarekin, dela HEBko ezagutza esplizitua, dela deduzitu beharrekoa. Ezagutza-eskuratze prozesu dinamiko bat abiarazten dela orduan esan daiteke.

Arestian azaldu ditugun prozesuen eta mekanismo hauen arteko diferentziarik nagusia zera da, alegia lehengo horien xedea ezagutza inplizitua HEBan esplizitu bihurtzea dela —ez interaktiboki ezpada batch-ean—, eta beste hauena aldiz, erabiltzailearekiko elkarrekintza gauzatzen deneko *ezagutza-dedukzio dinamikoa* litzateke, ezagutza deduzitu hori HEBan zertan gorde beharrik ez dagoelarik<sup>2</sup>.

Horrela, mekanismo sail honi HIZTSUAren oinarritzko funtzioek duten dedukzio-ahalmena gehitzen badiogu (ikus IV. kapitulua), HEBaren *azpiegitura deduktiboa* izango

---

<sup>1</sup> Posible litzateke, beti ere, lexikografoarekiko elkarriketa baten bitartez, egitear dagoen dedukzioaren zuzentasuna egiaztatzea.

<sup>2</sup> Jakina, horrek ez du esan nahi kontrakoa ere: elkarrekintza hori lexikografo batekin gauzatzen denean, adibidez, sistemaren ikaste-prozesu gidatu bat imajina daiteke, non esplizitu bihurtuko bailitzateke —erabiltzaile lexikografo horren baimenaz— ezagutza dinamikoki deduzitu hori.

dugu. Hemen azalduko dugun azpiegitura horren alderdia ezagutza deklaratioaz osaturik dago batik bat; oinarriko funtzioei dagokien ezagutza berriz, prozeduralagoa dela esan dezakegu. Aitortu beharra dago proiektuaren parte hau ez dagoela oso aurreratua eta, beraz, ondoko orriotan laburki azalduko ditugu gure ereduak hiztegi-ezagutzaren errepresentaziorako alde horretatik eskainitako aukerak.

Komentatuko ditugun mekanismoek ondoko gaiok ukitzen dituzte:

- Erlazio lexikalen konposaketa.
- Erlazioen eta THESAURUSeko kontzeptuen arteko loturak.
- Ziurtasun-ezaren tratamendua: HEBan dirauen anbiguetatea.

Erabiltzailearen eskaripean abiaraziko diren prozesu deduktibo hauen inplementaziorako, adimen artifizialeko teknika ezagunez baliatzea pentsatzen dugu, hala nola erregelatan oinarritutako arrazoiketa —aurreranzko zein atzeranzko kateamenduan abiaraz daitekeena—, esleipen prozedurala (iratxoak) eta "mundu" desberdinetango arrazoiketa hipotetikoak. Hori dena HEBaren gainean dago eraikia, non STRUCTURESen definitutako objektu-klaseen portaera prozedurala deskribatzeko objektu horiei lotutako metodoak erabiltzen baitira.

### **III.3.1 Erlazio lexiko-semantikoen konposaketa.**

Sistemaren dedukzio-ahalmenaren alderditik ikusita, erlazio lexiko-semantikoen propietateen ustiapena —erlazio horiek metalengoaia lexikografikotikoak izan ala ez— bide da bere zereginik inportanteenetako bat. Ongi zehaztu behar da zeintzuk diren errepresentatutako erlazio bakoitzaren propietateak, propietate horiek zenbateraino ematen duten biderik hiztegian implizitu dagoen ezagutzarik eskuratzeko, eta ezagutza deduzitu hori noiz interesatzen zaigun eta noiz ez. Erlazio-propietateen artean, erlazio batzuen konbinazio bezala adieraz daitezkeenak hartu ditugu bereziki kontuan.

M.W. Evens-ek argitaratutako liburuaren sarreran (1988), autorea erlazio lexiko-semantikoen propietateei buruz mintzo da: alderantzizko erlazioa, erreflexibotasuna (partea-osa erlazioak lukeena, gauza oro baita bere parte), simetria (sinonimiak), trantsitibotasuna (hiperonimia/hiponimiak) eta "one-to-one" delakoa (zerbaiten-ezkerretara-justu-egon erlazioak lukeena, adibidez). Ondoren dio, propietate guzti horien artean trantsitibotasuna dela —erlazio taxonomiko zein meronimikoena— gehien erabili dena: atributu-herentzia, adimen artifizialeko hainbat sistematan halako garrantzia duen printzipioa, nondik dator bada, ez bada erlazio taxonomikoaren trantsitibotasunetik? Simetria eta antisimetria ere propietate interesgarriak omen dira dedukzioaren ikuspuntutik. Evens-ek bere liburuaren

### III. kapitulua

hurrengo atala eskaintzen dio erlazioek elkarrekin dituzten harremanak aztertzeari, adibide ugari emanaz.

R.J. Byrd-ek (1990) ere aipatzen du konposaketa lexikaleko mekanismoak erabiliz erlazio inplizituak deduzitzen ahal direla: bere kasuan, sinonimian oinarritutako antonimoen dedukzioa da (antonimoen sinonimoak antonimoak dira).

Erlazioon trantsitibotasuna ustiatzeak problema interesgarriak aurkezten ditu (Winston *et al.*, 87), bereziki "trantsitibotasuna" silogismo bateko premisa eta konklusioetan ageri diren erlazio batzuei aplikatzen zaienean<sup>1</sup>; lehen ere aipatua dugun erlazio meronimiko desberdinen trantsitibotasunaren kasua litzateke. Trantsitibotasun mota hau ongi espresatu eta inplementatu beharra dago eta, horregatik, gure sisteman oraintxe azalduko dugun konposaketa lexikalerako mekanismoa asmatu dugu.

Gure sisteman bada, egitate berriak —kontzeptuarteko erlazio inplizituak— konposaketa lexikaleko erregelak aplikatuz deduzitzeko aukera HIZTSUAren azpiegitura deduktiboaren parte bezala inplementatu da. Hemendik aurrera, konposaketa lexikala diogunean, erlazioen propietateak<sup>2</sup> nahiz erlazio lexiko-semantiko desberdinen konposaketa adierazi nahi ditugu.

Gure proposamena honi buruz nahiko sinplea da eta problemari aurre egiteko bi modu eskaintzen ditu:

- *Trantsitibotasun-erregelen sorkuntza dinamikoa.*
- *Konposaketa lexikaleko erregela esplizituak.*

Lehen ebazpidea beraz, *trantsitibotasun-erregelen sorkuntza dinamikoan* datza. Sorkuntza dinamiko hori posible da sortutako erregelak beti forma berdinekoak direlako, trantsitibotasun-erregela baten forma tipikokoak alegia. Trantsitibotasun-erregelon sorkuntza, erregelan parte hartzen duten erlazioak zehazten dituzten hirukoteen erazagupenean oinarritzen da.

Hirukote horiek,  $(R_1 R_2 R_3)$  formakoak, ondoko eskema orokorra duten erregelak sortarazten dituzte:

**If  $X R_1 Y$  and  $Y R_2 Z$  then  $X R_3 Z$**

Hirukote-erazagupenak —bat edo gehiago egin daitezke, atributu bakoitzeko— hirugarren atributu erlazionalaren *COMPOSITION-LEXICALE* fazetan egiten dira, atributu

---

<sup>1</sup> Baldin  $(x R_1 y)$  eta  $(y R_2 z)$  orduan  $(x R_3 z)$ , non  $R_1$ ,  $R_2$  eta  $R_3$ -k ez duten zertan erlazio bera izan.

<sup>2</sup> Simetria, ikusi dugu, alderantzizko erlazioen mekanismoaren bitartez inplementatu dugu (ikus § III.2.1).

horrek errepresentatzen baitu sortuko den trantsitibotasun-erregelaren emaitzari dagokion erlazioa. Arrazoa sinplea da: ( $R_1 R_2 R_3$ ) hirukotea  $R_3$  atributuan erazagutzen da, azken hori baita balio deduzituak har ditzakeen atributua, konposaketa lexikala martxan jarritakoan.

Atributu erlazional guztiek dute AVCOMP-LEX izeneko irakurketa-irratxo bat, atributuaren balioa atzitu nahi den bakoitzean exekutarazten dena. Irratxo horren ardura da, exekutatzen denean, erregela sortzea eta arrazoiketari ekitea. Hortaz, atributua irakurtzeko atzitutakoan, atributuari lotutako irratxo horrek konposaketa lexikaleko erregela bat sortu —lehendik ez badago sortua, noski— eta atzeranzko kateamendua askatuko du, atributurako balio berriak deduzituz (onenean). Balio deduzitu horiek atributuan zeuden balio esplizituei gehituko zaizkie, nahiz eta ez HEBan iraunkorki grabatu. Konposaketa lexikaleko erregela hauek, REGLES-COMPOSITION (ikus D eranskina) klaseko elementu, ondorio bezala lortutako *egitateak mundu berrian baiesten* dituzte, eta horretara erabiltzaileak beti jakin dezake zer alde dagoen HEBko ezagutza iraunkorraren eta erregelak eguneratu mundu edo testuinguru berri horretan baieztatutakoaren artean<sup>1</sup>.

Ikus dezagun, adibide gisa, konposaketa-erregela hauen bitartez nola ebatz daitekeen problema tipiko bat. Erlazio meronimikoen "trantsitibotasunaren"<sup>2</sup> problema da (Cruse, 86; Winston *et al.*, 87), II. kapituluari jada aipatua. Aipatu artikuluetan, egileek diote erlazio meronimikoak trantsitiboak direla silogismoaren bi premisetan ageri den meronimia mota berekoa denean, baina bi meronimia desberdin konbinatzen direnean trantsitibotasun hori ez dela gertatzen. Hiztegi-definizioetatik erauzitako PARTIE-DE erlazioak beti meronimi mota bereko erlazioa denotatzen duela emanik, erazagut dezakegu —ENTITESeko PARTIE-DE atributuan— (PARTIE-DE PARTIE-DE PARTIE-DE) hirukotea, baina ezin erazagutuko dugu, adibidez, (PARTIE-DE MEMBRE-DE PARTIE-DE) hirukote konbinatua.

Erlazio hauek meronimiko ez diren beste batzuekin konposatu ahal izateko —bereziki taxonomikoekiko edota partekotasun espaziala adierazten duten erlazioekiko<sup>3</sup>— erlazio meronimiko eta ez-meronimikoak hierarkia mistoetan konbinatuz egin daiteke (Winston *et al.*, 87). Hierarkion bitartez adieraziko litzateke, hain zuzen ere, erlazio desberdinen

---

<sup>1</sup> Zoritxarrez, erabilitako softwarearen bertsioak (KEE 3.1) ez du uzten "mundu berrian" egiturazko egitaterik (klase/azpiklase erlaziorik) edo klase-sloten bitartez adierazitako egitaterik baiesten. Jakina denez, gure atributu erlazional guztiak klase-sloten bitartez adieraziak dira eta, beraz, ezin izan ditugu konposaketa lexikaleko erregela hauek nahi bezala (hor goian azaldu bezala) inplementatu. Hartaz, prototipoaren egungo bertsioan erregela horien ondorioak HEBan balioak fisikoki eguneratuz gauzatzen dira.

<sup>2</sup> Esan dugun bezala, erlazio bakar baten trantsitibotasunaz baino erlazio meronimikoen bertsio desberdinen konposaketaz dihardute autoreok: "partea-osoa" ez ezik, "osagaia-objetua", "elementua-multzoa", "gaia-objetua", "tokia-ingurunea" eta abar. Kontua da, orduan, determinatzea zein den X eta Z-ren artean dagoen  $R_3$  erlazio meronimikoa,  $R_1$  eta  $R_2$ -ren konbinazioaren emaitza dela jakinik (azken bi erlazio horiek X eta Y, eta Y eta Z-ren artekoak izaki, hurrenez hurren).

<sup>3</sup> "-n egon" espresioaz adieraz genezakeen partekotasun espazialeko erlazioa gure LOCATIF atributuari legokioke.

### III. kapitulua

konposaketatik eratortzen den erlazioaren mota zein den. Guk proposaturiko konposaketa lexikaleko hirukoteak badira, nolabait, hierarkia horiek adierazteko modu bat. Izan ere, (PARTIE-DE LOCATIF LOCATIF), (LOCATIF HYPERONYME LOCATIF) edo (HYPERONYME PARTIE-DE PARTIE-DE)<sup>1</sup> hirukoteak erazagutzeak adierazten duena zera da, alegia partekotasun-erlazioen silogismoak egiazko izateko derrigorrezko baldintza ondorioko erlazioa premisetako "beherena" izatea dela, erlazio mota desberdinen artean ezarritako hierarkia CLASS INCLUSION > MEROLOGICAL INCLUSION > SPATIAL INCLUSION<sup>2</sup> dela emanik (Winston *et al.*, 87:435). Horren inguruan, M.W Evens-ek ere (1988) aipatzen du erlazio meronimiko eta taxonomikoen konbinazio bat, gure formalismoan (MEMBRE-DE HYPERONYME MEMBRE-DE) tankerako hirukoteaz adierazten ahal dena, eta arestian azaldu dugun erlazio ordenatze horrekin bat letorkeena. Kontuan hartu behar da ordea, hirukote hauetako batzuk, HYPERONYME erlazioa lehen tokian eta herentzidun erlazio bat bigarrenean dutenak hain zuzen, ez ditugula geuk horrela erazagutu HEBko inplementazioan, herentzi mekanismoa bera aski baita horrelako konposaketa lexikalez lor genitzakeenak lortzeko.

HEBko errepresentazioan hiztegi-metalengoiatik eratorritako erlazioak ere badaudenez, erlazio lexiko-semantikoen konposaketarako mekanismo hau are baliagarriago bihurtzen da. Metalengoiatzko erlazio horien arteko konbinazio posibilitateak nahiko handiak dira, izan ere. Ikus dezagun, adibidez, zein erregela sortuko litzatekeen |géranium I I| kontzeptuaren OBJECTIF atributua atzitzerakoan, atributu honen COMPOSITION-LEXICALE fazetan (OBJECTIF CE-QUI OBJECTIF) hirukotea erazagutu dugula emanik. Atzipena egiterakoan bada, AVCOMP-LEX irakurketa-iratxoak konposaketa lexikaleko erregela bat sortuko luke, REGLES-COMPOSITION klasean, eta arrazoiketa askaraziko luke. Hona hemen sortuko litzatekeen erregela eta berorren aplikazioaren ondorioa kasu honetan:

<b>If</b>	X OBJECTIF Y and Y CE-QUI Z	;;; X-ren helburua Y da (entitate bat) ;;; Y "est ce qui" Z (egintza bat)
<b>then</b>	X OBJECTIF Z	;;; X-ren helburua Z da (egintza)

---

<sup>1</sup> Azaldu beharra dago hemen HYPERONYME —eta orobat HYPONYME— erlazioaren agerpen hori, hirukoteetan. Esana baitugu erlazio horiek, THESAURUSen, sistemaren klase/azpiklase erlazioen bitartez inplementatuak izan direla, eta horrela, kontzeptu baten hiperonimo zein hiponimoak ez direla atributu baten balio, beste erlazioen balioak ez bezala (hori KEEri eta ez beste ezeri dagokion kontua da, zeren kidako beste software batzuk ez baitute horrelako bereizketarik egiten). Hori benetako oztopoa zen guretzat, HEBko erlazio taxonomikoen tratamendua beste erlazio batzuenarekiko oso era desberdinean pentsatzea zekarrelako, inplementazioari begira, atzipen-funtzio desberdinak, eta abar. Irtenbidea atributu erlazional "alegiatzko" batzuk sortzea izan da —CONCEPTS-TYPEn definitutako HYPERONYME, HYPONYME, ANCETRE eta DESCENDANT—, berezko baliorik izan ez arren beharrezko denean kalkulatzeko dutenak (irakurketa-iratxo bati esker). Horrela, hiperonimoak —berekalakoak zein "arbasoak"— eta hiponimoak —berekalakoak zein "ondorengoak"— bestelako erlazioen balioak bezalaxe lor daitezke; hori, gainera, interesgarria da oso oinarrizko funtzioetan erlazioak era homogeneo batean tratatu ahal izateko (ikus hurrengo kapituluan).

<sup>2</sup> CLASS INCLUSION > MEROLOGICAL INCLUSION > SPATIAL INCLUSION, gurean, gutxi gorabehera, HYPERONYME > PARTIE-DE > LOCATIF litzatekeena, HEBko atributuak erabiliz.

zeinek, |géranium I 1|en OBJECTIF atributuan aplikatuz gero eta atributu horren balioa |ornement I 1| dela jakinik (ikus § III.1.3.4en), |orner I 1| balio berria lortuko bailuke atributu honetarako, erregela honela instantziatuz:

**If** |géranium I 1| OBJECTIF |ornement I 1| **and**  
 |ornement I 1| CE-QUI |orner I 1|  
**then**  
 |géranium I 1| OBJECTIF |orner I 1|<sup>1</sup>

Hirukoteen mekanismoa are adierazkorrago bihurtzen da alderantzizko erlazioak erabiltzeko posibilitatea hartzen badugu kontuan. Meronimiaren eta hiperonimiaren arteko interakzioa, (Evens ed., 88:25)en “*If for all x, y, and z, x ISA y and z PART y, then z PART x.*” espresioaz emana, aise adieraz daiteke (HYPERONYME PARTIE-DE+INV PARTIE-DE+INV) hirukotearen bitartez<sup>2</sup>, zeinek ondoko erregela sortaraziko bailuke PARTIE-DE+INV atributua atzitzerakoan:

**If** X HYPERONYME Y **and** ;;; X-ren hiperonimo bat Y da  
 Y PARTIE-DE+INV Z ;;; Z Y-ren parte bat da  
**then**  
 X PARTIE-DE+INV Z ;;; Z X-ren parte bat da

Prototipoko atributu erlazionalen *COMPOSITION-LEXICALE* fazetan ezarritako hirukoteen artean aipa ditzakegu, adibide batzuk ematearren, (CARACTERISTIQUE QUI-A POSSESSION), (CARACTERISTIQUE QUI QUI), (CARACTERISTIQUE SANS CARENCE)<sup>3</sup>, eta abar. Definitutako hirukoteen zerrenda osoa ikusteko jo beza irakurleak txosten honen D eranskinera, STRUCTURES azpi-EBaren listatua daukanera alegia.

Esan beharra dago, prototipoaren egungo egoera kontuan harturik, eta ikusita bertako ambiguetate lexikala aski hedatua dela, trantsitibitate-erregelaren emaitza ez dugula onartzen tarteko kontzeptua lexikalki ambigua baldin bada. Segurtasun neurri honekin egitate faltsuak ondorioztatzea ekidin nahi genuke.

Hirukoteen mekanismoa eta erregelen sorkuntza dinamikoa nahiko konponbide malgua dela iruditzen zaigu, konbinazio lexikaleko fenomeno horien ezarpenari aurre egiteko. Mekanismoa sinplea da oso, deklaratioa, eta oso aise egin daitezke aldaketak, nahi izanez

<sup>1</sup> Inferentzia honek —*orner I 1*: *ce qui orne* definizioari esker egina— *géranium*en helburua *orner* egintza dela adieraziko luke.

<sup>2</sup> Izatez, adibideko kasu konkretua gauzatzeko ez legoke hori egin beharrik gure sisteman, inferentzia hori herentzi mekanismoari esker egingo baita: PARTIE-DE+INV atributuaren herentzi rola *VERRIDE.VALUES* da

<sup>3</sup> *CARACTERISTIQUE*, *QUI-A*, *POSSESSION* eta *CARENCE* ENTITES klasean definituriko atributu erlazionalak dira, *SANS QUALITES*en eta *QUI* bai ENTITESen eta bai QUALITESen. Hirukote horien alderantzizkoak honakoak dira: (*QUI-A+INV* *CARACTERISTIQUE+INV* *POSSESSION+INV*), (*QUI+INV* *CARACTERISTIQUE+INV* *QUI+INV*) eta (*SANS+INV* *CARACTERISTIQUE+INV* *CARENCE+INV*). Alderantzizko hirukote horiek ere erazagutu dira dagozkien erlazioen atributuetan (hirukoteko hirugarren elementua).

### III. kapitulua

gero. Aukera ematen du, besteak beste, hirukote desberdinekin jolasteko, ea nolako emaitzak sortzen diren ikusteko.

Bukatzeko, aipa dezagun "arrazoiketa sinonimiko" esaten diogun dedukzio-estrategia, hurrengo kapituluko hainbat funtzionalitatearen ebazpidean erabili asmo dugun estrategia orokorra denez. Arrazoiketa sinonimikoa diogunean, kontzeptu baten sinonimoak kontzeptu horren ordeztu erabiltzea esan nahi dugu, edozein arrazoitze-prozesutan aurrera jarraitu ahal izateko estrategia bezala. Arestian azaldu dugun mekanismoa egokia da horrelako arrazoiketa modelizatzeko: nahikoa da, izan ere, atributu erlazional guztien *COMPOSITION-LEXICALE* fazetan (R SYNONYMES R) hirukotea eranstea, non R-k edozein atributu erlazional adierazten baitu. Horrela, X eta Y kontzeptuen arteko R erlazioa Y-ren sinonimoetara hedatzea lortzen dugu, aski modu dotore eta erosoan.

Erlazioen konposaketaren problema honentzako bigarren aukera *konposaketa lexikaleko erregela esplizituak* idaztean datza. Konposaketa kasu "konplexuenak" adierazteko aldeztu aurretik idatzitako erregelaz baliatu gaitzake; horrelakoak, noski, hirukote sinple baten bitartez adierazi ezin diren kasuetarako utziko genituzke (premisia gehiago, ondorio konplexuagoa, eta abar). Erregela hauekiko arrazoiketa ere atributuari loturiko irakurketa-irratxo berberak askaraziko luke, arestiko erregela dinamikoki sortuen kasuan bezalaxe. Horrelako erregelak *COMPOSITION-EXPLICITE* klaseko elementu izango dira<sup>1</sup> (ikus *REGLES-COMPOSITION* klasearen azpian, D eranskinean).

#### III.3.2 Propietateen herentzia.

Kontzeptuen errepresentazioaz aritu garenean azaldu dugunez, kontzeptuak hierarkia bateko klase bezala errepresentatzeak permititzen du kontzeptu horien propietate eta erlazioen herentzia —errekurtsiboki— beren hiponimoetara (haien azpiklase bezala errepresentaturik). Hiponimoetaranzko balio-herentzia hori atributu bakoitzaren kasuan era konkretu batean gauza daiteke, beti ere atributu horrentzat definitutako herentziaren arabera. Ohargarria da herentziaren mekanismoak hiperonimoen erlazio eta propietateak fintzeko bide bat eskain dezakeela (ikus gorago § III.2.1en, *OVERRIDE.VALUES* rolari buruzkoa).

Herentzia HEBaren aberasbidetzat hartu badugu ere, erabiltzailearen eskaripeko inferentziaz ari garen honetan ere aipagarria iruditzen zaigu, HIZTSUAN ezarritako arrazoiketa-mekanismoen funtsezko aspektu bat delako.

Propietate eta erlazioen herentzia hiperonimia/hiponimiaren trantsitibotasunaren ondorio zuzena dugu; gure ustez, herentzia hori funtsean herentzia lehenetsitzat hartu behar genuke,

---

<sup>1</sup> Horrelako erregelarik ez da oraingoz definitu.



hots, informazio heredatua kontuan hartzeko bertako informaziorik —lokalik— ez egotea eskatuz. Kasu batzuetan, zehaztutako herentzi rola dela medio —*UNION* delako rolaren kasuan, esaterako— informazio lokala eta heredatuaren arteko bereizketa egin beharrik suerta daiteke, bi informazio multzoak batera aurkeztuko baitzaizkigu atributuaren balio gisa. Herentziaren aspektu hori kontuan hartu beharrekoa da erabiltzailearen kontsultak eragindako arrazoiketari ekin behar zaionean. Bereizketa hori egin ahal izateko, kontzeptu baten erlazioak itzultzen dituen metodoak lista desberdinetan biltzen ditu multzo bakoitzeko balioak: batean balio lokalak bakarrik eta beste batean balio heredatuak. Horretara, martxan dagoen arrazoitze-prozesuaren arabera aukera daitezke zeintzuk balio interesatzen zaizkigun.

Alderantzizko erlazioen implementazioan sortutako arazo bat —eta balioen herentziarekin zerikusirik daukana— komentatu nahi genuke amaitzeko. Esana dugu, alderantzizko erlaziook implementatzeko, atributu batean balio bat gehitzerakoan —edota kentzerakoan— exekutarazten diren idazketa-iratxoak erabili ditugula. Orobat esan dugu alderantzizko erlazio horiek implementatzeko nahiago izango genuela egiaren mantenimendurako sistemaren (TMS) justifikazioei loturiko egitate bezala implementatzea, erlazio horiek balio lokal bezala baino balio eratorri bezala ezarriz. Baina hori ez dela posible izan sistemaren mugengatik, eta horrexegatik hartu behar izan dugula exekuzio automatikoko prozeduren bidea. Aukera horren alde txarra ordea, da, alderantzizko balioak balio lokal bezala gordetzen direla, nahiz eta balio heredatu baten kariaz sortuak izan. Eta horixe da, hain zuzen ere, batere gustuko ez duguna, zeren eta horrela balio heredatuetatik datozen alderantzizkoak jatorrizkook baino era "iraunkorragoan" erregistratzen baitira HEBan; horren ondorioz, jatorrizko balioa desagertuko balitz —balio lokal bat sortu delako, adibidez— alderantzizko erlazioak bere hartan iraungo luke.

Arazo horri irtenbide bat emateko, *RELATION-INVERSE* iratxoaren exekuzioa galaraztea erabaki dugu jatorrizko balioa balio heredatu bat den kasuetan. Horrek badu, noski, bere alde txarra ere, zeren balio heredatuaren kasuan ez baita segurtatzen erlazioak bi zentzuetan daudenik (esplizituki); hala eta guztiz ere, horrela jokatzeko nahiago izan dugu. Azken problema hori ebatz daiteke dinamikoki erabiltzailearen eskariari erantzuterakoan, erlazio lexiko-semantikoen konposaketa mekanismoaren bitartez<sup>1</sup>. Horrela ez genuke dedukzio-ahalmenik galduko ezpada dinamikoki lortuko, HEBan fisikoki grabatutako ezagutza "hondatzeke".

---

<sup>1</sup> Nahikoa litzateke (R HYPONYME R) hirukotea atributu erlazional guztietan gehitzea.

### III.3.3 Erlazioen eta THESAURUSeko kontzeptuen arteko loturak.

HEBaren azpiegitura deduktiboari dagokion beste aspektu bat, STRUCTURES azpi-EBan definitutako erlazioen eta THESAURUSeko kontzeptuen arteko harremanena da.

Gurean aldeaz aurretik finkatutako erlazioak —hizkuntzako hitzekin etiketatuak, hala nola, CE-QUI, PARTIE-DE, MANIERE-DE, DEVENIR, etc.— erabiltzearen aldetik jo badugu ere, interesgarria da erlazio horien eta THESAURUSen errepresentaturiko hizkuntz kontzeptuen arteko loturak zehaztea.

MANIERE-DE atributu erlazionala eta |manière I 1| kontzeptua nolabait lotuko bagenu, atributu hori daukan kontzeptua "zerbait egiteko era edo modu bat (*manière*)" dela adieraziko liguke; era berean, CARACTERISTIQUE |propriété I 3|<sup>1</sup> kontzeptuarekin lotuz gero, atributu horren balioak bere entitateen propietateak (*propriété*) direla jakiteko balioko luke.

LPPLk *petit I 1* sarrerarentzat ematen duen definizioa —*de faible dimension*— (413. orr.) THESAURUSeko maila erlazionalan RELATIF-A<sup>2</sup> atributu erlazionalaren bidez errepresentatzen da, erlazioa |petit I 1| kontzeptua eta '*faible dimension*' izen-sintagmaren adierazpide den |dimension I 1#2| konfigurazioaren artean ezarriz. Hori interesgarria da, zeren eta *petit I 1*-en definizioari esker jakin baitakigu, adjetiboa, kualifikatzen duen entitatearen tamainari (*dimension*) dagokiola eta ez beste edozeri. Eta hori dena esateko, agian interesgarria litzatekeela RELATIF-A |propriété I 3| kontzeptuarekin nolabait erlazionatzea —atributua daukan kontzeptuak propietate bat (*propriété*) denotatzen duela adieraziz oraingoan—, baina kasu honetan lotura izan litekeela are aberatsagoa edozein propietate denotatzen duela adierazi baino neurriari dagokiola zehazki adieraziko bagenu.

Bistakoa denez, adibide bezala eman ditugun erlazio eta kontzeptuen arteko harreman horiek askotarikoak ditugu. Zenbaitetan atributuaren jabe den kontzeptuaren ezaugarri bat adierazten dute, bestetan balioaren ezaugarri bat, eta askoz ere gauza konplexuagoak ere bai beste batzuetan. Beraz, arazoa atributuen eta hizkuntzako kontzeptuen artean zer-nolako loturak ezarri behar ditugun, eta lotura horien semantika zein den ongi zehaztean datza, ezagutza horietan oin hartuko duten balizko arrazoiketak zuzenak izan daitezen nahi baldin badugu.

Lotura horiek beharrezkoak ditugu HIZTSUAri erabiltzailearen kontsultarako malgutasun handiago bat emateko. Oinarrizko funtzionalitateen mailan ere (ikus IV. kapitulua) nahi genuke HIZTSUA gai izatea erlazioa kontzeptu baten bidez adierazia duten eskariak

---

<sup>1</sup> *caractéristique* hitza ez da LPPLko sarrera!

<sup>2</sup> *de* hitzarekin hasi eta ondotik izen-sintagma bat duten adjetibo-definizioak errepresentatzeko DEF-RELATIFA atributu definitorioa erabiltzen da, balioa berehala RELATIF-A erlazionalera goratzen delarik.

onartzeko, erlazio aurrefinkatuak dituztenak bezalaxe. Sistemak horrelako gaitasuna izatea nahi badugu, ezinbestekoa da hizkuntz kontzeptuak eta erlazioak errepresentatzeko erabili atributuak nolabait lotzea. Horrela modelizatuko dugun erlazioei buruzko metaezagutzak sistemaren dedukzio-ahalmena handiagotu beharko luke, azken finean erlazioen "esangura" deskribatzen ari baikara, horretarako HIZTSUA den hiztegi-tresna adimendunaz baliatuz.

Prototipoan honen inguruan zer egin dugun azaltzea besterik ez zaigu geratzen. Gorago aipatu dugunez (ikus § III.1.3.2.2), erlazio-kontzeptu lokailu mota bat baino ez dugu ezarri oraingoz prototipoan, atributu erlazionalen fazeta batez baliatuz: *CONCEPT-TYPE-CORRESPONDANT*. Fazeta honek THESAURUSeko kontzeptu zenbaiten erreferentzia eduki dezake: erreferentzia horiek eskari batean erlazioaren lekuan erabiliak direlarik, ezarritako loturek adieraziko diote sistemari zein atribututan aurki daitezkeen eskatutako balioak.

Horrela bada, erabiltzailearen eskari batean zerbaiten propietateren (*propriété*) bat nahi bada, sistemak jakin dezake *CHARACTERISTIQUE* atributuan bila dezakeela propietate hori, atributu hori |*propriété* I 3| kontzeptuari loturik dagoenez; era berean, zerbait egiteko modua edo era baldin bada eskatzen dena (*manière, façon*), |*manière* I 1|<sup>1</sup> kontzeptuarekin lotutako *MANIERE-DE+INV* atributura jo dezake informazioaren bila.

Txostenaren D eranskinera jo daiteke prototipoaren egungo bertsioan ezarritako erlazio-kontzeptu lotura horiek zein diren ikusteko (erreparatu atributu erlazionaletako *CONCEPT-TYPE-CORRESPONDANT* fazetan).

### **III.3.4 Ziurtasun-ezaren tratamendua: HEBan dirauen anbiguetatea.**

HIZTSUAren azpiegitura deduktiboa esaten diogun horren inguruan ahotan hartu nahi genukeen azken aspektua, HEBan dirauen anbiguetatearena da. Anbiguitasun lexikal eta sintaktikoari dagozkion arazoak azaldu ditugu hemen goian (§ III.2.2.3), HEBaren bigarren aberaste-faseaz ihardun dugunean. Aberaste-fase horretako prozesuak batch-ean exekutaturiko prozesuak dira, eta beren emaitzak —onak ala ez hain onak— HEBan geratzen dira. Horiek gorabehera, dudarik ez da HEBan badela oraindik anbiguaterik, bai lexikala eta bai sintaktikoa. Eta hori ezin da ahaztu HIZTSUAren kontsultei erantzuterakoan.

Orain interesatzen zaiguna ordea, da, ikustea ea zer posibilitate ditugun "zerbait deduzitzeko" anbiguetateren batekin topatzen garenean. Erabiltzailearen eskaripean sistemak dituzkeen arrazoitze-gaitasunei buruz ari gara bada.

---

<sup>1</sup> Hemen ere alderantzizko erlazioek eskaintzen duten malgutasuna interesgarria suertatzen da: *MANIERE-DE+INV* atributua da modua edo era adierazten duten balioak eduki ditzakeena, eta beraz, horri dagokio |*manière* I 1| kontzeptuarekiko lotura.

### III. kapitulua

Ez anbiguetate lexikalak, kontzeptu anbiguen bidez errepresentatuak, eta ez anbiguetate sintaktikoak, jeneralean konfigurazioetako atributu sintagmatikoen bitartez errepresentatuak, ez dute inongo dedukziotarako aukerarik ematen; ez, ezarritako arrazoitze-estrategia orokorraz baliatuz bederen. Alde batetik, dedukzio-kateamendu guztiak gelditu egiten dira, printzipioz, kontzeptu anbiguo bat topatutakoan; bestetik, atributu sintagmatikoetan oinarritutako arrazoiketarik ez da egiten. Baina, ez ote dago ezer deduzitzerik, nahiz eta dedukzio ez oso "ziurrak" izan, anbiguetate kasu horien aurrean gaudenean? Horixe da azpiatal honetan erantzuten ahaleginduko garen galdera. Hala ere, problemari aurre egiteko hiru estrategia aipatu besterik ez dugu egingo:

- *Erabiltzaileari laguntza eskatu*: erabiltzailearekiko elkarrekintzan egiten diren arrazoiketetan sistemak eska diezaioke erabiltzaileari laguntzeko, auzitan dauden kontzeptuak edota erlazioak desanbiguatzen.
- *Erantzun ez-ziurrak*: atributu sintagmatikoetako fazeta erlazioaletan dauden ziurtasun-balioetan oinarriturik, sistema irits daiteke inferentzia ez-ziurrak egitera, eta erabiltzaileari eskaini ziurtasun-ez hori esplizituki adieraziz. Horretarako, sistema inferentzia ez-ziurreko erregelez eta ziurtasun-ezaren kalkulurako arauetz hornitu behar litzateke.
- *Desanbiguketa-heuristikoez baliatu*: HEBaren bigarren aberaste-fasean aipatutako heuristikoez dinamikoki baliatzea litzateke. Batch-eko aberasketan porrot egindako heuristikoak izanik ere (desanbiguketa konkretu honi dagokionez, noski), beren erabilera dinamikoa egin nahi genuke, aurreko bi estrategiekin osatuz. Hau da, erabiltzailearen laguntzaz baliatuz eta ziurtasun maila bateko dedukzioak egiteko posibilitatetarekin.

Estrategia horiek egungo prototipoan ezartzeari begira garatutakoa nahiko mugatua dela esan behar da.

Anbiguetate sintaktikoaren tramenduari dagokionez berriz, erlazioak eta atributu sintagmatikoak lotzeko mekanismo bat ezarri dugu: mekanismo horri esker, erlazio batera gora daitezkeen atributu sintagmatikoak erlazio horren atributuari lotzen zaizkio, eta horrela sistemak badaki "nora jo dezakeen falta zaion erlazio horren bila". Honela bada, atributu erlazional guztiek dute *SYNTAGMATIQUES-PROMOUVABLES* izeneko fazeta, non atributu erlazional horretara goragarriak diren balioak zein atributu sintagmatiko edo metalengoaiazkotan aurki daitezkeen adierazten baita. Dedukzio-prozesu batean, aurrera egiteko C kontzeptuaren R erlazioaren balioa behar delarik, sistemak erreparatuko du — fazeta horretako informazioari esker— ea C-ren edota bere sinonimoren baten definiensa errepresentatzen duen konfigurazioan (definizio klasikoaren kasuan, jakina) ba ote dagoen R-rako goragarria den atributu sintagmatikorik; hala izatera, sistemak bi aukera litzuzke, edo

erabiltzaileari eskatu, egin bide daitekeen goratze horren zuzentasuna berma dezan, edo erantzun ez-ziur bat eman. Erabiltzailearen erantzunak bide emango luke, printzipioz, arrazoiketari jarraitzeko eta baita, erabiltzaile aditu baten kasuan behintzat, sistemak ezagutza berriak ikas ditzan.

Atributu erlazionalei esleitu zaien prozedura batek —AVSYNT-PROM irakurketa-iratxoa— atzitzen du aipatu fazeta eta anbiguetateak ebazten saiatzen da, horretarako aberasketaren bigarren fasean definitutako desanbiguaketa sintaktikoko erregelak martxan jarritz (ASK.USER moduan, hots, erabiltzaileari baieztapena eskatzen zaiola).

Azterkizun geratu zaigun beste aukera bat, erabiltzaileari erantzun atributu sintagmatikoetan zuzenean oinarrituak ematearena da. Azken finean, hemen goian azaldutako soluziobideen antzekoa litzateke, eta, gainera, horrelako atributuetan finkatutako arrazoiketak dukeen oinarri falta genuke oztopo nagusia.

Anbiguetate lexikalen ebazpen dinamikoaren aldetik berriz, ez da ezer inplementatu prototipoan. Zentzuzkoa da pentsatzea anbiguetate sintaktikoen kasurako urratutako bideon antzekoak egoki litezkeela kasu honetan ere.

### **III.4 Prototipoaren egungo egoera.**

Aberaste-fase honen ondoren prototipoa zein egoeratan geratu den azaldu nahi genuke hemen, labur-labur.

Bigarren aberaste-fasea buruturik —sinonimiaren ustiatzea eta eraikuntza taxonomikoa, batik bat<sup>1</sup>—, sareko loturen kopuru osoa 21800era iritsi da (vs. 19691 hasierako eraikuntzaren ondoren), hots, %10.7ko gehikuntza. Gehikuntza hori ez da berdinki banatua: lotura taxonomikoak %21.9 gehitu diren bitartean (8537 vs. 7003 lotura), esplizituen kopurua %4.47 baino ez da hazi (12054 vs. 11538 lotura)<sup>2</sup>; herentziaz lortutako loturak berriz, hasierako eraikuntzan zeuden 1150 haietatik 1209 izatera pasa dira aberasketaren ondoren (%5.1eko gehikuntza)<sup>3</sup>.

---

<sup>1</sup> HEBko desanbiguate-prozesuak ez dira oraindik burutu, nahiz eta erregela batzuk inplementaturik eta probaturik egon dagoeneko.

<sup>2</sup> Sinonimiaren trantsitibotasunaren ustiatzeak ekarritako gehikuntza.

<sup>3</sup> Herentziazko lotura horien gehikuntza txiki samarra izate horrek badu bere esplikazioa: izan ere, definizio sinonimikoen eta "klasikotzat" jotakoen arteko disjuntzioa ia erabatekoa baita iturburu-hiztegian. Disjuntzio horren kariaz, eraikuntza taxonomikoaren hedatze prozesua ezin da oso eraginkorra izan herentziari dagokionean, nahiz eta bestela emaitza politak lortu (oro har, oso gutxi ezagutzen da "genusaren sinonimoez", eta ez da ezagutzen "sinonimoen genusa").

### *III. kapitulua*

THESAURUSEko kontzeptu bakoitzak 3.55 lotura ditu sareko beste kontzeptu batzuekiko<sup>1</sup>. REFERENCESEkoek berriz, 3.86 lotura bakoitzak, batez-besteko.

Konposaketa lexikalaren mekanismoa prototipoko kontzeptuen lagin batekin martxan jarrita egindako kalkulu txiki batek erakutsi duenez, THESAURUSEko lotura kopuru osoak %5-10eko gehikuntza izango luke horrela (oraingoz erazagaturik dauden hirukoteekin, jakina). D eranskinean ikus daitezke kalkulu hori egiteko atributu erlazioaletan erazagututako hirukoteak; hirukote horiek konposaketa lexikaleko beste horrenbeste erregela sortarazi dute estatistika hauek ateratzean.

---

<sup>1</sup> Oraingoa konfigurazio-unitateak kontatzeke, ENTITSEko ume (berehalako) diren kontzeptuek —REFERENCESEko edota DEFINITIONSen azpiklaseren bateko elementuak— batez-beste 4.2 lotura dituzte THESAURUSEko beste kontzeptu batzuekin; ACTIONS/EVENEMENTSEkoek 3.62, QUALITSEkoek 3.53 eta ETATSEkoek 1.89. Konfigurazioek, beren aldetik, 2.73na lotura dituzte batez-beste.

# BIGARREN PARTEA: HIZTEGI-SISTEMA URGAZLEAREN ALDERDI FUNTZIONALA

## **IV. HIZTSUAren funtzionalitatea.**

Laugarren kapitulu honen xede nagusia HIZTSUAren oinarrizko funtzioen multzoa aurkeztea da. Multzo hori erabiltzaile "arruntarentzat" pentsatutako hiztegi-kontsultarako funtzionalitateek osatua da, eta, mugatua izanik ere, ohizko sistemen funtzionalitateak baino zerbait gehiago eskaini asmoz sortua da. Horrela bada, kapitulu honetan aurkeztuko dugunak hiztegi-sistema urgazlearen alderdi funtzionala islatu nahi luke, bere egungo egoeran azalduz.

Sistemak izan behar lituzkeen ezaugarri funtzionalen artean, honakoak nahi genituzke:

- Informazio-galde eta informazio-egiaztapeneko funtzionalitateak izatea.
- Ezagutza atzitzeko modu desberdinak eskaintzea, hitz bidezkoa eta "esangura" bidezkoa batik bat.
- Funtzionalitate desberdinak elkarrekin konposatzeko posibilitateak izatea, eskari konplexuagoak adierazi ahal izateko.

Azpimarratu beharra dago funtzionalitate horiek HIZTSUAren Dedukzio-Moduluaren parte direla eta sistemaren azpiegitura deduktiboaren "gainetik" ezarriak daudela (ikus III. kapitulua); horretara, HIZTSUA kontsultatzean eskuratutako ezagutza inplizitua — ezarritako mekanismoek eskuratzen duten neurrian, noski— azpiegitura deduktibo horri esker lortzen da.

## *IV. kapitulua*

Jakin badakigu funtzionalitate multzo hau ez dela osoa eta, dudarik gabe, etorkizunean osatu beharra egongo dela. Funtzionalitate honen sorburua, partez behintzat, hiztegi-kontsulten ezaugarriak aztertzeraz zuzendutako zenbait saiotan dago —irakurketa-saio, batez ere—, Donostiako Informatika Fakultateko lantaldeko partaideon artean horretarako propio egindako ikersaio eta behaketa enpirikoetan hain zuzen. Erabiltzaile motari buruz zerbait esatekotan, hizkuntza ongi ez dakien erabiltzailea —batik bat— izan dugula gogoan esan behar, jaiotzatiko hizkuntza frantsesa ez duena alegia. Bistakoa da tipologia horrek ez dituela bere baitan hartzen hizkuntz hiztegi baten erabiltzaile posible guztiak, baina, hala ere, hiztegi erabilerari dagokionean nahiko erabiltzaile mota nabarmena dela deritzagu.

Aurrena, paperezko zein makinako hiztegietako informazioa atzitzeko ohizko bideei buruzko sarrera bat egingo dugu (§ IV.1) eta, ondoren, HIZTSUAren oinarritzko funtzioak azaltzera pasako gara (§ IV.2).

### **IV.1 Hiztegietako informazioaren atzipenerako ohizko bideak.**

Hiztegi-laguntzako sistema adimendun batek segurtatu behar lukeen gauzarik inportanteenetako bat, bertako ezagutzak atzitzeko aukera zabalak eskaintzea litzateke: halako sistema batek ohizko sistema baten atzibide mugatuak gainditu egin behar lituzke. Atal honetan ohizko hiztegi-sistemek eskaintzen dituzten atzibideok ikusiko ditugu, bai paperezkoenak eta bai euskarri informatikoan daudenak.

Hizkuntz hiztegi klasikoaren arazoa —hitzen atzipenari dagokionez— da hitzaren forma idatzia ezagutu ezean hitza aurkitzea ezinezkoa dela; hau da, hitza nola idazten den ahaztu egin baldin bazaigu, esaterako, erabat ezinezkoa egingo zaigu hitz hori hiztegian topatzea, nahiz eta dozena bat hitz antzeko edo, are, sinonimo jakin. Ohizko hiztegia "hitza ulertaraztera" zuzendua baitago, "esan nahi den hori esateko behar den hitz egokia bilatzen" laguntzera baino. Arazo honen irtenbide klasikoa hiztegiko hitzak beste era batera antolatzea izaten da, sarreren ordena alfabetikoa ez den beste irizpide baten arabera alegia: horrelaxe sortzen dira, hain zuzen ere, thesaurusak.

#### **IV.1.1 Paperezko hiztegia.**

Hizkuntz hiztegi mota ezagunena eta usatuena, hitzak ordena alfabetikoaren arabera sailkatuak dituena da. Horrelako hiztegietan, sarrera bakoitzeko honako hau aurkitu ohi dugu: nabarmenki definizioa —hiztetik ideiaranzko bidea—, baina baita bestelako hainbat informazio ere, hala nola, kategoria gramatikala, erabileremua, adibideak, eta abar. Beraz, hiztegi hauen ezaugarriak behinena, erabiltzaileari eskainitako atzibideei dagokienean, kontzeptuak beren adierazlearen arabera (forma idatzia) ordenatuak egotearena da;



ordenaketa horren kariaz —hitzen *antolamendu semasiologikoa* deitua—, erabiltzaileak kontsulta garaian ondokoa egin behar du: (1) hitzaren forma idatzia ezagutu, eta (2) sailean-sailean liburuan bilatu aurkitu arte. Horrelako hiztegi-tako bilaketa-prozesu tipikoetan maiz suertatu ohi da bilaketa batek beste batzuk ekartzea berarekin, batez ere aurkitutako hitzaren artikuluan egiten diren aipuak direla medio: sinonimoak, definizioan topaturiko hitz ezezagunak, etab.

Bigarren hiztegi eredu handia, askoz ere ezezagunagoa baldin bada ere, *antolamendu onomasiologikoa* duena da, hau da, hitzak esanguraren arabera antolatuak dauzkana. Horrelako hiztegi-tan, hitzen esangurari buruzko informazioa baino esanahi-eremu jakin bateko hitzei buruzko informazio egon ohi da. Hiztegi horiei "kontzeptualak" ere esaten zaie, eta, Lenders-ek (1990) dioen bezala, lexikoaren sailkapen kontzeptualerako sistema formal bat izan beharra daukate. Hiztegi-gintza tradizionalan, "ideiatik hitzerantzko atzibidearen" arazoari eman izan zaion irtenbidea thesaurus edo hiztegi ideologikoak egitea izan da: hiztegi horien asmoa erabiltzaileari hitz baten bilaketan laguntzea da, erabiltzaileak hitz horren hurbilekoak edota bere esanahia gutxi gorabehera ezagutzen dituela emanik.

Azken multzo horretan ditugu bada, thesaurusak; thesaurus horietan hitzak "ideien" arabera sailkaturik daude, erabiltzaileak, horrelako edo halako ideiarekin bat adierazi nahi duelarik, aise topa ditzan. Mota horretako lehen hiztegia, eta ezagunena ere bai inondik ere, Roget-ena da (1852). Roget-en thesaurusean hitzak sei kategoriatan daude sailkaturik: erlazio abstraktuak, espazioa, mundu materiala, adimendua, nahimena, eta gaitasun moralei eta zentzumenari dagozkienak. Kategoria zabal horiek azpisailetan banatuak daude, maila desberdinetan. Autoreak berak aitortzen zuen, liburuaren lehen argitalpenean, hitzak kategorietan sailkatzea lan zaila zela eta, zenbait, sail batean baino gehiagotan sailkatu beharra izan zuela. Hori horrela izanik, aurkibide moduko bat ipini zuen —bere iritziz batere beharrezko ez litzatekeena, ideien sailkapena hobeki eginga balego—, irakurleak nahi izandako esangurak aiseago aurki zitzaizkion. Aurkibide hori bada, hasieran alferrikakotzat joa, argitaralpenik argitaralpen hedatuz joan da etengabe<sup>1</sup>, eta horrexek erakusten du, S.I. Landau-k dioen bezala (1984), kontzeptuen sailkapen eta hierarkia ongi pentsatua eta artez moldatua izanagatik ere, nekez lortuko dela hizkuntza baten lexikoaren sailkatze kontzeptuala jaiotzatiko hiztunentzat naturala den era batean egitea.

Beste hiztegi motarik ere bada hiztegi "kontzeptualon" sailean, hala nola hiztegi ideologikoak —Casares-ena (1988) bezalakoak— edota sinonimo-hiztegi batzuk. Hala ere, sinonimo-hiztegiak direla eta, esan beharra dago gehienak hitzez antolatuak izan ohi direla, sarreren ordena alfabetikoaren arabera ordenatuak alegia: beraz, hiztegi hauen berezitasuna

---

<sup>1</sup> 1977ko argitaralpenean, bilaketan aurkibide horrexetatik has dadin gomendatzen zaio irakurleari (Landau, 84).

#### IV. kapitulua

ez datza eskaintzen duten atzibidean —ohizko hiztegien berdina—, baizik eta kontzeptuak esplikatzeko duten moduan, definizioak gabe beste hitzen erreferentziak edo aipuak erabiltzen baitituzte (sinonimoak, eta zenbaitetan antonimoak edota antzeko hitzak ere bai). Erreferentzia horiekin erabiltzailea konforme gera liteke edo, bestela, bilaketari berrekin, hiztegi berean edota besteren batean.

Hiztegi-iturri desberdinen erabilera eta lankidetzaren posibilitatea da horrek idarokitzen duena, erabiltzaile bizkor eta adimentsu baten eskuetan emaitza ederrik eman lezakeena. Hori ote da ordea, hiztegiak erabiltzean gertatzen da kasurik normalena? Hiztegi-erabiltzaileak ez daki, maiz, hiztegia behar den bezala erabiltzen, ez daki nola atera etekinik handiena esku artean duen tresnari: batzuetan argibide, ikur edo laburduren gehiegia izaten da arazo; beste batzuetan, batetik besterako erreferentziak dira hiztegiko bilaketak lan nekezak bihurtzen dituztenak. (Nagao *et al.*, 82:54)en dioten bezalaxe, hiztegietan ugari diren konbentzio tipografiko zehaztasun eskasekoak direla eta, paperezko hiztegi bat “. . . *relies greatly on intuitive human understanding.*”.

Huts egiteko beldur handirik gabe esango genuke, hiztegiak zentzuz eta adimentsuki erabilia ere, badagoela beti atziezina den informaziorik, eskainitako atzipen-bideak direla medio, hain zuzen, atziezina geratzen dena. Eta ezinezkoa da "ezkutuko" informazio hori azaleratzea, ohizko hiztegiek eskaini bitartekoez behinik behin. Argitaratzaileak ahalegintzen dira hiztegi "kontsultaerrazak" egiten baina, hala eta guztiz ere, hiztegiak gehienetan gaizki —edo gutxiegi— erabiliak direla esango genuke.

Badirudi hiztegiak euskarri informatikoetan ezartzeko aukerak bide berriak zabaldu behar lituzkeela. Kontua ez da, noski, papereko bide estu eta murriz berberak euskarri berrirako proposatzea. Zeintzuk dira baina, hiztegi-erabiltzaileari eskaini behar litzaizkikeen posibilitate berri horiek? Argi dago hiztegien erabileraz ikertu beharra dagoela, maila kognitiboan: erabiltzaile mota desberdinak, erabilinguru desberdinak, kontsulta-modu diferenteak, eta beste, hartu behar lirateke kontuan ikerketa horietan, erabilera horiek ezaugarritu eta hiztegi-kontsultarako sistema orok eman beharko lituzkeen aukerak zein diren jakitearren.

Horri buruz interesgarria da N. Calzolarik (1986; 90) dioena irakurtzea, lexiko automatizatu baten antolamenduaz ari dela egiturazko eta atzipenezko arazoak aipatzen dituenean, non lexiko-antolaketa eta -atzipenari dagozkion prozesu mentalak simulatzeak ekarriko dituen arazo psikolinguistikoak aztertzearen interesa azpimarratzen baitu; edota G. Marchionini-k (1989) entziklopedia batean bila dabilen irakurlearen eredu mentalari buruz egindako azterketa ikustea, non estudiatzen baitu irakurle horrek nola egiten dituen bilaketak eta zer eskuratzen duen, ondoren liburuko eta euskarri magnetikoko kontsulta-

prozesu desberdinak alderatzeko, informazioaren atzigarritasunaren eta erabilgarritasunaren aldetik.

#### **IV.1.2 Hiztegia euskarri informatikoan.**

Hiztegi bat informatizatzen hasiz gero berehala topatuko dugun problema informazioaren atzipenari dagokiona izango da: egin baitaiteke, izan ere, hitzak atzitzeko bide alfabetiko tradizionala baino eskainiko ez duen hiztegi elektronikorik, hau da, erabiltzaileari informazioa eskuratzeko beste biderik emango ez diona. Benetan eskasa litzateke horrelako hiztegia. Aurreko azpiatalaren bukaeran esan bezala, gizakiok hiztegiak nola erabiltzen ditugun ikertu beharra dago, hortik ateratzeko atzipen-mekanismo eta -bide berriak, hiztegi-sistemetan integratu beharko diren mekanismo eta bideak.

Hiztegi bat euskarri magnetikoan grabaturik daukagun puntutik —MRD—, erabiltzen ahal da jada testu-corpus bezala, testu-tramendurako programa baten laguntzaz esate baterako. Hitz bilaketak egin daitezke, eta kasu honetan gainera ez dugu zertan mugatu behar sarreren bilaketara; posible da, izan ere, hiztegi-testu osoaren gaineko bilaketarik egitea, flexioa eta eratorpenari dagozkion arazo morfologiko tipikoak kontuan izanik, noski. Dena den, beti ere karaktere-kateen mailako bilaketak izango dira egin ahal izango ditugunak.

Orain, daukagun softwarea informazio dokumentalaren bilaketan espezializaturiko softwarea bada, orduan bilaketa posibleak askoz ere sofistikatuagoak izan litezke, nahiz eta kasu horretan ere ez gainditu karaktere-kateen maila. Baina sistemak thesaurus konbentzional bat baldin badu, eta thesaurus horri esker sinonimo bidezko bilaketak egiten ahal badira, adibidez, orduan sistemaren ahalmena handiagoa izango da.

Horrez gain, hiztegi-testuari nolabaiteko egitura bat ezartzen baldin badiogu, datu-baseen gestiorako ohizko sistema batean sartuz, esate baterako, orduan aurreko pasarteetan aipatu kontsulta eta atzipenerako bideei hainbat eta hainbat posibilitate berri erantsiko dizkiegu. Esan dugunaren adibide, jakina, hor ditugu gure proiektuan LPPLrekin egindako HDBaren gainean burutu ahal izan diren lan guztiak (ikus II. kapitulua).

B. Boguraev-ek (1988), datu-baseetan grabaturiko hiztegien funtzionalitateaz diharduela, bi funtzio-multzo handi bereizten ditu: (1) informazioa eskuratzeko galde-funtzioak; (2) hiztegien sorkuntza eta mantenimendurako funtzioak. Lehen multzoan —eta hori da hemen interesatzen zaiguna, kapitulu honetan aurkezten ditugun funtzionalitateak kontuan harturik— bilaketa-funtzioak aipatzen ditu, eduki espezifikoetarako buruzko eskarien araberrako listatuak emango lituzketenak, eta "nabigazio" funtzioak. Beretzat bilaketakoak bezain

#### IV. kapitulua

garrantzitsu diren azkeneko horiek direla eta, patroï edo eredu konkretu bat zehaztu eta bilatzea, eta berorren inguruan "nabigatzeko" posibilitateak idarokitzen ditu.

Gaur egun salgai dauden hiztegi informatikoeak eskaintzen dituzte azaldu ditugunak bezalako funtzionalitateak: karaktere mailako bilaketak —pattern-matching-a barne—, datu-baseetako eremu bidezko bilaketa tipikoak —sarrera, kategoria gramatikala, erabileremua, etimologia, definizioa, sinonimoak, aipuak, adibideak—, konektore logikoak erabiltzeko posibilitatea, irteeran informazioa bahetzeko aukerak, eta abar. Horrelako sistemen konfigurazio ohizkoenak bi osagai izaten ditu: datuak batetik, hiztegi-informazioa behar den bezala egituratu, indexatu eta grabaturik dagoen ingurunea —gero eta gehiago CD-ROMetan—, eta atzipenerako softwarea, interfaze moderno eta atsegin baten bitartez erabiltzaileari sistemaren kontsulta-funtzionalitateak aurkezten dizkiona.

Adibide gisa, dagoeneko merkatuan dagoen hiztegi bat azalduko dugu hemen: CD-ROMeko *The Oxford English Dictionary* (Oxford: Clarendon Press, 1992). Hiztegiaren argitalpena daukan disko trinkoarekin batera (20 liburuki paperezko argitalpenean!), argitaratzaileak informazioaren bilaketa eta berreskurapenerako programa bat banatzen du —argitalpen honetarako bereziki egina—. Beren publizitateari kasu egiten badiogu, egileek bereizten dituzte erabiltzaile arruntarentzat pentsatutako funtzioak eta hizkuntzalari edo ikerlariari begira diseinatutakoak. Aurreneko multzoan testu mailako bilaketak egiteko aukerak ditugu gehienbat —sarrerren gainekoak ez ezik, forma fonetiko, definizio-testu, adibide eta aipuen gainekoak ere—, dataz, kategoria gramatikaz eta etimologiaz bahetzeko posibilitateekin, eta pattern-matching oinarritzko batekin; pattern-matching-eko mekanismoak "zernahitarako" bi item ditu, '?', karaktere bakarrarekin parekatzen dena, eta '\*', edozein karaktere kopururekin parekatzen dena. Bestalde, aukera ematen du hiztegiko erreferentzia gurutzatuetan zehar "nabigatzeko", egindako bidearen arrastoa sistemak gordetzen duela. Ikerlariarentzat pentsatutako funtzionalitateen artean eskari-lengoaia klasiko samarra eskaintzen dute —ahalmentsua eta malgua, beraien hitzetan—, erabiltzaileak arestian aipatu eskariak baino eskari konplexuagoak egin ahal ditzan. Eskari horietan erabiltzaileak konbina ditzake eskari sinpleagoak konektore logikoak erabiliz, emaitzak gako desberdinen arabera ordenatu edo fitxategietara bideratu, eta abar. Hala eta guztiz ere, bilaketa hauek karaktere-kateetan edo aipatu bahe-eremuetan oinarrituak dira. Interfazeak erabiltzailearentzat itxura nahiko atsegina du, bistaratutako informazioa hautatzeko klikatzeko aukerak eta guzti, nahiz eta, dirudienez, erraztasun horien bitartez egin daitekeena sarrera bat aukeratu —errepresentazio-unitate oinarritzkoa— eta bistaraztera mugatzen den. Sistemak badu laguntza-sistema linean, eta aukera ematen du listatuak lortzeko, eskari komentatuak beste baterako gordetzeko, leiho-multzoa pertsonalizatzeko, eta abar, eta abar.

Bestelako hiztegi motak direla eta, esan behar da hiztegi analogiko, ideologiko, thesaurus eta abarren informatizazioak, konbentzionala izanda ere, eta erabilingurune egokietan integratu eta linean ezartzeak laguntza handikoa izan behar lukeela erabiltzaileentzat, kasu honetan bilaketak karaktere-kateen maila gainditu egingo bailuke, hiztegi mota horietan hitzen arteko erlazioak erlazio lexiko-semanticoen mailakoak diren neurrian.

W. Lenders-ek (1990) esaten duenez, aurreko azpiatalean aipatutako hiztegi kontzeptualok —hiru erlazio semantikoren arabera antolatuak gehienetan, hiponimia, sinonimia eta antonimia— usatuagoak dira lengoiaia berezitueta lengoiaia naturalean baino. Horien artean, informazio dokumentalaren bilaketa-sistemetan erabili ohi diren thesaurusak aipatzen ditu.

Lehen kapituluan aipatu dugu jada MEMODATA (Caen) enpresa frantsesak iragarritako produktua: DICOLOGIQUE™. Hiztegi "interaktibo" horretan, egileen esanetan, hitza ideiatik abiatuz (eta alderantziz) bila daiteke, eta konektore logikoen erabilera permititzen duten bilaketa-funtzio ahalmentsuak ditu. Baina, eta hau bide da inportanteena, hiztegi-informazioen artean ezarritako erlazioak maila lexiko-semanticoko batean daude. DICOLOGIQUEren diseinuan hiru alderdi hauek izan dira kontuan: lexikoaren sinonimo bidezko egituraketa, hitzen hainbat taldekatze analogiko eta termino bakoitzaren definiziorako hurbilbide izan nahi duen sailkapena. Hiztegi hau, autoreek analogiko eta sinonimiko esaten diotena, hitz bakoitzaren deskonposaketa semantikoan dago oinarritua, eta deskonposaketa hori hiztegi definizioak eskuz aztertutik eginga da.

Lineango hiztegi-informazioaren atzipenari buruz, hainbat eta hainbat autore (Webber, 84; Landau, 84<sup>1</sup>; Byrd, 90) bat dator sistemek esangura edo definizio bidezko atzipena eskaini behar luketela esatean, atzibide arruntagoak —alfabetikoak— edota hitzen eskuratzea atzipen fonetikoaren algoritmoen bitartez bideratuko luketenak ahaztu gabe.

Eskariak formulatzeko garaiko laguntzak ere aztergai dira hainbat lanetan, sistema hauen interfazea ahalmentu eta sendo bihurtuko luketen neurrian (Fox, 80; Marchionini, 89): konektore logikoen bitartez konbinaturiko eskariak, ezagutza sinonimikoaren erabilera, karaktere-kateen mailako pattern-matching-a eta forma flexionatuen onarpena atzipen-gako gisa<sup>2</sup>.

---

<sup>1</sup> Alabaina, autore honek badu hainbat zalantza lineango hiztegiok paperezkoekiko dituzten abantailak, informazioaren atzipenaren hainbat alderditatik ikusita, gainera.

<sup>2</sup> Ohargarria da hitzen gaineko "kontrol morfologikoa" derrigorrezko zerbait izango dela hiztegi-tresna bat erabilinguru jakin batean integratzeko garaian: testu-idazketa dela, ordenadorez lagunduriko irakurketarako postua dela, eta abar. Hiztegi-interfazeak onartu beharko lituzke forma flexionatu edota eratorriak eta behar den bezala analizatu, hiztegi-atzipena zuzen burutuko bada. Problema hori are larriagoa da, noski, euskara bezalako flexio maila handiko hizkuntzetan.

Hona hemen, amaitzeko, A. Zampolli-ren hitzak (1987:328), hiztegi-definizioen analisiak informazio lexikalaren atzipenaren ikuspuntutik ekar ditzakeen hobekuntzak dituela aipagai:

For example, the procedures by which computational linguists process the definitions in order to extract different semantic relations, and the resulting subsets of structured lexical units, can improve considerably for the common user the access to the lexical information. The improvement is even more evident with regard to the consultation of a bilingual dictionary.

## IV.2 Oinarrizko funtzionalitatea.

Atal honetan deskribatuko dugu HIZTSUArentzat asmatutako oinarrizko funtzionalitateen multzoa (Arregi *et al.*, 91). Multzo hori da, izan ere, arrazoiketaren aldetik ikusita sistemak duen alderdirik garrantzitsuenetako bat, erabiltzailearenganako aspektu funtzionala erakusten duena alegia. Aurreko ataletan azaldutako mekanismo guztiak —sistemaren azpiegitura inferentziala—, HIZTSUAren xede nagusiari begira daude diseinatuak: erabiltzailearen eskariaren aurrean sistema adimendun bat bezala azaltzea.

Esan gabe doa funtzio hauek diseinatzean ez ditugula kontuan hartu hiztegi-sistema orok izan behar lukeen hainbat eta hainbat funtzio; besteak beste, merkatuan dauden sistemetan aurki daitezkeen horiek, hain zuzen. Beraz, oro har, irakurleak ez du hemen aurkituko, konparazio batera, hiztegi-sarrera baten definizioa edota beste edozein atal bistaratzeko funtziorik, funtzio horietako batzuk gure oinarrizko funtzioen aldamenean ezartzea aise egingo litzatekeen arren.

Horiek horrela, oinarrizko funtzioen diseinuaren aldetik interesatu zaiguna zera izan da, alegia aurreko kapituluan azaldutako definizioen semantikaren errepresentazioa "lanean" jarriko duten funtzionalitateak zehaztea. Beraz, funtzio multzo horretan bai aurki daitekeela bi kontzepturen arteko lotura bilatzen duen funtziorik —horretarako sarea korrituz—, edo arrazoiketa sinonimikoaz baliatzen den funtziorik, kontzeptu baten sinonimoak ematera mugatzen den funtzioa baino areago; orobat, aurki daiteke murriztapen sail bat betetzen duten kontzeptuen zerrenda emango digun funtzioa, murriztapenok kontzeptuarterako erlazioen mailan (lexiko-semantikoan) adieraziak daudela; edota baita definizio-espresio batetik abiatuak dagozkion kontzeptuetara eramango gaituenik ere.

Atzera gogorarazi nahi genuke hemen funtzio hauek hiztegi-erabiltzaile arrunta gogoan izanik pentsatutako funtzioak direla. Jakina, HIZTSUAk eskaini behar lituzkeen funtzioen artean, bestelako erabiltzaileentzat pentsatutakoak ere egon behar lukete, hala nola, oso maila desberdinetan baina, lexikografoarentzat edota HEBaren ingeniari eraikitzaile-administratzailearentzat sortutako funtzionalitateak.

Oinarrizko funtzionalitate hau inplementatu egin da, partez bada ere, aurreko kapituluan azaldu dugun HEBaren prototipoaren gainean. Funtzionalitate hauetako zenbait diseinatzeko eta inplementatzeko garaian, objetuei zuzendutako programazioaren paradigmaren baitan kokatu gara esan behar, (Yokoyama & Hanakata, 86)n egiten duten moduan edo; izan ere, funtzio asko inplementatzeko metodo-esleipena aukeratu baitugu, funtzioaren ezarpena den metodoa STRUCTURESeko CONCEPTS-TYPE klaseari —gehienetan— lotuz. Ez da ahaztu behar ordea, oinarrizko funtzioen multzo hau HIZTSUAren azpiegitura deduktiboaren gainean ezarria dela eta, beraz, funtzio hauek exekutatzek arrazoiketamekanismo sail bat martxan jartzea ekarri ohi duela, hartara funtzioaren exekuzioa mezu-bidaltze simple eta ondoriozko metodo egikaritza baino zerbait konplexuago bihurtuz.

Aurrera baino lehen, azpimarratu nahi genuke ondorengo ataletan egiten den oinarrizko funtzioen sintaxi-zehaztapena HIZTSUAren Dedukzio-Moduluak eskatzen duenaren terminoetan egin dela. Horrek ez du esan nahi, inola ere, erabiltzaileari sintaxi hori ezagutu eta erabiltzeko eskatuko zaionik. Erabiltzaile-sistema elkarrekintza Komunikazio-Moduluak kudeatzen duenez, eta horren barne Eskari-Azterketako Moduluak, erabiltzaileak hemen azalduko ez den modu atseginean egin ahal izango ditu bere eskariak (ikus V. kapitulua). Eta orobat esan genezake sistemaren erantzunen adierazpenari dagokionez. Hortaz, ez dago zertan harritu eskariak adierazteko erabiltzen den sintaxi zurruna ikustean, eta ezta ere adibideetako erreferentzia kontzeptual guztiak hitzen adiera konkretuen erreferentzia direla, edota erlazioak adierazteko sistemaren erlazio primitiboak —atributu erlazonalak— erabiltzen direla ikustean. Adibideetako beraz, barne-lengoaia bat bezala hartu behar da, eta berorren helburua ez da, berez, erabiltzaile eta sistemaren arteko komunikazioa.

Erabiltzaileak aukera izan behar luke bere eskariak modu atsegin batez espresatu ahal izateko, lengoaia zurrun batek mugatua egon gabe. Hitzak bezala adierak erabiltzeko aukera eman beharko zaio, eskarien arabera noski, eta definizioak, kontzeptuen propietateak, murriztapenak, kontzeptu arteko elkarbide edota aldeak, eta abar adierazi ahal izateko modurik "naturalena" eskaini beharko zaio<sup>1</sup>, guzti horiek, azkenean, oinarrizko funtzioen argumentu bihurtuko badira ere.

Ez dira lan honen helburuen arteko erabiltzailearen eskarien aurretratamendua —Eskari-Azterketako moduluari dagokiona— inplementatzea edo sistemaren erantzunak erabiltzaileari aurkeztu baino lehen tratatzea —Erantzun-Sorkuntzako moduluak egin beharrekoa—. Beraz, eta lehen esan bezala, ondoren aurkezten diren funtzioen sintaxia berariaz sortutako *galdeketa eta erantzunen barnelengoiari* dagozkio, zein, sistemaren prototipoa probatzeko, partzialki inplementatua izan baita.

---

<sup>1</sup> Lengoaia naturala ote?

#### IV. kapitulua

Aurrenik, IV.2.1 azpiatalean azalduko ditugu eragiketa primitiboak; ondoren oinarritzko funtzioak, galde-funtzio eta egiaztapen-funtzioetan sailkatu eta banan-banan aurkeztuz joango gara (§ IV.2.2), beren nondik norakoak, sarrera/irteerako zehaztapena, sintaxia, burutzen duten prozesua, erantzunera iristeko proposatzen ditugun estrategiak eta zenbait adibide azalduz.

### IV.2.1 Eragiketa primitiboak.

Lehen-lehenik zenbait eragiketa primitibo definituko dugu, (Reimer & Hahn, 83)n egiten den moduan. HEBaren eraikuntza —DICTIONNAIRE eta THESAURUS azpi-EBena, batik bat— eragiketa eraikitzaileen multzo bat bezala ikus daiteke, emaitzatat frameetan oinarritutako datu-eredu bat eman duena. Azpiatal honetan azalduko ditugun eragiketak datu-eredu horrekiko eragiketak dira, eragiketa konplexuagoen, eta bereziki, oinarritzko funtzioen euskarri.

Eragiketa hauen bilduma ez da osoa, baina hala ere saiatu gara hemen karakteristikoak edo azaltzen.

Ondokoan bada, eragiketa hauen zehaztapena emango dugu, denotatzen dituen eragilea emanaz (ilunxeago), domeinua eta kodomeinua<sup>1</sup>, eta beren semantikari buruzko azalpen labur bat.

- Atributu baten aplikazioa kontzeptu bati

$$\sigma : \text{atributua} \infty \text{kontzeptua} \emptyset \text{atributu-balioa}$$

$A, C \emptyset A \sigma C = V$ , non  $V \subset C$  kontzeptuaren  $A$  atributuaren balioa baita. Hemendik aurrera,  $V$  beste kontzeptu baten erreferentzia denean, lotura-hirukotearen idazkera ere erabiliko dugu eragiketa hau denotatzeko:  $(C \ A \ V)^2$ .

- Kontzeptu baten definizio erlazionala      **def** :  $\text{kontzeptua} \emptyset \text{definizioa}$

$C \emptyset \text{def}(C)$ , non  $\text{def}(C)$ -k  $C$  kontzeptuaren definizio erlazionala adierazten baitu, hots,  $C$ -k beste kontzeptu batzuekin atributu erlazionalen bitartez dituen erlazioen multzoa<sup>3</sup>.

---

<sup>1</sup> Balio-mota hauei buruz jakiteko errepara § IV.2.2n (IV.1 irudia) eta IV.2.2.1en (oinarritzko datu-egituren sintaxia).

<sup>2</sup> Ikus § IV.2.2.1en.

<sup>3</sup> Zuzeneko erlazioak zein alderantzizkoak, balio lokalak, heredatuak edo bestela deduzitutakoak ere barne.



- Ahaidetasun taxonomikoa

$\Delta : kontzeptua \infty kontzeptua \emptyset boolearra \perp$

$C_1, C_2 \emptyset C_1 \Delta C_2 =$  egiazkoa, bsb bai  $C_1$  eta bai  $C_2$  taxonomia berekoak badira (THESAURUS azpi-EBan).

- Hurbiltasun semantikoa

$\equiv : definizio-espresioa \infty definizio-espresioa \emptyset boolearra \perp$

$D_1, D_2 \emptyset D_1 \equiv D_2 =$  egiazkoa, bsb bi definizio-espresioak berdinak edo "semantikoki hurbilak" badira.

Definizio-espresioon arteko hurbiltasuna dagoenentz determinatzeko irizpideak zehaztu behar dira. Hara  $D_1$  eta  $D_2$ -ren arteko hurbiltasun semantikoa egon dadin beharrezkoa den zenbait baldintza:

- definitutako kontzeptuen arteko ahaidetasun taxonomikoa egotea, "genus et differentia" erako definizio klasikoetan: (HYPERONYME  $\sigma C_1$ )  $\Delta$  (HYPERONYME  $\sigma C_2$ ),  $C_1$  eta  $C_2$   $D_1$  eta  $D_2$ -k definitutako kontzeptuak izanik, hurrenez hurren;
- (HYPERONYME  $\sigma C_1$ ) eta (HYPERONYME  $\sigma C_2$ )-ren propietateak bateragarriak izatea (lehen hurbilketa batean eskatuko da antonimoak ez izatea, gutxienez).

- Murritzapen-betegarritasuna

$| : murritzapen-espresioa \infty kontzeptua \emptyset boolearra \perp$

$R, C \emptyset C | R =$  egiazkoa, bsb  $\text{def}(C)$ -k  $R$  murritzapena betetzen badu.

- Kontzeptu arteko bide baten bilaketa

**chemin** :  $kontzeptua \infty kontzeptua \emptyset sareko-bidea$

$C_1, C_2 \emptyset \text{chemin}(C_1, C_2)$ , non  $\text{chemin}(C_1, C_2)$ -k  $C_1$  eta  $C_2$ -ren arteko bide bat denotatzen baitu, THESAURUS azpi-EBkoa; bide horren espresioa lotura-hirukote lista bat izango da, lehenengo lotura  $C_1$ -ekin hasi eta azkena  $C_2$ -rekin bukatzen dela, lotura bakoitzeko helburu-kontzeptua hurrengo loturako sorburu-kontzeptua dela, eta, kasu guztietan, tarteko elementua atributu erlazional bat izanik.

Inplementazio oso bat egitekotan, jakina, zehaztu beharko dira beste eragiketa primitibo asko; esate baterako, kategoria gramatikalarren edo hitzen bestelako ezaugarriren baten galdea (gorago aipatuak), edota karaktere-kateen gaineko eragiketak, zeintzuei esker

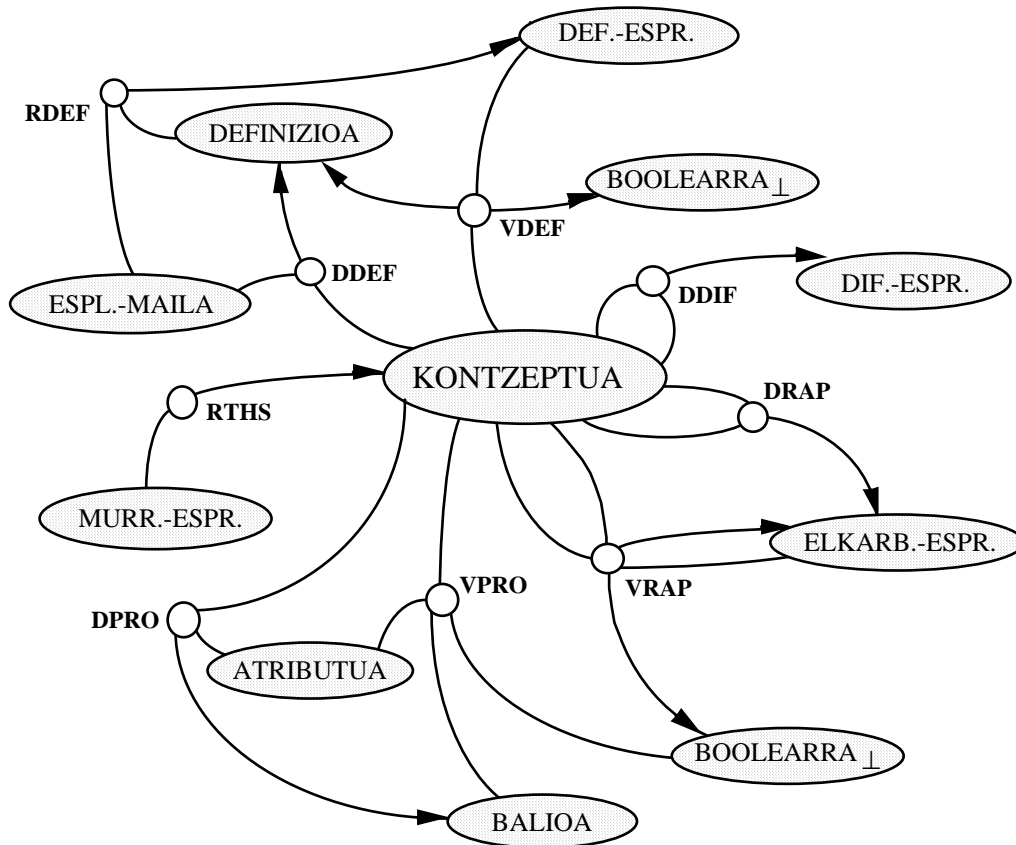
murriztapen orto-grafemikoak adierazi ahal izango bailirateke, HIZTSUA gurutzegramazaleentzat tresna interesgarria bihurtuz.

### **IV.2.2 Oinarrizko funtzioen zehaztapena. Inplementaziorako estrategiak.**

Kapituluaaren sarreran azaldu ditugun ezaugarri funtzionalen arabera, diseinatu ditugun funtzionalitateek eskaintzen diote erabiltzaileari ezagutza atzitzeko modu eta bide desberdinik.

Funtzionalitate multzo honetan (ikus IV.1 irudia) bada kontzeptu bati buruzko informaziorik ematen duen funtziorik (DDEF, DPRO); beste funtzio batzuek kontzeptu pare bat hartzen dute datutzat eta emaitza gisa bien arteko elkarbidea (DRAP) edo diferentziak (DDIF) itzultzen dituzte; hirugarren multzo batean egongo lirateke erabiltzaileak kontzeptu bat edo batzuei buruz duen ezagutza egiaztatzerazuzendutako funtzioak, sarreratzat kontzeptua(k) eta definizio-espresio bat (VDEF), elkarbide-espresio bat (VRAP) edo atributu-balio bikote batzuk (VPRO) hartzen dituztenak, erabiltzaileak duen ezagutza horren adierazpide gisa; eta azkenik ditugu bilaketa tesaurikoa esaten dioguna burutzen duen funtzioa (RTHS), murriztapen-espresio bat datutzat hartu eta kontzeptu zerrenda bat ematen duena, eta RDEF, definizio bat sarreran hartu eta definizio-espresio birformulatua itzultzen duena.

IV.1 irudiko diagraman ikus daitezke HIZTSUAren oinarrizko funtziook, datu-mota abstruekin lotutako eragiketa gisa aurkeztuak. Irudiko korapilune handiek (ilunxeago) datu-mota desberdinak errepresentatzen dituzte, arkuak eragiketen emaitzaren mota —gezi-muturrak— eta domeinua, eta korapilune zirkular txikiek eragiketak berak adierazten dituzte, bakoitzaren identifikadorea aldamenean dutela.



**IV.1 irudia.-** HIZTSUAren oinarriko ingurune funtzionala.

Horrela bada, HIZTSUAk, ideiatik edo esanguratik hitzerako inongo biderik eskaintzen ez duen ohizko hiztegi batekin eraikitako sistema izanik, lagunduko lioke erabiltzaileari "ideiatik hitzerako" bide horretan. Gure iritziz, hiztegi-sistema batek, adimenduna izatea nahi baldin badugu, erabiltzaileari lagundu egin behar dio buruan duena zehazten, behar dituen hitz egoki horien bilaketan, hain zuzen. Eta funtzionalitate hauek, bide horretan emaniko lehen urratzat har ditzakegu.

Ondokoan, funtzio hauek sailkatzeko ardatza galde/egiaztapenarena izan da. Horrela, IV.1 irudiko bederatzi funtzioak bi talde handitan banatu ditugu:

- *Galde-funtzioak:*
  - 1) Definizio-galdea (**DDEF**).
  - 2) Definizio-birformulaketa (**RDEF**).
  - 3) Kontzeptu baten propietateen galdea (**DPRO**).
  - 4) Bi kontzepturen arteko elkarbidearen bilaketa (**DRAP**).
  - 5) Bi kontzepturen arteko diferentzien galdea (**DDIF**).

#### IV. kapitulua

6) Kontzeptu-bilaketa tesaurikoa (**RTHS**).

• *Egiaztapen-funtzioak:*

7) Definizio-egiaztapena (**VDEF**).

8) Kontzeptu baten propietateen egiaztapena (**VPRO**).

9) Bi kontzepturen arteko elkarbidearen egiaztapena (**VRAP**).

Ondoren azalduko ditugu, bada, funtzio hauek banan-banan. Deskribapenarekin batera emango dugu bakoitzaren zehaztapena, sintaxia, beren funtzionamenduari buruzko ohar zenbait, eta erabilera-adibideak. Esan behar da adibide hauek hipotetikoak direla, funtzioek eskariak egiteko dituzten aukera desberdinak erakustera zuzenduak batez ere, batzuetan sistemagandik espero den erantzuna ere ematen dela; horrez gain, adibideotan emango ditugu eskarien sorburuan egon daitekeen espresioa hizkuntza arruntan, eta orobat sistemaren erantzunaren parafrase moduko bat, lengoia naturalean hori ere. Kasu guztietan esan daiteke erantzunak LPPLtik ateratzeko modukoak direla<sup>1</sup>.

Funtzioen sintaxia adierazteko EBNF formalismoa erabiliko dugu. Hona hemen lehenengo bi erregelak, galde- eta egiaztapen-eskari desberdinen zerrenda ematen dutenak:

*galde-eskaria*= *definizio-galdea* | *definizio-birformulaketa* | *propietate-galdea* | *elkarbide-bilaketa* | *diferentzi-galdea* | *bilaketa-tesaurikoa*.

*egiaztapen-eskaria*= *definizio-erlazionalaren-egiaztapena* | *propietateen-egiaztapena* | *elkarbidearen-egiaztapena*.

Funtzio bakoitza deskribatzen hasi baino lehen ordea, ondoko azpiatalean funtzioon zehaztapenean erabiliko diren oinarritzko datu-egiturak azalduko ditugu.

##### IV.2.2.1 Oinarritzko datu-egiturak: deskribapena eta sintaxia.

Azpiatal honetan azalduko ditugu ondorengoetan oinarritzko funtzionalitateak deskribatzeko erabiliko diren oinarritzko datu-egiturak, bakoitzaren sintaxia emanez.

Egitura hauek duten abstrakzio mailari erreparatuz, ondoko sailkapena egitera iritsi gara:

- Datu-mota abstraktuak, hala nola *definizio-*, *murriztapen-* eta *elkarbide-espresioak*, *definizioak*, *kontzeptuak*, *atributuak*, *atributu-balioak*, *boolearra*  $\perp$  mota eta *esplikazio maila*.

---

<sup>1</sup> HIZTSUaren prototipoarekin egindako benetako erabilera-adibideak txostenaren F eranskinean aurki daitezke.

- THESAURUS kontzeptu elkarlotuz osatutako saretzat hartzearekin zerikusia duten egiturak: batez ere *lotura* eta berorren ingurukoak (loturez eta loturetako elementuez egindako egiturak).
- HEBan erreferentzia "fisikoren" bat duten egitura oinarritzkoak.

Datu-egitura hauek deskribatzeko goitik beherako bide bati jarraituko gatzaizkio, hau da, egitura abstraktuenetatik hasi eta abstrakzio maila txikiagoa dutenatarantz joanez. Irakurleak ere, hortaz, aurrera egin beharko du testuan, abstraktuenen zehaztapen sintaktikoetan azaltzen diren egiturarik oinarritzkoenak aurkitu nahi baditu.

Ikus dezagun bada, lehenbizi, IV.1 irudian azaltzen diren datu-mota abstraktuei dagokien sintaxia:

- *Definizio-espresioa*: egitura hau, normalean, erabiltzaileak definizio bat espresatu nahi duenean usatzen da. Horretarako bi modu eskaintzen zaizkio, edo *lotura-espresioa* batez baliatzea (ikus beherago), edo kontzeptu bakarreko espresio bat erabiltzea (kasu horretan, espresioa definitutako kontzeptuaren sinonimotzat hartuko da —edo, batzuetan, hiperonimotzat, VDEFen bezala—).

*definizio-espresioa* = *lotura-espresioa* / *kontzeptua* .

- *Murriztapen-espresioa*: egitura honen bitartez erabiltzaileak murriztapen sail bat adieraz dezake, *lotura-espresio* bat erabiliz.

*murriztapen-espresioa* = *lotura-espresioa*.

- *Kontzeptuarteko elkarbidearen espresioa*: egitura honen bitartez bi kontzepturen arteko harremana edo elkarbidea espresa daiteke, horretarako edo bide-egitura bat usatuz (ikus beherago) —gehi, aukeran, murriztapen-espresio bat—, edo kontzeptu bakar bat<sup>1</sup>.

*elkarbide-espresioa* = "( "and" sareko-bidea [murriztapen-espresioa] )" / *kontzeptua* .

- *Kontzeptuen arteko diferentzien espresioa*: egitura honi esker bi kontzepturen arteko aldeak adieraz daitezke, horretarako *lotura-bikote* konjuntzioa erabiliz (ikus beherago). *Lotura-bikote* bakoitza atributu bati dagokio, bi kontzeptuek amankomunean duten atributua baina balio desberdinekin.

*diferentzi-espresioa* = "( "and" *lotura-bikotea* {*lotura-bikotea*} )" .

---

<sup>1</sup> Bi kontzepturen arteko erlazioa adierazteko hirugarren kontzeptu bat usatzeko aukera hori erabiltzaileari emandako aukera litzateke, baina ez sistemak bere erantzunak emateko erabil dezakeen zerbit.

#### IV. kapitulua

- *Definizioa*: egitura hau, sistemaren funtzio esplikatzaile nagusiaren emaitza — DDEF—, kontzeptu baten definizioa lotura sail baten konjuntzio bezala adierazteko asmatua da. Konjuntzio-egitura horretan, lotura guztien sorburu-kontzeptua kontzeptu definitua bera izango da<sup>1</sup>.

***definizioa***= *lotura-konjuntzioa*.

Ohargarria da, azaldu ditugun egiturek oraintxe ikusi behar dugun *lotura* funtsezko egituran dutela beren oinarria. Izan ere, oinarrizko funtzionalitate hauek adierazteko asmatutako notazioan, kontzeptuarteko loturen espresioak erabiltzen baitira definizio-, murriztapen-, elkarbide- eta diferentzi espresioak errepresentatzeko. Jarrai dezagun orain gainerako egitura abstraktuon azalpenarekin:

- *Kontzeptua*: kontzeptu bat errepresentatzen duen egitura, izan daiteke nahiz THESAURUSeko kontzeptu baten erreferentzia nahiz THESAURUSeko kontzeptu batekin instantziatuko den aldagai bat.

***kontzeptua*** = *thesauruseko-kontzeptua* / *aldagaia*.

- *Atributua*: kontzeptu baten espresibidetzat onartzen da, noski, sistemaren atributu primitibo bat (ikus beherago) edo aldagai bat erabiltzea, baina baita ere beste kontzeptu bat. Azken kasu horretan, sistemaren dedukzio-mekanismoei dagokie kontzeptua sistemaren atributu primitiboren batekin erlazionatzea.

***atributua***= *atributu-primitiboa* / *kontzeptua* / *aldagaia*.

- *Atributu-balioa*: atributu baten balioa, izan daiteke nahiz THESAURUSeko kontzeptu bat (atributu erlazionalen kasua) nahiz bestelako balio primitiboa (ikus aurrerago).

***atributu-balioa***= *thesauruseko-kontzeptua* / *balio-primitiboa*.

- *Esplikazio maila*: DDEF eta RDEFen argumentu den esplikazio maila errepresentatzeko sortutako mota honek, definizio baten esplikazio maila desberdinak adierazten dituzten balioak hartzen ditu. Ondoko bi maila nagusiak definitu ditugu: *testuzkoa* ("textuel", MRDko —LPPL— testuaren bitartezkoa) eta *erlazionala* (kontzeptuak sareko beste kontzeptuekiko dituen erlazioen bidezkoa). Azken maila horren barruan, gainera, ondoko azpimailak definitu dira: *lokala* ("local", kontzeptuak bere definizioari zor dizkion erlazioak biltzen dituen), *heredatua* ("hérité", aurrekoetz gain sarean ezarritako egitura taxonomikoari zor

---

<sup>1</sup> Definizioetako loturetan —sistemak emandako definizioak dira— ez da aldagairik egongo.

zaizkion erlazio heredatuak ere biltzen dituen), eta *deduzitua* ("déduit", inferentzi mekanismoei esker deduzitutako erlazioak ere biltzen dituen).

*esplikazio-maila*= "textuel" / "local" / "hérité" / "déduit".

- *Boolearra*  $\perp$ : mota honek hiru balio ditu, egiazkoa, faltsua eta ezezaguna<sup>1</sup>. Mota horri esker trata daiteke ukapenaren arazoa, datu-baseetan egin ohi denaren antzera, hau da, mundu itxiaren hipotesipean (Reiter, 78).

*boolearra*  $\perp$ = "true" / "false" / "unknown".

Usatzen diren datu-motak deskribatu eta gero, ikus ditzagun mota horien azpian leudekeen egiturarik oinarritzkoenak. Arestian esan bezala, kontzeptuen eta beren arteko erlazioen errepresentaziorako hautatutako ereduak sare tipokoa izateak, oinarritzko egitura delako lotura-hirukote hori izatea dakar, hots, kontzeptua-atributu\_erlazionala-kontzeptua hirukotea. Orain azalduko ditugun egiturek —sareko bidea, eta lotura-espresio, -konjuntzio eta -bikotea— bada, kontzeptuarteko lotura-hirukote horretan dute beren euskarria.

- *Sareko bidea*: THESAURUSeko bi kontzepturen artean dagoen bide bat lotura-konjuntzio baten bitartez errepresentatzen duen egitura, non lotura bakoitzeko helburu-kontzeptua hurrengo loturako sorburu-kontzeptua baita.

*sareko-bidea*= *lotura-konjuntzioa*.

- *Lotura-espresioa*: konjuntzio-, disjuntzio- eta ukapen-eragile logikoen bitartez elkarrekin erlazionatutako lotura multzoa.

*lotura-espresioa*= "(" eragile-logikoa lotura {lotura} ")".

non,

*eragile-logikoa*= "not" / "or" / "and".

- *Lotura-konjuntzioa*: "and" eragilearen bidez erlazionatutako lotura multzoa (aurreko egituraren kasu berezia).

*lotura-konjuntzioa*= "(" "and" lotura {lotura} ")".

- *Lotura-bikotea*: atributu erlazional primitibo berbera duten lotura pareak (ikus beheerago, diferentzi espresioari buruzkoa).

*lotura-bikotea*= "(" "and" lotura lotura ")".

---

<sup>1</sup> Funtzio baten emaitza "ezezaguna" izango da argumentu ezezagun bat dagoenean edota dedukzio-prozesuak huts egiten duenean.

#### IV. kapitulua

- *Lotura*: hirukotea, bi kontzeptuz eta bi kontzeptu horiek elkarlotzen dituen atributuaz osatua. Atributu hori atributu erlazional primitiboa izango da sistemaren erantzunetan; erabiltzaileari aldiz, permitituko zaio erlazioa kontzeptu baten bitartez adieraztea (ikus gorago esandakoa, atributuaren sintaxiari buruz).

*lotura* = "(" *sorburu-kontzeptua* *atributua* *helburu-kontzeptua* ")".

non,

*sorburu-kontzeptua* = *kontzeptua*.

*helburu-kontzeptua* = *kontzeptua*.<sup>1</sup>

Eta amaitzeko, ikus ditzagun HEBaren inplementazioan erreferentzia "fisikoren" bat duten egitura oinarritzkoak:

- *THESAURUSeko kontzeptua*: THESAURUS azpi-EBko kontzeptua —unitatea—.

*thesauruseko-kontzeptua* = "/" *sarrera* *homografo-adierazgailua* *adiera-zenbakia* ["#" *konfigurazio-zenbakia*] "/".

- *Atributu primitiboa*: THESAURUSeko klase bat —kontzeptua— ezaugarritzen duen atributu bati dagokion sinboloa (jeneralean erlazionala).
- *Balio primitiboa*: atributu ez-erlazional baten balio-klaseko balioei deritzegu honela.
- *Aldagaiak*: Eskari-erantzun batean instantziatu beharreko aldagaia.

*aldagaia* = "?" *identifikadorea*.

Aldagaion artean guk aldagai murriztua deitzen duguna bereizi behar da. *Aldagai murriztu* bat konfigurazio multzo jakin bateko unitateekin baino ezingo da instantziatu: izan ere, "/" ikurraren ondoko kontzeptuaren konfigurazioekin bakarrik instantziazten ahal da horrelako aldagai bat. Esan beharrik ez dago, horrelakoetan, kontzeptu hori unitate ez-konfigurazional bat izango dela derrigor.

*aldagai-murriztua* = *aldagaia* "/" *thesauruseko-kontzeptua*.

Eta orain ekin diezaiokegu oinarritzko funtzionalitateak banan-banan azaltzeari. Berorien sintaxia zehazteko, arestian deskribatu egitura oinarritzko horiez baliatuko gara.

---

<sup>1</sup> Aldagaiak erabiltzen ahal dira (ikus gorago, kontzeptuaren sintaxiari buruzkoa) loturetako sorburu-kontzeptuaren eta helburu-kontzeptuaren espresioetan ere.



#### IV.2.2.2 Galde-funtzioak.

DDEF, RDEF, DPRO, DRAP, DDIF eta RTHS dira HIZTSUAN definitu ditugun galde-eskariak.

Eskari horiei dagokien arrazoiketari dagokionez, ondoko estrategia orokorrak zehaztu ditugu:

- III. kapituluaz azaldu konposaketa lexikaleko erregelak martxan jartzea ala ez aukeran ematea erabiltzaileari.
- Hautazko desanbiguaketa dinamikoa —lexikala nahiz sintaktikoa—, erabiltzailearekiko elkarrekintzan.

Bi estrategia hauek lan-modu desberdinak ezartzen dituzte, HIZTSUAren kontsulta egiten denean.

Konposaketa lexikala gaituz gero<sup>1</sup>, HEBko atributu erlazionalen atzipenak konposaketa lexikaleko erregelen sorkuntza dinamikoa eragingo du —horretarako *COMPOSITION-LEXICALE* fazetako informazioaz baliatuz—; halaber, definitutako konposaketa-erregela esplizituekiko arrazoiketari ere ekingo litzaioke (ikus § III.3.1). Estrategia honi esker, § III.3.1 azpiatalean azaldutako arrazoiketa sinonimikoa ere gauzatuko litzateke.

HIZTSUA kontsultatzerakoan desanbiguaketa dinamikoa martxan jarritz gero berriz, erabiltzaileak prest egon beharko du sistemarekin lankidetzan aritzeko, suertatuko diren desanbiguatze lanetan parte hartzeko alegia. Arrazoitze-prozesuaren garaiko lankidetzari hori sistemak kudeatuko du, sistema-erabiltzaile elkarrekintzarako moduluaren bitartez (ikus § III.3.4).

##### IV.2.2.2.1 DDEF: Definizio-galdea.

DDEF funtzioa HIZTSUAren *funtzionalitate esplikatzaile* bezala bururatu da. Hizkuntza hiztegi baten erabilerari begira funtsezko funtzioa da hau: kontzeptu baten definizioa eskatzea.

Gure ikuspuntutik, DDEFek "azalpen" desberdinak ematen ditu, aukeratutako *esplikazio maila* zein den. Esplikazio maila testuala ("textuel") aukeratuz gero, funtzioak adierari dagokion definizio-testua itzuliko du. Beste edozein kasutan baina, kontzeptu baten definizioa *kontzeptuak beste kontzeptu batzuekiko dituen erlazioen multzo* bezala hartzen da.

---

<sup>1</sup> Sistemaren aldagai orokor batez kontrolatua.

#### IV. kapitulua

Kontzeptu baten definizioaren ikuspegi hau, beste kontzeptuekiko dituen erlazioen multzoarena alegia, autore anitzek konpartitzen du. Txostenaren III. kapituluan aldatu dugu jada N. Calzolariren pasarte bat (1984a), non autore horrek aldarrikatzen baitzuen, erlazional tipoko errepresentazio batean, kontzeptu bati, bere esangura, elkarrerlazio lexikalen unibertso horretan daukan kokaguneak ematen diola. Orobat dio R. Alterman-ek (1985:154), ondokoa baieztan duela:

The dictionary may be thought as a semantic net, with each term gaining its relative meaning from its position in the net.

(Papegaaij *et al.*, 86)n kontzeptu-definizioak "menpekotasun-bikotez"<sup>1</sup> eraikitzen dira: bikote bakoitza bi lexemak osatzen dute, batak bestearekiko duen rola adierazten duen "erlatoze" batez lotuak.

Gurean berriz, kontzeptu baten "definizio erlazionala" lotura-konjuntzio batek emana dator (ikus sintaxia gorago), kontzeptu definituak THESAURUSeko beste kontzeptu batzuekin dituen loturen konjuntzioa alegia.

Lehenbiziko esplikazio mailatzat hartutako testuzko maila alde batera utzirik bada, DDEFek itzultzen duen emaitza ondokoa da, esplikazio maila erlazional desberdinen arabera:

- *Lokala* ("local"): Esplikazio maila honetan DDEFek, kontzeptuak bere definizioari esker edota besteren definiziotan duen partehartzeagatik (alderantzizko erlazioei esker) jasotako erlazioak itzultzen ditu.
- *Hereditua* ("hérité"): aurrekoaz gain, kontzeptuak sareko egitura taxonomikotik jasotako erlazio heredituak ere itzultzen ditu DDEFek.
- *Deduzitua* ("déduit"): maila honetan DDEFek, kontsulta garaian HEBko mekanismo inferentzialei esker deduzitutako erlazioak ere ematen ditu.

Lau maila hauek esplikazio maila "progresibo" bezala daude bururatuak, hau da, pentsatzekoa da sistemak emango duen esplikazioa gero eta aberatsagoa izango dela, esplikazio maila gora doan neurrian (hor goian ezarritako ordenaren arabera).

Hona hemen DDEF funtzio honen zehaztapena:

Datuak:        *C* kontzeptua ; *N* esplikazio-maila

Emaitzak:     *D* definizioa <sup>1</sup>

---

<sup>1</sup> "dependency pairs".

Sintaxia:

*definizio-galdea*= "DDEF" "(" thesauruseko-kontzeptua "," esplikazio-maila ","  
*definizioa*<sup>2</sup> ")".

Prozesua: C-ren "esanahia" kontsultatu HEBan.

Adibideak:

1. **Erabiltzailea.- DDEF** (|guêpe I 1|, textuel, ?D)

*Donne-moi la définition de guêpe (niveau d'explication "textuel").*<sup>3</sup>

**Sistema.-** D= 'insecte hyménoptère à aiguillon'

2. **E.- DDEF** (|guêpe I 1|, local, ?D)

*Donne-moi la définition de guêpe (niveau d'explication "local").*

S.-D= (and (|guêpe I 1| HYPERONYME |insecte I 1|)  
(|guêpe I 1| CARACTERISTIQUE |hyménoptère I 1|)  
(|guêpe I 1| POSSESSION |aiguillon I 1|)  
(|guêpe I 1| HYPONYME |frelon I 1|)  
(|guêpe I 1| POSSESSEUR |guêpier I 1|))

*La guêpe est un insecte hyménoptère qui possède un aiguillon. Un frelon est une guêpe.  
Un guêpier "possède" des guêpes.*

non ikus daitekeen nola, esplikazio maila lokalean ere, erabiltzaileak maila testualekoa baino aberatsagoa den azalpen bat jasoko lukeen, kasu honetan |frelon I 1| eta |guêpier I 1|<sup>4</sup> kontzeptuen frameetan ezarritako erlazioen alderantzizkoei esker.

<sup>1</sup> D **def**(C)-ren azpimultzo bat izango da, oro har. Esplikazio maila testuala delarik, definizioa karaktere-katea bat izango da.

<sup>2</sup> Azpimarratutako parametroak emaitza-parametroak dira.

<sup>3</sup> Hemendik aurrera, eskari kodetu bakoitzaren azpian eta sistemaren erantzunen azpian esaldi bat jarriko dugu frantsesez, espresio kodetuari hizkuntza arruntean legokiokeena edo adieraziz.

<sup>4</sup> **frelon I 1**: *grosse guêpe* (LPPL, 245. orr.).  
**guêpier I 1**: *nid de guêpes* (LPPL, 266. orr.).

3. E.- **DDEF** (|guêpe I 1|, hérité, ?D)

*Donne-moi la définition de guêpe (niveau d'explication "hérité").*

S.-D= (and (|guêpe I 1| HYPERONYME |insecte I 1|)  
 (|guêpe I 1| CARACTERISTIQUE |hyménoptère I 1|)  
 (|guêpe I 1| CARACTERISTIQUE |articuler I 1#m|)  
 (|guêpe I 1| POSSESSION |aiguillon I 1|)  
 (|guêpe I 1| POSSESSION |patte I 1#n|)  
 (|guêpe I 1| HYPONYME |frelon I 1|)  
 (|guêpe I 1| POSSESSEUR |guêpier I 1|))

*La guêpe est un insecte hyménoptère et articulé qui possède un aiguillon et six pattes. Un frelon est une guêpe. Un guêpier "possède" des guêpes.*

non |articuler I 1#m| |articuler I 1|en m-garren konfigurazioa litzatekeen, 'articulé' partizipioaren adierazpide, eta |patte I 1#n| |patte I 1|en n-garrena, 'six pattes' sintagma errepresentatzen duena. Bai *articulé* ezaugarria eta bai *guêpek six pattes* edukitzea ere bere hiperonimotik heredatutako egitateak edo erlazioak lirateke, alegia |insecte I 1|<sup>1</sup> kontzeptutik. Ez dira hor jarri, noski, kontzeptuak bere hiperonimoetatik hereda ditzakeen propietate guzti-guztiak.

4. E.- **DDEF** (|guêpe I 1|, déduit, ?D)

*Donne-moi la définition de guêpe (niveau d'explication "déduit").*

S.-D= (and (|guêpe I 1| HYPERONYME |insecte I 1|)  
 (|guêpe I 1| ANCETRE |animal I 2|)  
 (|guêpe I 1| ANCETRE |être II 1|)  
 (|guêpe I 1| CARACTERISTIQUE |hyménoptère I 1|)  
 (|guêpe I 1| CARACTERISTIQUE |articuler I 1#m|)  
 (|guêpe I 1| POSSESSION |aiguillon I 1|)  
 (|guêpe I 1| POSSESSION |patte I 1#n|)  
 (|guêpe I 1| POSSESSION |aile I 1#p|)  
 (|guêpe I 1| HYPONYME |frelon I 1|)  
 (|guêpe I 1| POSSESSEUR |guêpier I 1|))

*La guêpe est un être, animal, insecte hyménoptère et articulé qui possède un aiguillon, six pattes et des ailes membraneuses. Un frelon est une guêpe. Un guêpier "possède" des guêpes.*

non |aile I 1#p| |aile I 1|en p-garren konfigurazioa litzatekeen, 'ailes membraneuses' sintagmaren adierazpide. *guêpe* kontzeptuak *ailes membraneuses* edukitzea |hyménoptère I 1|en definiziotik deduzitutako egitatea litzateke (horretarako adjetibo gisa analizaturik). |animal I 2| hiperonimo "arbasoa" |insecte I 1|en definiziotik eratorria da, eta |être II 1| |animal I 2|renetik<sup>2</sup>.

<sup>1</sup> *insecte I 1: animal articulé à six pattes* (LPPL, 299. orr.).

<sup>2</sup> *hyménoptère I 1: ... à ailes membraneuses...* (LPPL, 280. orr.).

#### IV.2.2.2.2 **RDEF:** Definizio-birformulaketa.

Funtzio hau ere hizkuntz kontzeptuen ulerkuntzarako azalpen-tresna lagungarri bezala asmatua da. Funtzio sekundariotzat daukagu, beti ere definizio-galde baten ondoren (DDEF) erabiliko delako. Bere helburua da sistemak emandako *kontzeptu-definizio bat birformulatzea, eskaripean*.

Sistema balia daiteke definizio jakin baten esplikazio mailaz egin beharreko birformulaketa nola egin erabakitzeko. Horrela, zenbait kasutan, birformulaketa definizioaren esplikazio maila "igotzea" besterik ez da izango (aurreko azpiatalean azaldu dugun zentzuan). Beste batzuetan ordea, estrategia sofistikatuagorik erabili beharko da definizioa behar den bezala birformulatzeko.

Hona zehaztapena:

Datuak:  $D$  definizioa;  $N$  esplikazio-maila

Emaitzak:  $D'$  definizio-espresioa, non  $D' \cong D$

Sintaxia:

*definizio-birformulaketa* = "RDEF" "(" definizioa "," esplikazio-maila ","  
*definizio-birformulatua* ")".<sup>1</sup>

non definizio birformulatu bat definizio-espresio baten bitartez adierazten baita?<sup>2</sup>:

*definizio-birformulatua* = *definizio-espresioa*.

Prozesua eta estrategiak:

D' sortu, ondokoez baliatuz:

- a) Sinonimia: D-ko kontzepturik "inportanteenak" beren sinonimoekin ordeztu. Estrategia hau inplementatzeko konposaketa lexikalaz baliatzen gara (ikus § III.3.1en, arrazoiketa sinonimikoari buruzkoa).
- b) Hedapena: D-ko kontzeptu bat edo batzuen ordeztu beren definizioa erabili. Estreinako estrategia bezala, sistema saiaturiko da D-k definitzen duen kontzeptuaren hiperonimoa "hedatzen".

---

*animal I 2: être organisé, doué de mouvement et de sensibilité* (LPPL, 30. orr.).

<sup>1</sup> Funtzio honetan definizioa —datua— sistemak emandako definizio bat da, aurretiazko DDEF eskari batek, normalean. Beraz, bertan ez da aldagairik azalduko.

<sup>2</sup> Bereizten dira hemen DDEF batek emandako definizioak eta definizio birformulatuak. Definizio birformulatuok, sistemaren partetik birformulatze-prozesu bat nozitu dutenez, definizio-espresioen bitartez eratzen dira, erabiltzaileak, VDEFen eta, espresatutako "definizioen" sintaxi bera erabiliz.

#### IV. kapitulua

- c) Integrazioa: D-ren parte bat ordeztu —lotura-konjuntzioa— parte horrek definitzen duen kontzeptuarekin (komeni baldin bada).

#### Adibideak:

1. E.- **DDEF** (|*email* I 1|, local, ?D)

*Donne-moi la définition de email.*

S.-D= (and (|*email* I 1| HYPERONYME |*enduit* I 1|)  
(|*email* I 1| CARACTERISTIQUE |*vitriifiable* I 1|)  
(|*email* I 1| POSSESSEUR |*faïence* I 1|))<sup>1</sup>

*L'email est un **enduit vitriifiable** de la **faïence**.*

E.- **RDEF** ( (and (|*email* I 1| HYPERONYME |*enduit* I 1|)  
(|*email* I 1| CARACTERISTIQUE |*vitriifiable* I 1|)  
(|*email* I 1| POSSESSEUR |*faïence* I 1|)),  
local, ?DR)

*Reformule la définition précédente.*

S.-DR= (and (|*email* I 1| HYPERONYME |*substance* I 1|)  
(|*email* I 1| CARACTERISTIQUE |*liquide* I 2|)  
(|*email* I 1| CARACTERISTIQUE |*vitriifiable* I 1|)  
(|*email* I 1| POSSESSEUR |*faïence* I 1|))

*L'email est une **substance liquide vitriifiable** de la **faïence**.*

horretarako **enduit I 1**-en definizioa —*substance liquide étendue sur* (LPPL, 202. orr.)— kontsultatuz, eta |*enduit* I 1| kontzeptua partzialki hedatuz.

#### IV.2.2.2.3 DPRO: Kontzeptu baten propietateen galdea.

Galde-funtzio honen helburua *kontzeptu jakin baten atributu batek zer balio dituen* ezagutzea da. Hona hemen DPRO funtzioaren zehaztapena:

Datuak: C kontzeptua ; A atributua

Emaitzak: V atributu-balioa , non  $A \sigma C = V$

Sintaxia:

***propietate-galdea**= "DPRO" "(" thesauruseko-kontzeptua "," atributua ","  
atributuaren-balio-lista ")".*

non

---

<sup>1</sup> *email I 1: **enduit vitriifiable** de la **faïence** (LPPL, 194. orr.).*

*atributuaren-balio-lista* = "(" { "(" atributu-primitiboa "(" balio-lista" )" } )".

*balio-lista* = atributu-balioa { atributu-balioa }.

Ohargarria da sintaxi honek permititzen duela atributu gisa aldagai bat erabiltzea — kontzeptuari dagozkion atributu erlazional primitibo guztien eskari bezala interpretatuko da—, edo kontzeptu bat (ikus atributuaren sintaxia gorago). Azken kasu horretan, sistemak gauza izan behar du bere atributu primitiboen eta THESAURUSeko hizkuntz kontzeptuen arteko loturak bilatzeko.

#### Prozesua:

Kontua da **def(C)**-n —edota "antzeko" kontzeptu baten definizioan— A atributuaren "erreferentziaren" bat bilatzea, eta balioak eskuratzea.

#### Estrategiak:

Ondoko estrategiak bururatu ditugu kontzeptu baten propietateen bilaketa honetarako:

- a) **def(C)**-ra jo eta saiatu bertan aurkitzen A esplizituki duen loturaren bat.
- b) A atributuaren definiziora jo eta saiatu aurreko estrategia aplikatzen, A-ren sinonimo eta hiperonimoez baliatuz. Hori egiteko, A atributuaren eta THESAURUSeko kontzeptu baten arteko lotura bilatu behar da, eta horretarako dago, hain zuzen ere, *CONCEPT-TYPE-CORRESPONDANT* fazeta (ikus § III.3.2).
- c) C-ren hiperonimoetatik<sup>1</sup> heredatutako propietateak aztertu.
- d) **def(C)**-ko balioak aztertu eta saiatu A atributuarekiko bateragarriak litezkeenak aurkitzen.
- e) Arrazoiketa sinonimikoa: tratatu —aurreko estrategiak aplikatuz— C kontzeptuaren sinonimoak.

---

<sup>1</sup> Hiperonimoez, oro har, dihardugunean, bai berehalako hiperonimoak eta bai urrunekoak —edo "arbasoak"— ere hartzen ditugu aintzakotzat.

#### IV. kapitulua

##### Adibideak:

1. E.- **DPRO** (|ambre I 1|, |couleur I 1|, ?LV)

*De quelle couleur est l'ambre?*

S.-LV= ((CHARACTERISTIQUE (|jaune I 1|)))<sup>1</sup>

*Jaune.*

2. E.- **DPRO** (|ambre I 1|, ?P, ?LV)

*Donne-moi les propriétés de l'ambre.*

S.-LV= ( (HYPERONYME (|résine I 1|))

(CHARACTERISTIQUE (|fossile I 1| |jaune I 1| |translucide I 1|)))

*Il s'agit d'une résine, fossile, jaune et translucide.*

3. E.- **DPRO** (|grand I 1|, SYNONYMES, ?LV)

*Donne-moi des synonymes de grand.*

S.-LV= ((SYNONYMES (|gros I 1| |énorme I 1|)))

*Gros, énorme.*

4. E.- **DPRO** (|brûler I 1|, |agent I 1|, ?LV)

*Quel est l'agent de l'action de brûler?*

S.-LV= ((AGENT (|feu I 1|)))

*Le feu.*

#### IV.2.2.2.4 DRAP: Bi kontzepturen arteko elkarbidearen bilaketa.

Funtzio honen helburua *bi kontzepturen arteko elkarbidea, erlazioa, bilatzea* da. Gurean, elkarbidetzat, *bi kontzeptuak THESAURUS sarean lotzen dituen bidea* hartuko dugu. Jakina, *bi kontzepturen arteko elkarbidea*, okerreanean, *bi kontzeptuak elkarlotzen dituen edozein bide* izango da, baina, interesgarria zera da, alegia sistema gai izatea bide posible batzuen artean hautatzeko. Eta horretarako zenbait estrategia definitu eta ezarri beharko da.

Hau da funtzio honen zehaztapena:

Datuak:  $C_1, C_2$  kontzeptua

Emaitzak:  $R$  elkarbide-espresioa, non  $R$ -k  $C_1$  eta  $C_2$ -ren arteko bide bat denotatzen baitu, batzuetan murriztapen-espresio batez lagundurik etor daitekeena.

---

<sup>1</sup> Sistemak ez dauka, kontzeptu baten atributu primitiboen artean kolorea adierazten duenik, baina gai izan behar du horrelako galderei erantzun bat emateko (beharrezko ezagutza baldin badu, jakina).



Sintaxia:

*elkarbide-bilaketa* = "DRAP" (" thesauruseko-kontzeptua ", " thesauruseko-kontzeptua ", " elkarbide-espresioa )".

Prozesua eta estrategiak:

Sare semantikoa C<sub>1</sub> eta C<sub>2</sub>-tik abiatzen den —eta biak elkarlotzen dituen— bide bat bilatu beharra dago<sup>1</sup>. Horrelako bilaketak kostu handikoak izan ohi dira (inplementazio eraginkor bati begira) eta, beraz, beharrezko izango da bilaketa-espazioa murriztuko duen heuristikorik ezartzea. Heuristiko horiek erlazioen analisi eraginkorra egiteko gaitasuna izan behar dute, batik bat erlazio taxonomiko eta sinonimikoena. Kontuan hartzekoa litzateke hemen erabiltzailearen parte hartzea, bilaketa-prozesu horretan gidari gisa.

Adibideak:

1. E.- **DRAP** (|guêpe I 1|, |miel I 1|, ?R)  
*Quel rapport y a-t-il entre guêpe et miel?*

S.-R= ((and (and (|guêpe I 1| HYPERONYME |insecte I 1|)  
(|insecte I 1| HYPONYME |abeille I 1|)  
(|abeille I 1| PREDICATION |produire I 1#n|)  
(|produire I 1#n| THEME |miel I 1|))  
(and (|abeille I 1| CARACTERISTIQUE |hyménoptère I 1|)  
(|guêpe I 1| CARACTERISTIQUE |hyménoptère I 1|))))

*Le miel est produit par l'abeille, qui est un insecte hyménoptère comme la guêpe.*

non |produire I 1#n| |produire I 1| en n-garren konfigurazioa baita, 'produire le miel' aditz-sintagmaren adierazpide.

IV.2.2.2.5 DDIF: Bi kontzepturen arteko diferentzien galdea.

Emandako *bi kontzepturen arteko aldeak edo diferentziak bilatzea* da funtzio honen xedea.

Ez da txantxetako definitzea zer den bi kontzeptu bereizten dituen, horretarako berorien hiztegi-definizioez baino ez baliatuz. Lehen hurbilketa batean, kontzeptu arteko diferentzi eskaria taxonomia bereko diren kontzeptuen artean baino ezingo da egin (ikus hemen beheko aurrebaldintza). Horrela bada, diferentzia bi kontzeptuen definizioetan komuna ez den parteari esango diogu, amankomuneko zerbait egongo dela suposatuz, biak taxonomia berekoak izaki.

---

<sup>1</sup> Funtzio honek, normala denez, **chemin** eragiketa primitiboa du oinarri.

#### IV. kapitulua

Hortaz, funtzio honen zehaztapena ondokoa da:

Datuak:  $C_1, C_2$  kontzeptua

Emaitzak:  $D$  *diferentzi-espresioa*, non  $\mathbf{def}(C_1)-D_1 \cong \mathbf{def}(C_2)-D_2$ <sup>1</sup>

Aurrebaldintza:  $C_1 \Delta C_2$

Sintaxia:

*diferentzi-galdea*= "DDIF" ("thesauruseko-kontzeptua ","  
thesauruseko-kontzeptua "," *diferentzi-espresioa* ").

Prozesua eta estrategiak:

$C_1$  eta  $C_2$  taxonomia berekoak direla emanik, biek amankomunean duten hiperonimotik beherantz aztertu beharko dira diferentziak, beren definizioetako zati "komuna" determinatuz, gainerakoa "diferentzia" bezala hartuko baita.

Pentsa liteke funtzio honetaz baliatzea sinonimia partzialeko kasuetan ere, hots, adiera sinonimoak dituzten bi hitzen arteko diferentziak eskatzen direlarik. Adiera ez-sinonimikoen zati komuna aztertu beharko litzateke kasu horretan, diferentziak bilatzearen.

Adibideak:

1. E.- **DDIF** (|bicyclette I 1|, |tricycle I 1|, ?DIF)

*Donne-moi les différences existant entre une **bicyclette** et un **tricycle**.*

S.-DIF= ( (and(and (|bicyclette I 1| POSSESSION |roue I 1#n|)  
(|tricycle I 1| POSSESSION |roue I 1#m|))))

*La **bicyclette** possède **deux roues** tandis que le **tricycle** possède **trois roues**.*

non |roue I 1#n| eta |roue I 1#m| '*deux roues*' eta '*trois roues*' sintagmen adierazpide diren konfigurazioak baitira, hurrenez hurren.

2. E.- **DDIF** (|congre I 1|, |perche I 1|, ?DIF)

*Donne-moi les différences qu'il y a entre un **congre** et une **perche**.*

S.-DIF= ( (and(and (|congre I 1| LOCATIF |mer I 1|)  
(|perche I 1| LOCATIF |eau I 2#n|))))

*Le **congre** est un **poisson** de **mer** tandis que la **perche** l'est d'**eau douce**.*

---

<sup>1</sup> '1' eragilea hemen kenketakoa da, definizioak biltzen dituen loturen eta diferentzi-espresioen arteko multzo-kendura adieraziz:  $D_1$ -ek  $D$ -ko lotura-bikoteetako "aurreneko" loturen multzoa denotatzen du,  $C_1$  kontzeptuari dagokiona, eta  $D_2$ -k  $C_2$ -ren loturen multzoa.

non |eau I 2#n| |eau I 2|ren n-garren konfigurazioa den, 'eau douce' sintagma errepresentatzen duena.

3. E.- **DDIF** (|laver I 1|, |balayer I 1|, ?DIF)  
*Quelle est la différence entre laver et balayer?*

S.-DIF= ( (and(and (|laver I 1| INSTRUMENTAL |liquide I 1|)  
(|balayer I 1| INSTRUMENTAL |balai I 1|))))  
*On lave avec un liquide, mais on balaye avec un balai.*

#### IV.2.2.2.6 **RTHS: Kontzeptu-bilaketa tesaurikoa.**

Funtzio honen helburua *murriztapen sail bat betetzen duten kontzeptuen multzoa lortzea* da Hauxe litzateke HIZTSUAren bertsio honetarako asmatutako funtzioen artean ahalmentsuena. Funtzio honek ematen dio erabiltzaileari ideiatik kontzepturako ibilbidea — edota definiziotik kontzepturakoa— egiteko aukera, "ideia" hori espresatzeko murriztapen-espresioaren sintaxia eskainiz.

Datuak: R murriztapen-espresioa

Emaitzak: LC= <C<sub>1</sub>, C<sub>2</sub>, ..., C<sub>n</sub>>, kontzeptu-lista, non edozein C<sub>i</sub>-rentzat — C<sub>i</sub> kontzeptua— beteko baita: C<sub>i</sub> | R (C<sub>i</sub>-k betetzen ditu R-k adierazi murriztapenak).

Sintaxia:

*bilaketa-tesaurikoa*= "RTHS" "(" murriztapen-espresioa ","  
emaitza-kontzeptuaren-aldagaia "," kontzeptu-lista ")".

non

*kontzeptu-lista*= "(" {thesauruseko-kontzeptua} ")".

*emaitza-kontzeptuaren-aldagaia*= aldagaia.

Emaitza-kontzeptuaren aldagaia murriztapen-espresioan derrigor egongo den aldagai baten adierazpena da. Hori baliagarria da murriztapen-espresioan aldagai bat baino gehiago agertzen diren kasuan<sup>1</sup>, emaitzako kontzeptu-lista osatzeko zein aldagaiz baliatuko garen adierazten baitigu.

<sup>1</sup> Aldagai anitzeko murriztapen-espresioak PROLOG eskarien moduan interpretatuko dira, hau da, aldagaiak balio desberdinekin instantziatzeko —kontzeptuekin jeneralean— baterakuntza-printzipioari jarraituko zaio.

Prozesua eta estrategiak:

Behin R-ko lehen murriztapena betetzen duten kontzeptu multzoa zehaztu eta gero, aurrera jarraitu beharko da, multzo hori gero eta gehiago mugatuz R-ko murriztapen desberdinen arabera. Esan beharrik ere ez dago, azkenean multzoa hutsa izango dela, R murriztapen sail osoa betetzen duen kontzepturik aurkitzen ez bada.

Ohargarria da R murriztapen-espresioa definizio-espresio bat izaten ahal dela, definizioa murriztapen multzo bezala harturik, eta kasu horretan, RTHS funtzioa "definizioetik" abiatuta kontzepturantz garamatzen funtzioa bihurtzen da (ikus hemen behean 4. adibidea).

§ IV.2.2.5 azpiatalean komentatuko ditugu zehazkiago funtzionalitate honen inplementazioan erabili ditugun estrategiak.

Adibideak:

1. E.- **RTHS** ((and (?X HYPERONYME |poisson I 1|)), ?X, ?LC)  
*Donne-moi les noms des différents poissons.*

S.-LC= (|anguille I 1| |bar I 2| |barbu I 2| |brochet I 1| |congre I 1| . . .)<sup>1</sup>  
*Anguille, bar, barbu, brochet, congre, ...*

2. E.- **RTHS** ((and (?X HYPERONYME |instrument I 1|)  
(?X OBJECTIF |mesurer I 1|)),  
?X, ?LC)  
*Donne-moi les noms des instruments utilisés pour mesurer.*

S.-LC= (|baromètre I 1| |dynamomètre I 1| |télémètre I 1| . . .)  
*Baromètre, dynamomètre, télémètre, ...*

3. E.- **RTHS** ( (and (|chat I 1| POSSESSION ?Y//|yeux I 1|)  
(?Y |couleur I 1| ?X)),  
?X, ?LC)  
*De quelle couleur sont les yeux du chat?*

non ?Y//|yeux I 1| bezalako aldagai murriztu baten bitartez (ikus sintaxia gorago), eskaria "zehatzagotu" egin baitaiteke, aldagai hori |yeux I 1|en konfigurazio batekin —taxonomikoki kontzeptu horren menpeko den batekin eta ez edozein kontzepturekin— instantziazera behartuz.

---

<sup>1</sup> Egungo inplementazioan, funtzioak itzultzen ditu murriztapen-espresioan agertzen diren aldagai guztien instantziazioak (aldagaiak baldin bada), emaitza-kontzeptuen listaz gain noski. Adibideotan ordea, ez ditugu instantziazio horiek azalduko, funtzioaren emaitza alferrik ez nahastearren.

4. E.- **RTHS** ((and (?X HYPERONYME |consumer I 1|)  
(?X AGENT |feu I 1|)),  
?X, ?LC)

*Donne-moi le mot pour dire "consumer par le feu".*

S.-LC= (|brûler I 1|)

*Brûler.*

#### IV.2.2.3 Egiaztapen-funtzioak.

HIZTSUArentzat sortutako bigarren funtzionalitate multzoa egiaztapen funtzioek osatzen dute. Funtzio horien xedea da, erabiltzaileari, hizkuntz kontzeptuen gainean duen ezagutzaren zuzentasuna egiaztatzeko, tresna espezifikorik ematea.

H. R. Webber-ek, bere hitzaldiaren laburpenean dioen bezala (1984:463) “. . . *a central function of any dictionary is in fact data verification*”. Gure ustez, erabiltzaileak *hizkuntzari buruz duen ezagutzaren egiaztapena* hiztegi batek eskaintzen duen —eta eskaini behar lukeen— funtzionalitatearen funtsezko aspektua da. Funtzionalitate hau erakusten duen kasurik "sinpleena" hitzen ortografi egiaztapenarena dugu, askotan jotzen baita hiztegi hitz bat nola idatzi behar den jakiteko; funtzionalitate hori testu-tratamenduaren arloan ere ezaguna da aspalditxo. Gurean, noski, hitzen semantikaren inguruko funtzionalitateak diseinatu ditugu.

Hona hemen egiaztapen-funtzio horiek zeintzuk diren: VDEF, VPRO eta VRAP.

Hasi baino lehen, esan dezagun funtzio hauen inplementazioan ere kontuan hartzekoak direla galde-funtzioei buruz hitz egin dugunean azaldu ditugun estrategia orokorrak (ikus § IV.2.2.2).

##### IV.2.2.3.1 VDEF: Definizioaren egiaztapena.

Funtzio multzo honetako lehenengoa definizio baten egiaztapena egiten duena da, erabiltzaileari, kontzeptu bati buruz duen "definizio" ustezkoa zuzena den ala ez ikusteko aukera ematen diona.

Datuak:            *C* kontzeptua ; *D* definizio-espresioa

Emaitzak:        *true* bsb **def**(C) ≅ D (funtzio honek **def**(C) ere itzuliko du)

Sintaxia:

*definizio-erlazionalaren-egiaztapena* = "VDEF" "(" *thesauruseko-kontzeptua* ","  
*definizio-espresioa* "," *egiaztapenaren-emaitza* "," *definizioa* ")".

#### IV. kapitulua

non

*egiaztapenaren-emaiza = boolearra ⊥.*

#### Prozesua eta estrategiak:

Kontua da erabiltzaileak kontzeptuaren definiziotzat emandakoa ametitzeko modukoa den ala ez ikustea. Bistakoa da ez dagoela espero izaterik erabiltzaileak emandako definizio hori eta sistemak daukana berdinak izatea. Beraz, sistemak gai izan behar du jasotako definizioa onargarria den ala ez determinatzeko. Horretarako, ondoko estrategiak zehaztu ditugu:

- a) **def(C)** eta D-ren arteko hurbiltasuna aztertu, hurbiltasun semantikoko eragiketaz baliatuz (§ IV.2.1).
- b) D' definizioak sortu, non  $D' \cong \mathbf{def}(C)$ , eta D' horien eta D-ren arteko hurbiltasuna egiaztatu. Definizio-sorkuntza hori burutzeko, **def(C)** eragiketa oinarritzat hartu eta estrategia sinonimiko eta taxonomikoak aplikatuko zaizkie bertako osagaiei.

#### Adibideak:

1. E.- **VDEF** ( |cantonner I 1|,  
(and (|cantonner I 1| HYPERONYME |établir I 1|)  
(|cantonner I 1| THEME |soldat I 1|)  
(|cantonner I 1| LOCATIF |ville I 1|))))

*Est-ce que **cantonner** signifie **établir** des **soldats** dans une **ville**?*

S.-True,

def(|cantonner I 1|) =  
(and (|cantonner I 1| HYPERONYME |installer I 1|)  
(|cantonner I 1| HYPERONYME |placer I 1|)  
(|cantonner I 1| HYPERONYME |établir I 1|)  
(|cantonner I 1| THEME |troupe I 3|)  
(|cantonner I 1| LOCATIF |localité I 1|))

*Oui, **cantonner** est **installer** (**placer**, **établir**) des **troupes** dans une **localité**.*

non |cantonner I 1| kontzeptuaren definizio erlazionalak islatzen ditu, besteak beste, bere jatorrizko erlazioak (lokalak) |installer I 1|, |troupe I 3| eta |localité I 1| kontzeptuekiko<sup>1</sup>.

---

<sup>1</sup> *cantonner I 1: installer des troupes dans une localité (LPPL, 89. orr.).*

IV.2.2.3.2 VPRO: Kontzeptu baten propietateen egiaztapena.

Kontzeptu baten propietateei buruz erabiltzaileak dakiena egiaztatzen laguntzeko pentsatutako funtzioa dugu hau.

Datuak:  $C$  kontzeptua ;  $L$  atributu-balio-bikote-lista

non  $L \langle A, V \rangle$  bikote lista den,  $A$  atributua eta  $V$  atributu-balioa

Emaitzak: *true* bsb  $L$ -ko bikote guztietan  $A \sigma C = V$  egiazkoa bada.

Sintaxia:

*propietateen-egiaztapena* = "VPRO" "(" thesauruseko-kontzeptua ","  
atributu-balio-bikote-lista "," egiaztapenaren-emaitza ")".

non

*atributu-balio-bikote-lista* = "(" "(" atributua atributu-balioa ")"  
{ "(" atributua<sup>1</sup> atributu-balioa " } ")".

Prozesua:

Listako bikote bakoitzeko  $A$  atributua bilatu behar da  $C$ -n, zuzenean ala § IV.2.2.2.3 azpiatalean aipatutako estrategiak erabiliz (sinonimia, atributuari lotutako kontzeptua, etc.), eta  $A \sigma C = V$  egiaztatu.

Estrategiak:

Estrategiak propietate-galdean erabilitakoen antzekoak izango dira, egiaztapen-funtzio honi behar den bezala egokituak.

Adibideak:

1. E.- **VPRO** (|or I 1|, ((|couleur I 1| |rouge I 2|)))

*Est-ce que la couleur de l'or est rouge?*

S.-False

---

*troupe I 3: réunion de soldats ...* (LPPL, 562. orr.).

*installer I 1: placer, établir* (LPPL, 300. orr.).

*localité I 1: petite ville, village...* (LPPL, 326. orr.).

<sup>1</sup> Atributuaren espresio gisa ametitzen da aldagai bat erabiltzea, "halako balioa ote da kontzeptu jakin baten edozein atributuaren balio?" moduko galderak egiteko aukera emanez.

#### IV. kapitulua

2. E.- **VPRO** (|grand I 1|, ((SYNONYMES |gros I 1|)))  
*Est-ce que **grand** et **gros** sont des synonymes?*

S.-True

3. E.- **VPRO** (|souris I 1|, ((|classe I 1| |rongeur I 2|)))  
*Est-ce que la **souris** est un **rongeur**?*

S.-True

4. E.- **VPRO** (|verser I 1|, ((THEME |liquide I 1|)))  
*Est-ce qu'on **verse** des **liquides**?*

S.-True<sup>1</sup>

#### IV.2.2.3.3 **VRAP**: Bi kontzepturen arteko elkarbidearen egiaztapena.

Hirugarren egiaztapen-funtzio honek —VRAP— *bi kontzepturen arteko balizko elkarbidea edo erlazioa egiaztatzea* du helburu. Erlazio hori adierazteko, elkarbide-espresioaren sintaxia dago erabiltzailearen eskumenean.

Datuak:  $C_1, C_2$  kontzeptua ;  $R$  elkarbide-espresioa

Emaitzak: *true* bsb  $C_1$  eta  $C_2$ -ren artean bide bat badago,  $R$  elkarbide-espresioaz adieraz daitekeena (funtzio honek itzuliko du **chemin**( $C_1, C_2$ ) ere)

Sintaxia:

*elkarbide-egiaztapena* = "VRAP" "(" *thesauruseko-kontzeptua* ","  
*thesauruseko-kontzeptua* "," *elkarbide-espresioa* <sup>2</sup> ","  
*egiaztapenaren-emaitza* "," *elkarbide-espresioa* ")"<sup>3</sup>.

Prozesua eta estrategiak:

DRAP elkarbide-bilaketakoen antzekoak; hargatik, kasu honetan sareko bilaketa "gidatua" izatearen abantaila dago, datu bezala, emandako  $R$  espresioak gidatua, hain zuzen.

---

<sup>1</sup> *verser I 1*: *passer un liquide dans un récipient* (LPPL, 577. orr.).

<sup>2</sup> Elkarbide-espresio bezala aldagai bat ere onartzen da (horrek eskari hau DRAP eskaria bihurtuko luke).

<sup>3</sup> Orobat, kontzeptuarteko elkarbide hauen espresioan ametitzen da aldagaien erabilera, RTHSn egiten denaren antzera edo. Hau da, elkarbideak adieraztean aldagi-identifikadore desberdinak erabil daitezke, baterakuntza-printzipioaren arabera instantziatuko direnak (PROLOGen moduan).



Adibideak:

1. E.- **VRAP** (|abeille I 1|, |miel I 1|,  
(and (and (|abeille I 1| PREDICATION ?X//|produire I 1|)  
(?X//|produire I 1| THEME |miel I 1|))))

*Est-ce que le rapport entre **abeille** et **miel** est que l'abeille **produit** du miel?*

S.-True,

chemin (|abeille I 1|, |miel I 1|) =  
(and (|abeille I 1| PREDICATION |produire I 1#n|)  
(|produire I 1#n| THEME |miel I 1|))

*Oui, l'abeille **produit** du miel.*

non ?X//|produire I 1| aldagai murriztuaren bitartez adierazten ahal den |produire I 1|en menpeko konfigurazio bat, kasu honetan berorren n-garrena litzatekeen |produire I 1#n|, '*produire le miel*' aditz-sintagmaren errepresentazioa.

2. E.- **VRAP** (|abeille I 1|, |miel I 1|, |produire I 1|)

S.-True,

chemin (|abeille I 1|, |miel I 1|) =  
(and (|abeille I 1| PREDICATION |produire I 1#n|)  
(|produire I 1#n| THEME |miel I 1|))

aurreko eskari bera egiteko beste modu bat, non kontzeptu baten espresioa erabili den —|produire I 1|— elkarbide-espresio bezala.

#### IV.2.2.4 Oinarrizko beste zenbait funtzio.

Kapituluaaren hasieran aipatu funtzionalitateari buruzko ikersaio enpirikoetatik sortutako funtzio pare bat komentatu nahi genuke orain, nahiz eta proiektuan ez ditugula garatu esan behar. Funtzio horiek, eta dudarik gabe beste hainbat, definitu eta gehitu beharko litzatezkieke benetako hiztegi-sistema urgazle adimendun bati. Parafraseaketarako laguntza eta hautapen-murriztapenen inguruko funtzioak dira.

- *Parafraseaketarako laguntza.* Funtzionalitate hau funtsezkoa iruditzen zaigu hiztegi-sistema urgazle batentzat eta bere xedea erabiltzaileak emandako esaldi bat birformulatzen laguntzea litzateke. RDEFen antza duen funtzioa, nahiz eta kasu honetan birformulatu beharrekoa ez izan derrigor definizio bat. Benetako parafraseatze-laguntza eman ahal izateko, sistemak eduki behar luke informazio morfosintaktiko aberatsagorik, hitzen kokamenduari buruzko ezagutzarik, eta abar. Funtzio lexikalei buruzko ezagutza ere ("lexical functions", Melçûk, 1982; 84), hitzen arteko erlazio lexikal "tipikoak" errepresentatzen dituztenak, interesgarria

#### IV. kapitulua

litzateke parafraseaketaren testuinguru honetan. Eta badirudi esaldi laburrak analizatu eta sortzeko sistema bat ere beharrezko litzatekeela.

- *Hautapen-murriztapenen galde eta egiaztapena* ("selectional restrictions"). Funtzionalitate hau, nahiz eta lengoaia naturalaren tratamendura zuzendua izan batik bat (Calzolari, 90), interes handikoa gerta liteke hiztegi-sistema batentzat, idazketarako laguntza emateko garaian: adjetibo bat zein izen klaseri aplikatzen dakioken jakiteko, aditz batek har litzakeen argumentuen mota ezagutzeko, eta abar.

#### IV.2.2.5 RTHS funtzioaren implementazioari buruzko zenbait ohar.

Gorago azaldutako RTHS funtzioaren implementazioa dela eta zenbait ohar egin nahi genuke hemen. Ondokoan bada, funtzio honen oinarritzko algoritmoari buruzko aspektu batzuk, erabilizko estrategia orokorrak, eta abar, hartuko ditugu ahotan.

**RTHS** funtzioaren implementazioa, egungo bertsioan, KEEK eskaintzen duen eskari-funtzio nagusian oinarritzen da: QUERY. Ezagutza-basetik egitateak eskuratzeko pentsatua den QUERY horren funtzionamenduari begiratzen badiogu, QUERY eskari bat PROLOG galdera baten antzekoa dela ikusten da: izan ere, sarreran ematen zaion formula egitate-basean egiazkoa dela frogatzen saiatzen da. Frogapen horretarako PROLOGen antzeko baterakuntza-mekanismo batez baliatuko da. Erantzuna NIL izango da formulak errepresentatu egitatea basean egiazkoa dena frogatu ezin baldin bada, eta T —edo aldagai libreen instantziak— bestela. Egitate hori frogatzera iritsi ahal izateko, eskaria egitean adieraz dakioken sistemari erregela bidezko arrazoiketa bati ekiteko —atzerantzko kateamenduan, noski—, arrazoiketa horretan parte hartuko duten erregelen klase edo klaseak zehaztuz. Funtzio honek duen beste aspektu interesgarri bat, exekutatzeko ari dela erabilizko galderak egiteko aukera da, egitate baten egiazkotasuna baieztatzeko eskatuz edota, baita ere, aldagaidun formula bat egiazko bihurtuko luketen balioak zehazteko aukera emanez.

RTHS eskarien kasuan, suposatzen da beti egongo dela, murriztapen-espresioan, gutxienez aldagai ez-instantziatu bat, emaitza-kontzeptuaren aldagaia hain zuzen ere (horrexen instantziak osatuko dute RTHSren emaitza den kontzeptu-lista).

Funtzioaren oinarritzko algoritmoak ondoko aspektuak ditu kontuan:

- Zein dedukzio-estrategia gaitu: konposaketa lexikala, desanbiguaketa dinamikoa (erabiltzaileak lagundua), eta abar<sup>1</sup>.
- Murriztapenen ordenaketa. Ondoko irizpideon arabera egiten da ordenatze hori:
  - 1) Egingarritasuna: murriztapenak behar bezala ordenatzen ez baldin badira, gerta liteke KEE eskaria egokiro egin ezina. Adibidez, murriztapen-espresioko lehen hirukotean bi aldagai egotea, lehen eta bigarren posizioetan (sorburu-kontzeptua eta atributuaren lekuetan, lotura batean), nahikoa da KEEk eskaria ez onartzeko<sup>2</sup>; gure soluzioa, kasu horretan, hirukoteari buelta ematea da, alderantzizko erlazioaz baliatuz. Esan beharrik ez dago, egindako buelta emate horien arrastorik gorde egin beharko dela, erantzuna eman aurretik atzera berriz lehengora itzuli ahal izateko.
  - 2) Eraginkortasuna: murriztapen-espresioan aldagai kopuru handia egoteak kostu handia erants liezaioke bilaketa tesauriko baten erantzuna lortzeko prozesuari. Baina posible da, murriztapenok egokiro ordenatuz gero, askoz ere eraginkorragoa izango den eskaria lortzea. Egungo inplementazioan, murriztapenak ordenatzeko irizpideak hauek dira: (a) aldagai libreen kopurua —beti ere kontuan hartuz ordenatze horrexen menpeko dela, hain zuzen, kopuru hori gutxiagotzeko aukera—, eta (b) lotura-espresioetan aldagai libreek duten kokalekua (atributua aldagai duen espresioa, aldagaia sorburu-kontzeptuaren lekuan duena baino nahiago da beti). Eskariaren optimizazio-eragiketa bat da hau, azken finean.

Optimizazio honen inguruko beste aspektu bat, RTHSren egungo bertsioan inplementatzeko dagoena, kontzeptu batetik "abiatzen" diren erlazioen kopurua kontuan hartzea litzateke, kontzeptua instantziaturik baina atributuaren lekuan aldagaia duten lotura-espresioetan erabilgarri. Agi danean, erlazio kopururik txikiena dutenak handiagoa dutenen aurretik ezarriz, lor daiteke eskari eraginkorrago bat (aldagai batzuen instantziazioak alde aurretik mugatuko baitira).

---

<sup>1</sup> Oraingoz konposaketa lexikalaren gaitzea edo desgaitzea besterik ez da kontrolatzen (sistemaren aldagai orokor batez baliatuz).

<sup>2</sup> Hori gertatzeko, edo hiru aldagai libre daukan formula da (murriztapen zentzugabea, zentzugaberik bada!) edo sorburu-kontzeptuari eta atributuari dagozkion aldagaiak instantziazteke dituen da. Jakina, problema ez da esistituko, aurreko murriztapenen bat bi aldagai horietako bat alde aurretik instantziazteko gauza baldin bada.

#### IV. kapitulua

- QUERY egokia formulatu eta abiarazi: horretarako, egokiro ordenatutako murriztapen-espresioa KEEren eskari-lengoaia formula ongi eratua bihurtu behar da.
- Erantzunaren eraketa: RTHSren erantzunean bi zati ageri dira. Lehenengoa murriztapenak betetzen dituzten kontzeptuen azpilista da (emaitza-lista); bigarrena beste azpilista bat da, murriztapen-espresioan agertutako aldagai bakoitzak izan dituen instantziak biltzen dituena. Emaitzaren bigarren zati hori interesgarria da, batez ere, funtzio hau beste batzuekin konposatzeko aukerari dagokionean.

### IV.2.3 Oinarrizko funtzioen konposagarritasuna. Kontsulta- eta eskari-lengoaia.

Oinarrizko funtzionalitatearen "konposagarritasuna" diogularik, eskariak formulatzerakoan funtzio horiek konbinatzeko erabiltzaileari ematen zaion aukera esan nahi dugu. Funtzio desberdinak eskari konplexu bakar batean konbinatzeko posibilitateak zabaldu egin behar ditu, ongi zabaldu ere, sistemaren gaitasunen aukerak, hiztegi-sistemaren kontsulta iharduera kualitatiboki hobetuz.

Kontua da zehaztea zeintzuk diren funtzio-konposaketarako eskaini behar diren egiturak eta bideak, erabiltzaileak izango duen espresibotasuna benetan handiagoa izan dadin. Asma liteke *hiztegi-kontsultarako lengoaia* moduko bat, malgua eta ahalmentsua, definitu berri ditugun oinarrizko funtziook funtzio primitibotzat hartu, eta, zenbait kontrol-egituraren bitartez, funtzio horiek konbinatzeko bideak eskainiko lituzkeena. Agintzaile erako lengoaia batean pentsatzen badugu, baldintzazko egiturak, iteratiboak, eta abar izango genituzke. Pentsa liteke baita ere datu-baseetako galdeketa-lengoaieen moduko zerbaitetan (SQL, adibidez), non lengoaia mota horretako egitura tipikoak izango baikenituzke konposaketarako egiturak.

(Neff *et al.*, 88; Neff & Boguraev, 89; Byrd, 90)en, autoreek LQL izeneko galdeketa lexikaleko lengoaia deskribatzen dute. (Neff *et al.*, 88)n diotenaren arabera, datu-base lexikal baten (Lexical Database, LDB) galdeketa-lengoaia bezala sortutako honek, ondoko baldintzak bete behar lituzke: (1) galdeketa mota desberdin ugari egiteko aukera eman, (2) erantzunak —pantailan bistartzeko zein beste fitxategi batzuetara igortzeko— formateatzeko bideak eskaini, eta (3) erabiltzaileari LDBan aldaketak egiteko eragiketak zehazten permititu. LQLz idatzitako programen estiloaz berriz, "naturalak eta oroiterrazak" izan behar dutela diote.

Hala, LQLk aukera ematen dio erabiltzaileari, datu-base lexikaletik eskuratu nahi duen informazioa zein den zehatz-mehatz adierazteko. Datu-base lexikala DAM (Dictionary

Access Method, ikus § II.1) atzipen-metodoaren pean erregistratua denez, LQLz adierazitako sarreraren bilaketan zehazten ahal dira hierarkiako nodoei buruzko hainbat eta hainbat baldintza<sup>1</sup>. Hortaz, LQL erabiliz eska litezke, konparazio batera, halako aditz-adiera edota halako hitz konkretua definizioan duten sarrerek (Byrd, 90). Gainera, erabiltzaileak kontrola dezake erantzunean zein atributu azaltzea nahi duen.

Ikus daitekeenez, LQL bereziki egituratutako datu-base lexikal baten galdeketa-lengoaia berezia da. Horrela bada, interesgarria da autoreek halako lengoaia batentzat egokitzen jotzen dituzten ezaugarriak eta eskakizunak zein diren ikustea, baina, bistan dago, gure helburua oso bestelakoa dela. Ez da ahaztu behar gure HEBan ezagutza-egituraketa kontzeptuarteko erlazio lexiko-semanticoen mailakoa dela eta, beraz, hiztegi-kontsultarako lengoaia diogunean, edota programazio-lengoaia, maila lexiko-semanticoko horretan eskari konplexuak adierazteko bideak eskainiko duen lengoaiak ari garela. Hori dela eta ezarri ditugu, maila erlazional horretan, definitu ditugun oinarrizko funtzioak ere.

Hiztegi-kontsulta —edo programazio— lengoaia honen bila bultzatzen gaituena arlo kognitibotik dator, nagusiki: hiztegi-erabiltzaileak, kontsulta bat egin behar duelarik, buruan darabilen "programa" espresatzeko gauza izango litzatekeen lengoaia da nahi genukeena. Hiztegi-sistema adimendun batek gai izan behar bailuke, erabiltzailearen "programa mental" horretatik abiatuz lengoaia horretan idatzitako eskaria zehazteko, erantzun egokia ematekotan. Ez dugu, noski, eskari moldatze horretatik kanpo uzten erabiltzaile eta sistemaren arteko elkarrekintza baten posibilitatea, zeinetan sistemak, erabiltzaileari, bere eskari-programa hori zehazteko laguntza emango bailioke.

#### **IV.2.3.1 Funtzio-konposaketa HIZTSUAren baitan.**

Oinarrizko funtzioen konposaketari gagozkiola, proiektuan funtzio horiek konbinatzeko zenbait bide zehaztu dugu, nahiz eta sinpleak izan oso.

Ondokoan komentatuko ditugu oinarrizko funtzionalitate batek duen espresibotasunari buruzko zenbait aspektu, eta funtzio desberdinen artean permititzen ditugun konbinazio aukerak:

- a) RTHS eskarien espresibotasuna;
- b) eskari-konjuntzioa (PROLOGen modura), eskari konplexuagoak formulatzeko bide gisa;

---

<sup>1</sup> Hemen "hierarkia" diogunean, DAMpeko sarreraren egituratze hierarkikoari gagozkio, non sarrerek erregistratzeko arbola-egitura bat erabiltzen baita, *entry*, *hdw*, *homograph*, *sense*, *pos*, *definition*, etab. bezalako nodo etiketatuek dituen.

c) disjuntzio- eta ukapen-eragile boolearren erabilera eskari konposatuetan.

Estreinako puntua (a) dela eta, esan behar da RTHS bilaketa tesaurikoko funtzionalitateak nahiko eskari konplexuak formulatzeko aukera ematen duela, murriztapen-espresioei eman zaien sintaxiari esker. Ikusiko dugu orain, zenbait adibide dela medio, nola adieraziko liratekeen RTHS honen bidez, lengoia agintzaile batean baldintzazko edota iteraziozko eraikuntzarik eskatuko luketen eskari batzuk:

1. E.- **RTHS** ( (and (?X HYPERONYME |couleur I 1|  
(?X ?R |ciel I 1|))  
?X ?LC)

*Donne-moi les noms des couleurs en rapport avec le ciel.*

Eskari honen azpian balegoke *couleurs* multzoko elementuen gaineko iterazio bat, kolore bakoitzeko *cielekiko* duen erlazioa egiaztatzen delarik.

2. E.- **RTHS** ( (and (|vacarme I 1| CARACTERISTIQUE |tumultueux I 1|  
(|vacarme I 1| SYNONYMES ?X))  
?X ?LC)

*Donne-moi les synonymes de vacarme, s'il s'agit de quelque chose de tumultueux.*

eta honek, berriz, baldintzazko egitura bat (if-then tipokoa) idarokitzen du, hori ere RTHSren bidez aise espresatuz.

Eskari desberdinak konposatzeko aukera dela eta, HIZTSUAN eskaini nahi genukeen mekanismoa *eskari-konjuntzioa* da (b), PROLOG estiloan idatzia eta aldagai-baterakuntzarekin<sup>1</sup>. Horrela, eskarien konjuntzioa (AND) espresatzeko, nahiko da komaz bereizitako eskari sinple batzuk segidan jartzea; hori interpretatuko da HEBan ezarritako arrazoiketa-mekanismoak martxan jarriz frogatu beharreko helburu bat bezala, non emaitzak berekin izango baititu eskarian agertu aldagaien instantziazioak.

Horrelako eskari konposatuetan, erabiltzaileak honako konbinazioak egin ahal izango ditu:

- DDEF funtzioa aplikatu RTHS eskari baten emaitzako kontzeptuen listari, horrela bilaketa tesaurikotik ateratako kontzeptuen definizioak lortuz.
- Antzera, DDIF edota DRAP (ikus hemen behean 3. adibidea) aplikatu RTHS baten emaitza-listari. Kasu horretan, erabiltzaileak adierazi beharko du beste kontzeptu bat, eskariaren bigarren partean, funtzio horien bigarren argumentua izango dena

---

<sup>1</sup> Inplementatzeke oraindik.

(eta kontzeptu horrexekiko diferentzia edota elkarbidea izango da sistemak eman beharrekoa).

- Atributu bat zehaztuz gero, erabiltzaileak aplika diezaioke DPRO RTHS batetik ateratako kontzeptu-listari, bilaketa tesaurikoaren emaitza diren kontzeptuen propietate jakin bat lortzeko asmotan.
- Egiaztapen-eskariak erabili —emaitza boolearrekoak— baldintzak adierazteko (ikus 4. adibidea hor behean).
- Bestelako eragiketa primitiboak erabili oinarritzko funtzioekin konbinatuz (ikus 5. adibidea hemen behean).

Hona zenbait adibide:

3. E.- **RTHS** ( (and (?X HYPERONYME |couleur I 1|)) ?X ?LC),  
**DRAP** (?X, |ciel I 1|, ?R)

*Donne-moi les noms des **couleurs** en rapport avec le **ciel** (et les rapports correspondants).*

1. adibidean ikusi dugunaren antzeko eskaria, RTHS eta DRAP konposatuz adierazia oraingoan.

4. E.- **VPRO** (|brochet I 1|, ((HYPERONYME |poisson I 1|))),  
**DDEF** (|brochet I 1|, local, ?D)

*Si le **brochet** est un **poisson**, donne-moi sa définition.<sup>1</sup>*

Adibide honetan ikus daiteke nola adierazi if-then baldintzak, egiaztapen-eskariez baliatuz.

Beste adibide bat ematearren, ikus dezagun nola erabiliko litzatekeen RTHS funtzioa ORTHO<sup>2</sup> delako karaktere-kateen pattern-matching-eko funtzio batekin konbinaturik, gurutzegramazaleentzat dukeen interesa azalduz<sup>3</sup>:

---

<sup>1</sup> Eskari honek, adibidean dagoen bezala espresatu ahal izateko, alde zuzenetik *brochet*en adiera bat aukeratu izana eskatzen du (*brochet* hitz polisemikoa izanez gero, noski). Adiera hautatze hori —hitzaren adiera desberdinei aplikatutako VPRO sail baten bitartez egina— izango litzateke Eskari-Analisiko moduluaren emaitza, izan ere horrek ez baitu, erabiltzailearen eskarian, hitza besterik hartuko (eta ez adierarik).

<sup>2</sup> Inplementatu gabe prototipoan. Horrelako espresioak onartu beharrak murriztapen-espresioen sintaxia zertxobait aldatzea ekarriko luke.

<sup>3</sup> Asma litezke helburu horrekiko beste eragiketa primitibo batzuk ere: hitz baten letra kopurua kalkulatzekoa, adibidez.

#### IV. kapitulua

5. E.- **RTHS** ( (and (?X HYPERONYME |personne I 1|)) ?X ?LC),  
ORTHO (?X, '\*eur')

*Donne-moi des dénominations de **personne** terminées par 'eur'.*

S.-LC= ( |chanteur I 1| |professeur I 1| |vendeur I 1| |joueur I 1| . . . )

*Chanteur, professeur, vendeur, joueur, ....*

Eta azkenik (c), egiaztapen-eskariak konbinatzeko OR eta NOT disjuntzio- eta ukapen-eragileen erabilera onartuz, hara nola geratzen den eskari konposatuon sintaxia:

**eskari-konposatua**= *eskari-sinplea* {"", " *eskari-sinplea*}.

non, eskari sinple baten sintaxia:

**eskari-sinplea**= *galde-eskaria* | *egiaztapen-eskaria* |  
"not" "(" *egiaztapen-eskaria* ")" |  
"or" "(" *egiaztapen-eskaria* {"", " *egiaztapen-eskaria* }")".

Eta honek ematen dio jada erabiltzaileari, eskari konplexuagoak formulatzeko era aski malgu eta aberatsik.

#### IV.2.4 HIZTSUAren erantzunak.

HIZTSUAren Erantzun-Sorkuntzako moduluari dagokio erabiltzaileari eman beharreko erantzunak egitea. Modulu hori ez dago oraindik garatua prototipoan eta, beraz, sistemak oraingoz ematen dituen erantzunak, azaldu ditugun oinarritzko funtzio horiek emandako berberak dira.

Sistema urgazle adimendun batean, erabiltzaileari uneoro "zer" eta "nola" erantzun behar zaion garrantzi handiko zerbait da. (Hartley & Smith, 88) linean dauden laguntza-sistemen erantzun eta azalpenei buruzko artikulua ondo irakur daitekeen bezala, eskarien erantzunek irakaste- eta ikaste-prozesuaren mikrokosmos bat osatzen dute, lan-inguruneak eta erabiltzaileari buruz ezagun denak ez ezik, ikaste-teoria bat islatzen duten irizpideek gidatua. Autore horien arabera, halako sistema baten azalpen-moduluak kontuan izan behar lituzke, laguntzarekin berarekin batera, helburu pedagogikoak ere, erantzun ulergarriak, sinesgarriak, lagungarriak eta koherenteak eman ahal izateko. Artikulu horren 358. orrialdean, ondokoa irakur daiteke:

The experience underlined the importance of having adequate educational tactics to control explanation giving, i.e. the content and quantity of material given in reply, its sequencing and its expression.

Oso bestelako maila batean, baina gure proiektuan egindakotik hurbilago, (Calzolari, 84b)n adimen artifizialeko hainbat sistematan erabili ohi diren erlazio primitiboak hizkuntza



desberdinetan parafraseatzeaz hitz egiten du egileak. Horrek idarokitzen du, inondik ere, HIZTSUAren prototipoa berehala hobetzeko bide bat, atributu erlazional bakoitzari bere frantsesezko parafrase bat esleituz<sup>1</sup>; horri esker, Erantzun-Sorkuntzako moduluak "lengoaia naturalezko" eskema —testu— aurredefinituak eman ahal izango lizkioke erabiltzaileari erantzunetan.

HIZTSUAren erantzunak direla eta, gainera, ez da ahaztu behar erabiltzaileari erantzuteko garaiko ziurtasun-ezaren tratamenduari dagokiona (ikus § III.3.3), erantzun ez-ziurak eskaini eta erabiltzailearekiko elkarrekintza bati ekiteko estrategia orokorra ezartzera eramán gaituena, sistemaren portaera betegarriagoa izan dadin pentsatua.

Alabaina, gorago esan bezala, HIZTSUAren prototipoan erantzunen zehaztapena, oraingoz, aurreko ataletan ikusitako oinarritzko funtzioei dagokien berbera da (ikus adibideak)

---

<sup>1</sup> Parafrase hori jasotzen ahal da atributu bakoitzaren fazeta berezi batean.



## V. Erabiltzailearekiko elkarrizketa.

Txosten honetan aipatua dugun Grosseto-ko (Italia) mintegian —"Automating the Lexicon", 1986—, azken gomendio gisa emandako ondorio batean, Zampolli-k ondokoa aipatzen du lan interesgarri bezala (1987:333):

Study how people interact with standard and electronic dictionaries and lexical databases to determine the most effective procedure for human/machine interaction.

Marchionini-k (1989) entziklopedia elektronikoa bi osagaiko bilaketa-sistematzat dauka: batetik datu-base bat, informazioaren gordailu, eta bestetik informazio horren atzipenerako interfazea. Interfazea dela eta, gizaki-makina elkarrekintzarako komunikabide bezala definitzen du, eta paper gaineko sistemen egonkortasunari bilaketa-sistema elektronikoek eduki beharreko ezaugarri dinamikoak kontrajartzen dizkio. Horien artean aipatzen ditu atzipen-erregelak, erabiltzaileari pattern-matching-eko teknikak eskaini behar dizkietenak, eta eskariak aiseago formulatu ahal izateko eragiketa logikoak, paperezko sistema konbentzional bati egiten ahal zaizkion eskariak baino benetan eskari sofistikatuagoak egin ahal izateko, noski. Beste aspektu bat ere seinalatzen du autore horrek artikuluan, sistema elektronikoa hauekiko elkarrizketak sistemari berari eman behar dion atzeraelikadurarena, hain zuzen. Eta amaitzeko esaten du sistema benetan "hostokagarriak" ("browsable") eraiki behar direla, bilaketa-estrategia simple baina txit interaktiboak imitatuko dituztenak (616. orr.):

Conducting a dialogue with the system is an essential difference between print and electronic information retrieval systems.

Informazio-bilaketako sistema elektronikoen aurrean (entziklopediak, bere kasuan) erabiltzaileek duten jokamoldea aztertutik, atzeraelikadura eta elkarrekintza jotzen ditu autore honek aspekturik behinentzat.

(Fox *et al.*, 80)en ere ageri da hiztegi-sistema elektronikoa batekiko interfazeak eduki beharreko sendotasun horren ideia, ortografiaren aldetiko sendotasuna behinik behin. Izan ere, egia baita funtzionalitate hori gabe zerbait galdu egingo genukeela atzigarritasunari begira, paperezko euskarriaren baitan ongi finkaturiko ezaugarri simplea izaki: erabiltzaileak hitz baten idazkera zehatza ezagutzen ez badu ere, posible du aurkitzea hiztegiaren, "ortografikoki hurbilak" diren hitzak berorren inguruan daudelako, ohizko ordena alfabetikoa dela eta. Hiztegi elektronikookiko interfazetarako lengoia naturala, menu bidezko aukeraketa eta alde zuzenetik ezarritako funtzioen hautapen simplea aipatzen dituzte autore horiek.

## V. kapitulua

(Hartley & Smith, 88)n berriz, autoreek linean dauden laguntza-sistema adimendunen interfaze era desberdinak hartzen dituzte aztergai. Lengoaia naturalarekin batera, "form-filling" metodoak, editore egituratuak, galderekiko menuak, eta abar aipatzen dituzte. Sistemaren erantzunak direla eta, aurreko kapituluaren bukaeran aipatu berri dugu autore horien arabera erantzunok ulergarriak, sinesgarriak, lagungarriak eta koherenteak izan behar dutela.

V. kapitulu honetan HIZTSUAren baitako erabiltzaile-sistema komunikazioaz arituko gara, laburki. Sarrerako kapituluan (I) esan bezala, erabiltzaile eta sistemaren arteko komunikazioaren gai hau ez da izan lanaren helburu nagusietakoa —proiektuaren fase honetan— eta beraz, orain, sistemaren benetako interfazea izan litekeenaren zirriborro bat baino ezin aurkeztu.

### **V.1 Erabiltzaile-sistema elkarrekintzaren ezaugarriak.**

Elkarrekintza honetan ongi bereiztu beharra dago zein erabilingurutan diharduen erabiltzaileak. Kontuan izan sistemaren erabiltzaile "arruntaz" ari garela beti, hau da, ez garela ari ez hiztegian aldaketak edo egingo lituzkeen lexikografoaz, eta ezta HIZTSUA osagai lexikaltzat hartuko lukeen LNPko programaz ere. Izan ere, lehenengo kasurako bestelako funtzionalitate baten premian baikeundeke, proiektuaren fase honetan izan ditugun xedeetarik urrun; HIZTSUA programa batetik erabili ahal izateko berriz, interfaze-lengoaia definitu behar litzateke ongi, txostenaren IV. kapituluan azaldutako oinarrizko funtzionalitateen —eta/edo besteren— terminotan edo.

HIZTSUAra jotzeko abiapuntua den erabilegoera eta sistema bera (HIZTSUA) independentetzat dauzkagu. Horrela, Interfaze-Moduluak aurkeztuko dizkio erabiltzaileari, HIZTSUAra jotakoan, sistemari berez dagozkion leihoez gain Erabilegoera delako leihoa, non islatuko baita zer ihardueratan ari den erabiltzailea: testu-idazketa dela, irakurketa dela, eta abar.

Erabilegoera desberdinak hartu ditugu kontuan:

- *Testu baten irakurketa eta ulerkuntza.* Erabiltzaileak testu bat irakurtzen dihardu —Erabilegoera leihoan bistaratutako testua— eta HIZTSUAra jotzen du testu-ulerkuntzari dagozkion eginkizunetarako hiztegi-laguntza eske: kontzeptu-esplikazioa, propietate edo ezaugarrien egiaztapena, eta abar.
- *Testu baten idazketa.* Erabiltzailea testu bat idazten —edota aldaketak egiten— ari da. Ordenadorez Lagundutako Idazketa ingurune baten kasua litzateke (CAW,

Computer Assisted Writing), non HIZTSUAK bere hiztegi-laguntza eskainiko lukeen. Ez da zaila HIZTSUA horrelako ingurune batean integraturik imajinatzea<sup>1</sup>.

- *Itzulpengintza*. Bistan da honelako testuinguruan erabiltzaileak baduela laguntza-sistema eleaniztun<sup>2</sup> eta hiztegianiztun baten premia; halako sistema batek, funtzionalitate elebakarrez gain, itzulpengintzari egokitutako funtzionalitate berezia ere eskainiko lioke erabiltzaileari: hiztegi elebidunak (bi zentzuetan), hiztegi teknikoak, eta abar. Horrelako zerbitzua ezinbestekoa da Ordenadorez Lagunduriko Giza Itzulpenaren arloan (MAHT, Machine Aided Human Translation).
- *HIZTSUAren berezko erabilera*. Hau da, hiztegi-jotzen da zerbaiten eske, baina erabilegoera konkretu batekin lotu gabe: erabiltzaileak, informazio bat edota esplikazio bat nahi duelarik, edo hitz baten idazkeraz edo esanahiaz duen susmoa egiaztatu nahi duelarik, edo etortzen ez zaion hitz horren bila dabilelarik, etc. egiten duen hiztegi erabileraren kasua litzateke. Izan ere, hiztegi-sistema urgazlea erabiltzailearen lexikoa aberasteko tresna didaktikoa bihur daiteke, hizkuntza irakasteko tresna azken finean, dituen funtzionalitatez baliatuz horretarako (Fox *et al.*, 80).

Erabilegoera delako leiho horrek leiho interaktibo bat izan behar du, erabiltzaileak bertan bistaratutako testuaren gainean zenbait eragiketa egiteko aukera izan behar duelarik. Hala, hitz bat, hitz pare bat edo sintagmatxo bat aukeratuz gero (gainean klikatuz edo), sistema eta erabiltzailearen arteko elkarrizketa bati ekin beharko zaio —elkarrizketa hauetan sistemak hartuko du aurrea, jeneralean—; elkarrizketa horiek aldakorak izan beharko dute, erabilegoeraren arabera, erabiltzaileak egindako ekintzaren arabera, etc. aldatuko direnak, alegia.

Erabiltzailearekiko interfazeak —Komunikazio-Moduluarekin lankidetzan— ondoko ezaugarriak izango ditu:

- Fisikoki, *leihoanitzeko interfazea* izango da, leiho aktibo eta pasiboak (read-only) izango dituena;
- sistemak gidatutako *menu eta elkarrizketetan oinarritua*, erabiltzailearentzat interfaze adiskidetsu eta atsegina izan dadin.

---

<sup>1</sup> Izan ere, badira dagoeneko ortografi egiaztapenez gain —zeinak hiztegi bat edukitzea eskatzen baitu, funtzionalitate mugatukoa izan arren— idazketarako laguntza diren bestelako funtzionalitateak eskaintzen dituzten testu-lanketako programak (WordPerfect™ eta Word, adibidez): sinonimo-bilaketak, antonimoak, erlazionatutako hitzak, eta abar.

<sup>2</sup> Testu-ulerkuntza —irakurketa— eta -sorkuntzako —idazketa— inguruneetan, erabiltzaileak badu hiztegi laguntza eleanitzaren premia ere, hizkuntza arrotz baten aurrean dagoenean behintzat (azken finean itzulpen iharduera bat da, nolabait).

- *Dinamikoa*: menuak hierarkizaturik egongo dira, eta horrela, sistema gai izango da kasuan kasuko egokiena dena aurkezteko, erabiltzailea zertan ari den kontuan harturik. Komunikazio-Moduluaren barneko Eskari-Azterketako modulua da — Laguntzako Elkarrizketakoarekin batera— menu hauek maneiatu eta erabiltzaile-sistema elkarrekintza aurrera eramateko ardura duena.
- Erabiltzailearekiko elkarrekintzaren garaian *hiztegi-ezagutza atzitu ahal izango duena* —Komunikazio-Modulari esker, berriz ere—, ondoko helburuekin:

- a) Erabiltzaileari lagundu bere eskariak zehazten, horretarako hiztegi-informazioa eskainiz: antzeko idazkera (ortografia) duten hitzen zerrenda eman, sinonimoak, sarrera bati dagozkion adiera desberdinen zerrenda, eta abar.

Halaber, pentsatua dugu sistemak duen ezagutza "semantikoagoaz" ere baliatu ahal izatea, erabiltzaileari bere eskariak egiten laguntzeko asmo horretan: gogoan ditugu, esate baterako, sistemak, erabiltzaileak nahi duen kontzeptuaren kontzeptu "hurbilen" lista, edota nolabaiteko harremana duten kontzeptuen lista, emango lukeeneko kasuak.

- b) Erabiltzailearen "hutsegiteak", errakuntzak atzeman berorren eskarietan<sup>1</sup>: batzuetan bistakoa da erabiltzaileak egindako ortografi huts bat dela medio eskaria guztiz zentzugabea dela (adibidez, erabiltzaileak sistemari galdetuko balio ea *côté* (sic) hitzak *bord de la mer* esan nahi duen). Beste batzuetan bestelako errakuntza kontzeptual bat erakusten du eskariak: esate baterako, eman dezagun erabiltzaileak eskatzen diola sistemari egiaztatzeko ea *guêpe* den *produit du miel* egintzaren egilea, eta sistemak erantzuten diola esanez *abeille* dela eta ez *guêpe* (alegia, erlea dela, eta ez listorra, ezta egiten duena).
- c) Erabiltzailearen laguntza eskatu eskari konplexuen kasuan, sistema berez erantzun egoki bat emateko gauza ez denean. Erabiltzailearen laguntza hauek<sup>2</sup> izan daitezke desanbiguaketak, thesauruseko bilaketetan orientazio eta gidatzeak, etc., zeintzuk interesgarriak baitira oso kontuan hartzen badugu ezagutza ez-osoarekin lan egitera behartuak gaudela. Erabiltzailearekiko elkarrekintza horretaz baliatuko gara orobat sistemak erantzun ez-ziurrak ematean, hau da, sistemak, HEBko eraikibidean

---

<sup>1</sup> Errore-diagnostikorako sistema adimendunei hurreratuko gintzaizkieke hemen, umilki bada ere.

<sup>2</sup> Erabiltzailearekiko elkarrekintza hori dela eta, sistemaren ezagutza *aberasteko* posibilitateak zabaldu egingo liriteke. Erabilaldiko *ikaste-prozesu* bat litzateke, elkarrekintza hori lexikografoarekin —erabiltzaile aditu, beraz— gertatzean ez ezik, erabiltzaile arruntarekin gertatzean ere kontuan hartu beharrekoa dena.

jarraitutako prozedurak direla eta, ematen duen informazioaren erabateko ziurtasunik ez daukanean (ikus § III.3.3).

Hiztegi-ezagutza atzitze honi esker, erabiltzaileak formulatzen ahal ditu bere eskariak bai hitzen terminotan eta bai kontzeptuen terminotan ere —adieren terminotan, hau da, sarrerako hitza eta homografo eta/edo adiera konkretuaren adierazgailua izango duen espresioaren bidez—, berehala interpretatu eta Dedukzio-Moduluko oinarritzko funtzionalitateek eskatu modura itzuliko direnak.

- Menuetako aukera desberdinek ekintza diferenteak abiaraziko dituzte:
  - a) Menu zehatzagoen aurkezpena —"turrustan"—, erabiltzailea eskaria egiten ari dela gidatzearren.
  - b) Elkarrizketa-laukien aurkezpena, non sistemak erabiltzaileari eskatuko baitio aukeratutako funtzionalitateek behar dituzten parametroak zehazteko. Elkarrizketa-lauki horietan pattern-matching-eko funtzioak erabiliko dira, interfazea sendoago izan dadin, erabiltzaileak hitzaren idazkera zuzena ezagutzen ez badu ere sistemak huts egin ez dezan; halaber, aurrikusitako funtzio-konposaketarako bide eman beharko dute.
  - c) Erabiltzaileari informazioa aurkezteko irakurleihoak azaldu, non erlazionatutako sarrera-listak, sarrera bati dagozkion adiera desberdinak, adibideak, eta beste, bistaratuko baitira. Leiho horietako batzuk aktiboak izango dira, eta erabiltzaileari aukera emango diote, adibidez, bistaratutako lista bateko hitzen artean hautatzeko, sistemak kasu bakoitzean abiaraziko duen ekintza, ekintza hori gertatzen den testuinguruaren arabera izango delarik.
  - d) Eta azkenik, emaitzen aurkezpenerako erantzun-leihoak. Horiek ere, batzuetan, aktiboak izango dira, eta horrela erabiltzaileari sistemarekiko elkarrizketan aurrera egiteko bidea —erantzunean bistaratutako objektu baten gainean klik eginez, adibidez— emango zaio.

Erabiltzaile-sistema elkarrizketa hori kudeatzeko estrategiak ongi definitu beharra dago, erabilegoera, erabiltzaileak hautatutako menu-aukerak, eta sistemak kasu bakoitzean daukan ezagutza kontuan izanik. Estrategia horiek gauzatzeko if-then erregela klasikoak ditugu buruan, ondorio gisa mezuak igorri (objetuetara zuzendutako programazioaren zentzuan), eta HEBko egitura desberdinei (hitzak, kontzeptuak, konfigurazioak, eta abar) edo interfazea osatzen duten objektuei (menuak, leihoak, eta abar) esleitu metodoak exekutaraziko dituztenak.

## V.2 HIZTSUAren interfazearen zirriborroa.

Atal honetan deskribatuko dugu erabiltzailearekiko interfazea zer-nolakoa izango den. Ezer baino lehen, esan beharra dago ondokoa HIZTSUAren interfazearen zirriborro bat baino ez dela, sistema urgazlearen arkitekturaren atal honi dagozkion lerro nagusien azalpena, alegia. Beraz, etorkizunean gauzatuko den interfazearen diseinu-azterketa bat da, zein aldetatik jo dugun erakusteko azaltzen duguna. Eta hala hartu beharra dago, hau da, gure aldetik dagoen asmo edo nahi bat bezala, sistema egungo prototipoan egindakotik harantzago eramateko asmo bezala alegia.

Erabilegoera desberdinen arabera aldatuko da sistemaren interfazea. Konparazio batera, itzulpen-egoera batean itzultzaileak bi testu-leiho izango ditu, bata iturburu-testua erakutsiko diona eta bestea helburu-testua idazteko; testu-tratamendu arruntean aitzitik, erabiltzaileak nahikoa izango du leiho bakarra, lantzen ari den testua edukitzeko. Horregatik, nahiz eta atal honetan azalduko dugun interfaze zirriborroa orokor izateko asmoz jaioa izan, argi dago erabilegoera desberdinetan integratzerakoan aldaketak edota egokitzeak izan beharko dituela. Hemen, orokor-nahi izate hori delako Erabilegoera leihoak laburbilduko du, HIZTSUA erabilia den testuinguruaren errepresentazio izateko sortua denez.

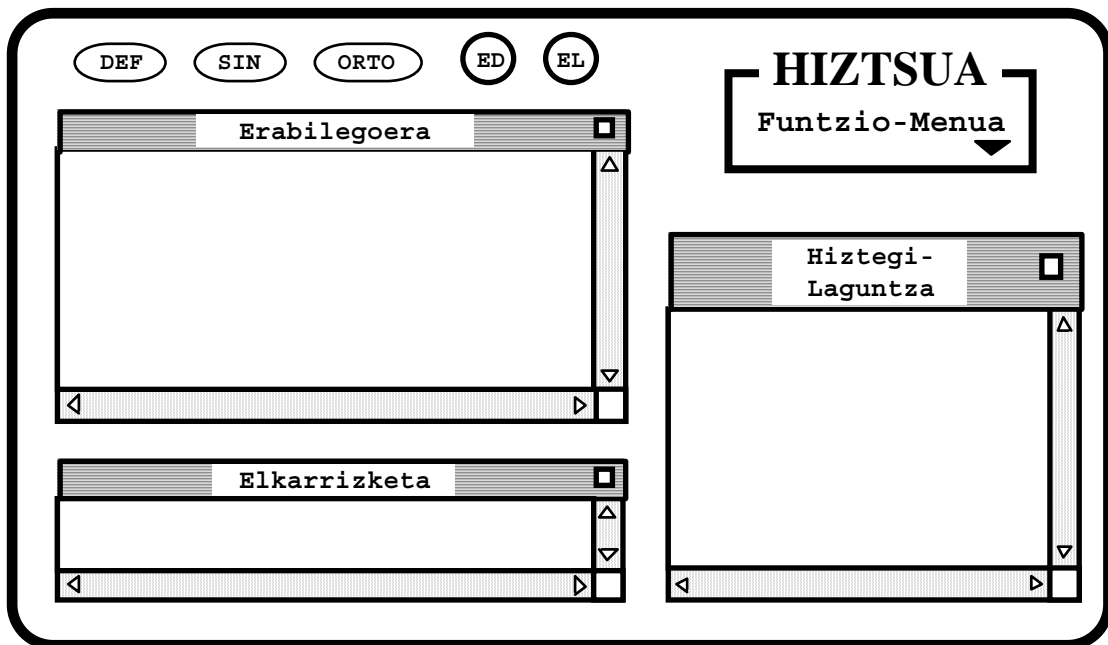
V.1 irudian dugu HIZTSUA erabiltzean agertuko diren leihoen multzoa. Ikus daitekeenez, hiru leiho dira:

- Aurreko atalean aipatutako *Erabilegoera*, hiztegi erabilera gertatzen den testuingurua islatzen duena.
- *Elkarrizketa* leihoa, non gauzatuko baita sistema eta erabiltzailearen arteko elkarrizketa. "Scrolling"dun leiho honek erregistratuko du sistema eta erabiltzailearen arteko elkarrekintzaren historia. Elkarrizketako leiho honen funtzio inportanteenetako bat erabiltzailearengandik beharrezko argumentuak lortzea da — sistemak hala eskatuta, noski—, Eskari-Azterketako moduluak erabiltzailearen eskariak HIZTSUAren oinarritzko funtzioen terminotan adierazi ahal izan ditzan. Halaber, sistema leiho honetaz baliatuko da, erantzun ahal izateko erabiltzailearengandiko informazioaren premia duelarik —desanbiguaketa interaktiborako laguntza behar duenean, batik bat—; horretarako, HIZTSUAK erabiltzailearekiko elkarrizketa berezi bati ekingo dio.
- Eta azkenik *Hiztegi-laguntza*, HIZTSUAK bere informazioa azaltzeko erabiliko duen leihoa: eskarien erantzunak, adibideak, erabiltzaileari aurkeztu beharreko lista desberdinak, eta abar. Beraz, leiho hori erantzunak aurkezteko leiho ez ezik, Elkarrizketakoaren leiho lagungarria ere bada, zeren honek, batzuetan, erabili ahal izango baitu hura sistemak erabiltzailearen ekintzaren bat eskatzen duen



informaziorik aurkezteko —hitz bat aukeratu, eskatutako informazioa zehatzago adierazi, etc.—.

Deskribatu berri ditugun hiru leiho horiek izango dute "scrolling"a, eta tamainaz aldagarriak, ezkutagarriak ("bury") eta "uzkurgarriak" ("shrink") izan beharko dute. Erabilegoerako leihoak, esan bezala egoeraren arabera desberdina izan daitekeenak<sup>1</sup>, horrelako hizkuntz softwareak izan ohi duen edizio ahalmena eskaini beharko du. HIZTSUAri dagokionean, leiho horrek permititu beharko du testu-zatiak saguaz klikatuz aukeratzeko, beste batzuek ordezkatzeko, eta abar. Hiztegi-Laguntzako leihoak ere permititu beharko du testu-zatiak aukeratu eta, saguaren laguntzaz, Erabilegoerakora eramaten, hitz ordezkapenak, parafraseaketak, testu-zatien birformulatzeak, eta abar, eragin ahal izateko.



**V.1 irudia.-** HIZTSUAren interfazearen ikuspegi orokorra.

Eragiketa guzti horiek eroso egin, eta orobat maiz usatzen diren funtzionalitateak agudo eta aise atzeman ahal izateko, pentsatu dugu erabiltzailearen eskueran botoi sail bat ipintzea (Erabilegoerako leihoaren gainean, V.1 irudian); botoi horiek ekintza konkretu bat izango dute esleitua, zein botoiok klikatutakoan exekutatuko baita. Gogoan ditugu definizio bat bistaratzeko funtzioa (DEF), hitz baten ortografia egiaztatzekoa (ORTO), hitz baten sinonimoen bilaketa (SIN), eta baita edizio-menuetako ohizko eragiketak (ED): kopiatu/erantsi, esate baterako. Beste botoi bati esleitu funtzioak (EL), "pop-up" erako menu

<sup>1</sup> Izan ere, leiho hori HIZTSUA integraturik edukiko lukeen "hizkuntz erabilerako" edozein inguruneren "errepresentazio bat" besterik ez baita.

bat azaleratu eta eragile logikoen zerrenda eskainiko dio erabiltzaileari, eskari konplexuak eta formulatzen ari dela lagungarri.

V.1 irudiko goialdean eskuinetara badago leiho txiki bat, "Funtzio-Menua" dioena, eta honen azpian gezi-mutur bat. Gezi-mutur horrek adierazi nahi du badela turrustan menu bat, leihotxo honen gainean klikatuz gero bistaratuko dena: HIZTSUaren *Funtzio-Menu Nagusia* da (ikus V.2 irudia). Menu nagusi honetan ondoko aukerak izango ditugu:



**V.2 irudia.-** HIZTSUaren Funtzio-Menu Nagusia<sup>1</sup>.

*Sarrera bistaratu* aukera, aukera "konbentzionaltzat" har dezakegu, edozein hiztegi-sistemak eskaini behar lukeena. Aukera horrek erabiltzailearekiko elkarrizketa bati emango dio hasiera, honek zein sarrera bistaratzea nahi duen adieraz dezan. Erabiltzaileak tekleatutakoa onartzeari dagokionez<sup>2</sup>, ondokoak izan behar dira kontuan:

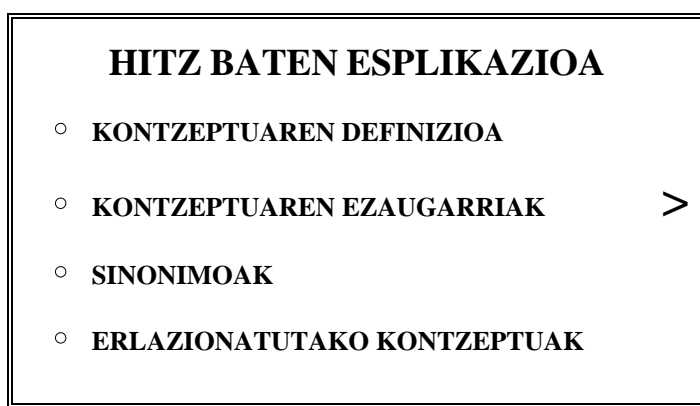
- Sistemak espresio ez-osoak onartu beharko ditu.
- Sistemak aukera eman beharko du patroi erako espresioak sartzeko, erabiltzaileak hitzaren ortografia zehatza ezagutzen ez duelarik eskariak egiterik izan dezan.
- Sistemak, eskariren batean sarrera "ezezagunik" jasoz gero, hurbileneko hitzekin — ortografia aldetik— osatutako lista bat erakutsi beharko dio erabiltzaileari; horrela, honek, lista horretan erreparatu eta nahi duen sarrera aukeratu ahal izango du.

<sup>1</sup> Aukera baten eskuinetara gezi-muturra egoteak turrustako menu bat dagoela adierazten du, aukera hori hautatuz gero bistaratuko dena.

<sup>2</sup> Erraztasun hauek, hemen ez ezik, erabiltzaileak hitzak tekleatu behar dituen guztietan izan beharko dira kontuan.

Aukera hori hauta dezan erabiltzaileak, normalean, hitzaren adieren definizio testualak hartuko ditu erantzuntzat, eta horiekin batera hiztegieta bestelako informazio tipikoa: kategoria, adibideak, eta abar.

HIZTSUA eta erabiltzailearen arteko elkarrekintza osoan, sistemak kontuan izan beharko du beti *erabilegoera* zertan den, hau da, beti ere "ezagutu" egin beharko du zein den egoera hori eta horren arabera jokatu. Esate baterako, erabiltzaileak testuko hitz bat klikatu eta aukeratu baldin badu, eta ondoren sarrera bistaratzeko eskatzen baldin badu menuan, bistan dago eskaria aukeratutako hitzari dagokiola eta, beraz, ez zaio galdetu beharko zein sarrera bistaratzea nahi duen.



### **V.3 irudia.-** Kontzeptu-Esplikazioaren Menua.

Erabiltzaileak menu nagusiko *Hitza esplikatu* aukera hautatzen duenean, sistemaren esplikazio-funtzioak dira martxan jarri beharrekoak. Segituan beste menu bat (V.3 irudia) aurkeztuko zaio erabiltzaileari, zer-nolako esplikazioa nahi duen galdetuz. Menu horretako aukera desberdinen hautaketak erabiltzailearekiko elkarrizketa bati ekingo dio, eta, printzipioz, HIZTSUAREN oinarritzko funtzionalitateei egingo die dei: DDEF, DPRO, DRAP. Elkarrizketa horren helburuak ondokoak izango dira:

- 1) erabiltzaileari hitzaren zein adiera edo adierak interesatzen zaizkion zehaztu<sup>1</sup>;
- 2) eskatutako funtzionalitateari dagozkion beharrezko argumentuak zehaztu<sup>2</sup>;
- 3) eskatutako kontzeptuaren edo kontzeptuen esplikazioa eman, eta kontsultarekin aurrera egiteko aukerak eskaini; adibidez, definizioaren birformulaketa (RDEFi

---

<sup>1</sup> Adiera zehazte hori lortzeko bide desberdinak erabiltzen ahal ditugu. Esate baterako, sistemak aurkez diezaioke erabiltzaileari hitzaren adiera desberdinen hiperonimoen zerrenda, edota sinonimoena, eta eskatu aukeratzeko; edota, erabiltzaileak aldeztu aurretik mintzagaia zein den adierazi baldin badu —hala eskatu zaiolako, adibidez—, sistema gai izan daiteke zentzu batzuk aukeratu eta beste batzuk baztertzeko.

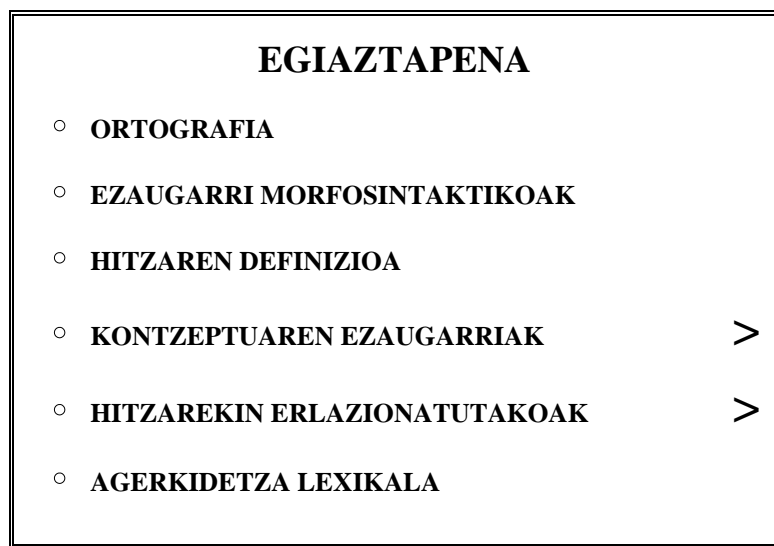
<sup>2</sup> Menu guztietan egongo dira definiturik balio lehenetsiak, erabiltzaileak aukera bati dagozkion argumentuak zehazten ez dituelarik erabili ahal izango direnak.

## V. kapitulua

deitu), beste era bateko esplikazioa (sinonimoak eskatu definizio-galde baten ondoren, etc.) nahi badu, eta abar<sup>1</sup>.

Menu nagusiko *Hitza bilatu* delako aukerak, jakina, bilaketa tesaurikoko funtzionalitateari egingo dio dei (RTHS), erabiltzaileak emaniko murriztapen-espresio batetik abiatuz murriztapen horiek betetzen dituzten kontzeptuen zerrenda lortzen duen funtzioari alegia. Hemen arazoa izango da erabiltzaileari murriztapen horiek zehazten laguntzea, zehazte hori interaktiboki eta ahalik eta modurik atseginean egin dadin.

V.4 irudian ikus daiteke *Egiaztapen-funtzioei* dagokien menua. Menu horretan ipini ditugu, oinarrizko egiaztapen-funtzioei dei egingo dieten aukerak ez ezik (VDEF, VPRO, VRAP), hiztegi orok eskaini behar lituzkeen bestelako egiaztapen-eragiketa arruntak ere, hala nola hitz baten ortografia edo ezaugarri morfosintaktikoen egiaztapena; halaber, bertan ezarri dugu garatu gabe dagoen agerkidetza lexikaleko murriztapenen egiaztapena egingo lukeen funtzioari dagokiona ere.



### V.4 irudia.- Egiaztapen-Funtzioen Menua.

Menu nagusiko *Parafraseaketa* aukera, IV. kapituluan axaletik baino azaldu ez dugun parafraseatze-laguntza eskaini asmoz sortua da.

Eta amaitzeko, V.5 irudian dago hitzen arteko *Erlazioak eta Diferentziak* bilatzeko erabili beharreko menua, DRAP eta DDIF funtzioetan oinarritutako aukera pareta dituen. Aukera horiek hautatutakoan, sistemak lagundu egin beharko dio erabiltzaileari hitz pare bat seinalatzen. Elkarbidearen bilaketan kontuan izatekoa da, bestalde, bilaketa-prozesuan

---

<sup>1</sup> Ez dugu hemen gehiago zehaztuko nolakoa izango den elkarriketaren jarraipen hori, eta ezta ere zer-nolako menuak azalduko lirakeen orduan.

erabiltzaileak parte hartzeko posibilitatea (ikus § IV.2.2.2.4en); kasu horretan, Komunikazio-Moduluari dagokio partehartze hori gauzatzeko elkarrizketa bideratzea.

## **ERLAZIOAK ETA DIFERENTZIAK**

- **HITZEN ARTEKO ELKARBIDEA**
- **HITZEN ARTEKO DIFERENTZIAK**

**V.5 irudia.-** Hitzen arteko Erlazioak eta/edo Diferentziak bilatzeko menua.

Oinarrizko funtzio desberdinen konposaketari dagokionean (ikus § IV.2.3.1), sistemak erraztasunak eman behar dizkio erabiltzaileari eskari konplexuak adierazteko. Adibidez, RTHS eskari baten emaitza den kontzeptu-lista erabiltzaileari aurkezteko, bertatik hitz bat edo batzuk hautatzeko aukera duen leiho batean egin beharko da, hautatu hitz horiei beste funtzio batzuk aplikatuz eskariarekin aurrera egin ahal izan dadin. Hori litzateke funtzio horiek era dinamiko batez "konposatzeko" bide bat, baina baliteke erabiltzaileak aldeztatik aurretik prestatutako duen eskari konposatu bat egin nahi izatea. Kasu horretan, eman beharko zaio espresio horietan agertu ohi diren eragile logikoak hautatu eta erabiltzeko biderik (ikus "EL" deituradun botoia, V.1 irudian), horretara eskariok aiseago egin ditzan.

## **V.3 Erabiltzaile-sistema komunikazio-prozesua eta moduluen arteko elkarrekintza.**

Erabiltzaile eta sistemaren arteko komunikazio-prozesua nola garatzen den, eta, komunikazio hori gertatzen delarik, HIZTSUAren moduluen arteko elkarrekintza zertan den azaldu nahi genuke kapituluko azken atal honetan.

Esana dugu, txostenaren lehen kapituluan, HIZTSUAren Komunikazio-Moduluari dagokiola —eta bere azpimodulu diren Eskari-Azterketa, Laguntzako Elkarrekintza eta Erantzun-Sorkuntzari— erabiltzailearen eskarien kariaz sortuko den elkarrizketaren kudeaketa; horretarako, Komunikazio-Moduluak Interfazea maneiaturiko du —aurreko atalean deskribatu berri dugun menu- eta leiho-sistema— eta sistemaren Dedukzio-Moduluarekin batera egingo du lan.

Eskari-Azterketako Moduluak aztertuko ditu erabiltzailearen eskariak —Interfaze-Modulutik datozenak— eta Dedukzio-Moduluari igorriko dizkio, oinarrizko funtzioen terminotan adierazita. Hori egin ahal izateko, Eskari-Azterketako moduluak ondoko funtzioak bete behar ditu:

## V. kapitulua

- "Hitza-adiera" arazoa ebatzi<sup>1</sup>, baldin badago, HEBko DICTIONNAIRE azpi-EBaz baliatuz eta, batzuetan, erabiltzailearen laguntzaz.
- Erabilegoera zein den kontuan eduki —kurtsoreak, aukeratutako zatiak, aldez aurretik HIZTSUArekin izandako elkarrekintza, etab.—, elkarrizketa era dinamikoago batez eraman ahal izateko, Interfaze-Moduluak unean uneko egoerari dagozkion aukerak baino azal ez diezazkion erabiltzaileari.
- Erabiltzailea dagoeneko erabilegoera eta sistemak daukan hiztegi-ezagutza integratu, sistemaren aldetik "adimendutsuagoa" izango den elkarrizketa egin asmoz. Eskari-azterketak beharko du, normalean, erabiltzailearekiko komunikaziorik, erabiltzaileak bere kontsulta zehatz dezan eskatzearren. Komunikazio hori Laguntzako Elkarrekintzako Moduluari esker bideratuko da, izan ere hori izango baita Eskari-Azterketakoa eta Interfazearen arteko bitartekaria.

Erabilegoera integratzeari begira bada, eta erabiltzaileari elkarrekintza interesgarriago bat eskaini asmoz, modulu honek hiztegi-ezagutza atzitzen ahal du, eta horrela, erabilegoera konkretuari eta HEBtik eskuratu ezagutzari egokituriko menuak azaleratu.

Dedukzio-Moduluak ere eska dezake Laguntzako Elkarrekintzako Moduluaren partehartzea, sistemak erabiltzaileari erantzun ahal izateko honekiko elkarrizketa baten premia duen kasuetan. Azkenik, Erantzun-Sorkuntzako Moduluari dagokio sistemaren erantzunak prestatu eta erabiltzaileari aurkeztea.

Interfaze-Moduluaren ardura da erabiltzailea eta HIZTSUAren arteko elkarrekintza fisikoki gauzatzea. Esan bezala, modulu hau Komunikazio-Moduluaren manupean dago eta erabiltzailearen ekintzen aurrean dinamikoki erantzun behar du, erabiltzaileak bere kontsultak era atsegin batez egin ditzan.

Eta bukatzeko, ikus dezagun labur-labur nola gauzatuko den sistema eta erabiltzaile baten arteko elkarrekintza.

HIZTSUA eta erabiltzaile baten arteko elkarrekintza, honen eskari batez hasiko da, normalean. Aldez aurretik, sistemak, bere Interfaze-Moduluaren bitartez, haren erabilegoerari dagozkion aukerak eskainiko zizkion, erabiltzaileak zertan diharduen eta eskaria egin aurretik zer-nolako ekintzak burutu dituen kontuan hartuz, batik bat. Behin eskaria adierazi eta gero, Interfaze-Moduluak Komunikazio-Moduluari egingo dio dei, erabiltzaileak egindako ekintzari eta erabilegoerari dagokion informazio guztia igorritz.

---

<sup>1</sup> Erabiltzailearen eskariak, normalean, hitzen funtzioan adieraziak izango direnez, modulu honi dagokio adiera edota zentzuen funtzioan (kontzeptuak) adierazitako espresioak lortzea. Hori, askotan, adiera bakar bati dagokion eskari bakuna bihurtuko da, eta beste batzuetan, adiera bakoitzeko eskari bat duen eskari saila.

Berehala Eskari-Azterketako Moduluak hartuko du kontrola, erabiltzailearen eskaria oinarritzko funtzionalitateen terminotan adierazi eta Dedukzio-Moduluari bidaltzearen. Horra iristerako, litekeena da Eskari-Azterketako moduluak HEBra jo behar izana, askotariko arazoak ebazteko: adiera konkretu bat zehaztu, erabiltzailearen errakuntza edota hutsegiteak atzeman, laguntza hobea eman asmoz listak eratu, eta beste. Orduan, Eskari-azterketa izango da hor erabiltzailearekin gertatuko den elkarrizketa —gertatzen bada, noski— kudeatuko duena. Egiteko horretan, modulu hori Laguntzazko Elkarrekintzakoaz baliatuko da, erabiltzaileari lortutako informazioa azaldu edota berorren hasierako eskaria hobeki zehazteko eskatzearen.

Prozesu horren ondorio gisa, Eskari-Azterketakoak eskari bat lortu eta Dedukzio-Moduluari bidali behar dio. Horrela abiarazten da dedukzio-prozesua modulu horren baitan, eta, printzipioz, modulu horri dagokio erantzun bat lortzea. Prozesu horrek, HEBaren gainean ezarritako dedukzio-mekanismoez baliatuz —oinarritzko eskari-funtzioak, herentzia, erlazio lexikalen konposaketa, eta abar—, elkarrekin lan egiten duten prozedura batzuk jartzen ditu martxan, DICTIONNAIRE eta, batez ere, THESAURUSEko ezagutza atzituko dutenak. Gorago esan bezala, Dedukzio-Moduluak, erantzuna lortzeko erabiltzailearen laguntza behar baldin badu, Laguntzazko Elkarrekintzakoaz baliatu eta erabiltzailearekiko beste elkarrizketa-azpiprosesu bati ekiten ahal dio.

Erantzuna prest dagoelarik, Dedukzio-Moduluak Erantzun-Sorkuntzakoari bidali eta honek, Interfazearen bitartez, aurkeztuko dio erabiltzaileari. Erabiltzailearekiko elkarrizketari jarraitu behar izanez gero, kontuan izan beharko da beti erantzunaren ondorengo erabilegoera berria.





# AURRERA BEGIRAKOAK ETA ONDORIOAK

## **VI. Zabaldutako ikerlerroak eta perspektibak.**

Tesi-proiektu honen etorkizuneko perspektibak zein diren azalduko dugu kapitulu honetan, labur-labur. Perspektiba horiek hiztegien eta, oro har, informatikaren munduan zabaldurik dauden ikerlerro desberdinetan kokatzen saiatuko gara. Miller-ek dioen bezala (1985:313), hiztegien mundu honetan jende guztiarentzat baitago lana, ikuspuntu askotatik begiratuta:

There is work here for everyone —linguists, philosophers, and psychologists, as well as computers scientists— and it is not abstract or impractical work.

Lan honetan zabaldutako ikerlerroak ere aski desberdinak dira. Ikerlerro horien ikuspegi orokor bat emango dugu ondokoan, nahiz eta alde zehatzetatik esan behar dugun ez zaizkigula denak berdin interesatzen. Izan ere, lerro horietako zenbaitetan dagoeneko lanean hasiak garen bitartean, beste batzuk aipatu baino ez ditugu egingo hemen, beren interesa eta proiektuan erabilitako gaiekin duten zerikusia azalduz.

Lehenik, hiru ataletan —VI.1, VI.2 eta VI.3— ondoko aspektuak hartuko ditugu ahotan: (1) HIZTSUA erabilinguruarekin eta erabiltzailearekin benetan integrazeari begirakoak, sistema, erabiltzailearekiko komunikazioaren aldetik behintzat, "adimendutsuagoa" egitearren; (2) HEBko ezagutzaren aberaste sakonago bat lortzeko eta bertan dedukzio-mekanismo berriak eta eraginkorragoak ezartzeko egin beharrekoak, eta (3) ingurune hiztegianitzeko eta funtzioaniztunaren ideiarri buruzkoak.

Azkeneko atalean berriz —VI.4—, hiztegi-sistema adimendun bat bestelako inguruneetan erabiltzeko aukeraz arituko gara, hala nola lengoia naturaleko sistemetan, dokumentuetako bilaketakoetan, eta abar.

## **VI.1 HIZTSUA erabilinguru eta erabiltzailearekin integratzea: sistema "aktiboago" eta "adimendutsuago" baterantz.**

Lan honen alderik ahulenetakoa, batez ere sistema laguntza-sistema adimendun bezala hartuta, erabilinguru eta erabiltzailearekiko integrazioari dagokiona da.

Sistema urgazle adimendun batean *erabilinguruarekiko integrazio "handiagoa"* nahi litzateke, sistemak, hiztegi-laguntza ematen den inguruan tratatzen ari den testua eta gertatzen ari den erabilerari dagozkion gorabeherak, sakonago ezagut ditzan. Beharrezko ote da ordea, benetako "laguntza adimenduna" eman ahal izateko, sistemak testua "ulertzea"?<sup>1</sup>

Erabiltzaile mota desberdinak, eta hizkuntzari buruz erabiltzaile desberdinek duten ezagutza maila desberdina ere kontuan hartzekoak dira, erantzunak era adimendutsuago batez emango baldin badira. Horrek sistema *erabiltzailearen eredu* batez hornitzera eramango gaitu eta, irakaskuntza-sistema adimendunetako (ICAI, Intelligent Computer Assisted Instruction) erabiltzailearen ereduaren gisan, sistema erabiltzaile bakoitzaren premietara hobeki egokitzen lagundu. Eredu hori eraikitzeko, sistemak kontuan izan beharko ditu erabiltzaileak dakiena eta ikasitakoen asimilazio maila, berorren ikasteko gaitasuna, interesa eta abarrekin batera.

Sistema urgazle adimendunei buruz lehenengo kapituluan aipatu baina proiektuan sakonki aztertu ez den zenbait aspektuk ere erakusten digu etorkizuneko ikerbidea. Batik bat, sistemak erabiltzailearekiko elkarrekintzan *jarrera "aktiboago" bat hartzeari* buruzkoa, eta planifikazioari eta erabiltzailearen planak ezagutzeari dagozkionak (ikus § I.2.2n).

Hori dela eta, *sistema planifikazio-teknikak erabiliz birdiseinatzea* kontuan hartzekoa izango litzateke; horrela, erabiltzailea eta sistemaren arteko ihardunari dagokion helburu- eta egintza-sekuentzia bat eraiki ahal izango da, dinamikoki. Lehenago ere esana dugu, sistemak gai izan behar lukeela erabiltzaileari laguntzeko honek "buruan duen programa" —hiztegitik erantzuna lortzeko "programa mentala"— sistemarentzat egokia den eskari bihurtzen. Sistemaren laguntzaz baliatuko den prozesu kognitiboaz ari gara, beraz.

Horretarako, sistemak eskaini beharko du lengoaiaren bat, delako "programa mental" horren hurbila eta ahalmentsua, ahalik eta naturalena. Sistema adimendunei buruz hitz egiten

---

<sup>1</sup> Quillian-en (1969) TLC sisteman bezala, agian?

denean beti aipatzen den puntu batera ekarri gaitu honek, azkenean: *erabiltzailearekiko elkarrekintza lengoaia naturalean*.

### **VI.1.1 Lexikografoa erabiltzaile. Hiztegi-gintzarako laguntza.**

Erabiltzaile desberdinak aipatuta, nabarmen ageri da horien artean lexikografoa. Txosten honetan behin baino gehiagotan aipatu dugu zer-nolako aukerak emango lituzkeen, HEBaren aberasketarako, sistema eta lexikografoa elkarrekin aritzeak. Ikaste hitza ere aipatu dugu, izan ere sistema, ezagutza berriak eskuratzeko, halako elkarrekintza batez baliatzen ahal baita. Beste aspektu bat, hiztegi-sistema adimendun bat hiztegi-gintzan diharduen lexikografoaren eskuetan jartzeari legokiokeena da. Miller-ek eta (Miller *et al.*, 88:557), artikulua ondo ondo azpimarratzen dute hiztegi handi bat euskarri elektronikoan edukitzearen erabilgarritasuna, helburu horrekin:

In addition, the electronic form of a comprehensively coded lexicon should allow other researchers to extend and examine the architecture of words in new and interesting ways.

Editore berezituak, Mel'cûk-en taldekoen lexikografi lanetarako egindakoa (Décary & Lapalme, 90), esate baterako, egokiak dira oso sarrerak idazten diharduen lexikografoarentzat, datuen zuzentasuna, kontsistentzia, bigizta-detekzioa, osotasuna, eta abar egiaztatzearen aldetik, eta hiztegi baliagarri eta koherente bat —teoria linguistiko jakin baten arabera— egiten laguntzen diote.

(Boguraev & Levin, 88)n, egileak kexu dira lexikografoek ez omen daukatelako sarrera kontsistente eta osoak egitea bermatuko lukeen teoria aski sofistikatorik, lexikoaren antolakuntzari buruz. Grosseto-ko mintegiaren ondorioen artean ere (Zampolli, 87:333), azpimarratzen da lexikografoen jokamoldea modelizatzearen beraien lana aztertzeak duen interesa: “*Study the work of lexicographers to model their behaviour, incorporating the results in knowledge-based systems that support lexicographic activities.*”.

Interesgarria litzateke ikustea ea, gurean hizkuntz lexikoa antolatzeko hartutako eredia baliagarri litzatekeen hiztegi-gintzan.

## **VI.2 HEBaren aberasketa: dedukzio-mekanismo berrien ezarrera.**

Txostenaren III. kapituluaz azaldu ditugu HEBaren aberasketa eta sisteman ezarritako dedukzio-mekanismoak.

Aberasketa dela eta, bi fase bereizi ditugu: HEBaren hasierako eraikuntzari dagokiona, eta bigarren fasea, non aberasketa horri begira berariaz exekutarazitako prozedura batzuek

## VI. kapitulua

baliatu baikara. Interesgarria deritzagu ezagutza hori aberasteko posibilitateak sakon aztertu eta HEBan aplikatzeari: buruan dugu, batik bat, HEBko elementu desberdinen *desanbiguaketa*, lexikala nahiz sintaktikoa. Lan honetan erakutsi dugu honi buruz urra daitekeen zenbait bide, baina uste dugu arazoa aztertzen jarraitu beharra dagoela, desanbiguatze-lan horietan laguntza eskainiko duten mekanismoak, seguruenik interaktiboak eta heuristikoetan oinarrituak, ezarri. Anbiguotasuna benetako arazo eta oztopoa da, izan ere, sistemaren inferentziarekin zerikusirik duen edozertarako.

Komentatu nahi genukeen beste puntu bat taxonomien finketari dagokiona da, eta finketa horren ondorioz egin nahi genukeen atributuen fintzeari buruzkoa. Ikusi denez, proiektuan eratutako hierarkietan "kontzeptu-klase gorenak" maila abstraktuegian geratu dira, gertuegi agian kategoria gramatikaletatik, eta horrela ezin dira —klase horietan— kontzeptu-klase bakoitza ezaugarritzen duten atributuak zehaztu. Klase horien finketa batek beraz, eta finketa horretarako sisteman dagoen hiztegi-ezagutzaz balia gaitzkeelakoan gaude, bide emango luke, gure ustez, errepresentazio zehatzago bat egin eta atributuen semantika hobeki zehazteko.

Erabiltzailearen eskaripean sistemak duen dedukzio-ahalmenari dagokionez berriz, *erlazio lexiko-semantikoen konposaketaren adierazpidea* sakon aztertu beharko litzatekeela uste dugu —eta konposaketa fenomeno horiek espresatzeko lengoia ahalmentsu bat asmatu, beharbada—; halaber, erlaziook hizkuntz kontzeptuekin dituzten harremanek ere merezi dute azterketa sakonagorik. Berririo HEBko anbiguotasunari gagozkiolarik, ikertzekoa litzateke baita ere sisteman *ziurtasun-ezarekin nola jokatu*, hots, erabiltzaileari erantzuterakoan sistemak aurkitzen dituen anbiguetateei nola egin aurre.

### **VI.3 Definizioen errepresentazio-ereduaren baliozkotasuna. Sistema hiztegianiztunak.**

Lan honetan aukeratutako *errepresentazio-eredua* beste hizkuntzetan idatzitako definizioak *errerepresentatzeko ere baliozkoa den ala ez* ikusteari erronka benetan interesgarria deritzagu. Hiztegi-definizioak direla medio kontzeptuen artean ezarritako erlazio lexiko-semantikoak, berberak ote dira edozein hizkuntzatan? Erlazio batzuen kasuan bistakoa da baietz: hitz bat bere sinonimoen zerrenda emanez definitzea ez da, noski, frantsesezko hiztegiatan bakarrik gertatzen den zerbait. Baina, metalengoaiatzko deitu ditugun erlazioen kasuan, zer? Horrelako erlazioen kasuan parte hartzen duten hitz edota espresioen aukera, hizkuntza jakin bateko tradizio lexikografikoaren menpeko ote da? Edo, hizkuntza baten eta bere metalengoaiaren arteko halako menpekotasun estu bat islatzen ote du, hizkuntza edozein dela ere?

Errepresentazio-eredua balioztatzearen aldetik, aztertu nahi genukeen beste aspektu bat hiztegi-sarreretakoa adibideen errepresentazioarena litzateke. Izan ere, adibideetan badago, gure ustez, hiztegia laguntza-sistema den aldetik ikusita kontuan hartzeko moduko ezagutza multzo garrantzitsua, hizkuntz ezagutza batik bat. Besteak beste, aztertu nahi genuke ea definizioen errepresentazioan erabilitako konfigurazio-unitateen ideia baliagarria ote den adibideak errepresentatzeko ere (maiz izen-, aditz-sintagma txikiak baitira, izan ere).

Galdera hauei erantzuna ematearren, ari gara jada proiektu berri batean lanean. Lan horretan ikergai, frantsesetik hain urruna den —egitura aldetik begiratuta, behinik behin— euskara hartu dugu. Euskal hiztegi elebakar zenbaitetatik harturiko ehun bat definizioak osatzen dute errepresentazio-eredua probatzeko lehen multzoa.

Proba horiek eta, testuinguru berri batean kokatzen ditugu: *ingurune eleanitzeko itzulpenerako laguntza-sistema hiztegianiztuna*. Sistema horren sorkuntzan, hiztegi desberdinen lankidetzak kontuan hartu beharrekoa da: elebakarrak eta elebidunak, orokorrak eta berezituak (teknikoak), sinonimo-hiztegiak, etimologikoak, eta abar. Halako sistema batek itzulpen-testuinguru batean eskaini beharreko funtzionalitatea aztertzea, eta "hiztegi-sistema aditu" desberdinon arteko lankidetzak zertan gauzatzen ahal den ikustea dirateke lanerako bide interesgarriak.

Horrekin zerikusia badu A. Zampolli-k (1987) deskribatutako Italiararen Datu-Base Lexikal Funtzioaniztunak (IMLDB, Italian Multifunctional Lexical Database). Ele bitako itzulpenerako laguntza-ingurune batean kokatutako sistema horren osagaiak ondokoak dirateke: hizkuntza bakoitzeko datu-base lexikal elebakar bana (MLDB), bi MLDBak elkarlotzeko "zubi" elebiduna, bi hizkuntzetako erreferentziako corpusak eta "testu elebidun kontrastatuak"<sup>1</sup> edukiko dituen testu datu-basea, eta giza erabiltzaile nahiz programatzaile datu-atzipenerako prozedura eta software multzoa.

Grossetoko mintegiaren ondorioetan (Zampolli, 87:335), partehartzaileek hiztegi desberdinen artean erlazioak ezartzeak duen interesa azpimarratzen dute:

Develop methodologies for interrelating monolingual and bilingual dictionaries; explore the possibility of combining technical monolingual dictionaries with bilingual general dictionaries to create technical bilingual dictionaries.

Zampolli-k (1987) berak ere seinalatzen du MLDB "neutroak" elkarrekin lotzeko metodologiak garatzearen interesa, abiapuntutzat hiztegi elebidunak hartuz.

---

<sup>1</sup> "Structure including a text in one language, its translation into another, plus a set of cross-references explicitly indicating the relations of (translational) equivalence between the corresponding elements of the two texts;" (Zampolli, 87:321).

## VI.4 HIZTSUA bestelako inguruneetan erabiltzeko aukerak.

HIZTSUA ingurune desberdinetan erabiltzeko posibilitateak ere lanerako bide berriak zabaltzen dizkigu.

Hiztegi-laguntza linean eskaintzen duen sistema bezala, interesgarria litzateke HIZTSUAK *testu-tratamenduan* eskain lezakeen funtzionalitatea aztertzea: erabiltzailea idazten ari dela sinonimo-zerrendak aurkeztea, parafraseaketarako laguntza linean (Chodorow *et al.*, 88), hitzen erabilera gramatikala egiaztatu eta zuzendu (Fox *et al.*, 80), eta abar.

Beste alde batetik, ordenadorez lagundutako irakurketan *erabiltzaileak testuaz jabetzeko egiten duen iharduerak* ere eskaintzen du HIZTSUAREN antzeko funtzionalitateez bali daitekeeneko ingurune bat. Izan ere, erabiltzaileak, testuaz jabetzen ari delarik, iharduera anotatibo dinamiko bat izango du (Evrard *et al.*, 90), zeinetan hiztegi-sistema baten eskarien moduko galderak egingo baititu: esplikazio-galdeak, elkarbide-bilaketak, esanahi-egiaztapenak, eta abar.

Halaber, HIZTSUA ikus daiteke *hizkuntza irakasteko eta ikasteko tresna* bezala ere. Hiztegia, hizkuntzari buruz halako ezagutza multzoa duen biltegia den neurrian, ongi izaten ahal da funtsezko osagai bat irakaskuntza-sistema batean. Hiztegi-sistema adimendun batek, irakaskuntza-sistema batean integraturik, honek domeinuari buruz duen ezagutza —hizkuntz ezagutza— aberasteko biderik eman behar luke, eta orobat ezagutza horren atzipena hobetu eta dedukzio-mekanismo ahalmentsuez hornitu. Halako tresna batek denboran zehar etengabeko bilakaera behar du izan, erabiltzailearen jakinminari eutsi eta erabiltzaile horren aurrerapena kontrolatu eta erregistratu ahal izateko. Ikasle-erabiltzailearen ereduaren gaiak are garrantzi handiagoa hartzen du testuinguru honetan.

*Informazio-bilaketako* sistemetan berariaz sortutako thesaurusen erabilera aski ezaguna da. Hizkuntz hiztegitik eratorritako thesaurus bat informazio-bilaketako prozesuetan erabiltzeko posibilitateak ere interesgarria dirudi (Fox *et al.*, 80; Chodorow *et al.*, 88; Evens *ed.*, 88; Fox *et al.*, 88). Informazio testualeko bilaketa-sistema horietako thesaurusei buruz, Krovetz-ek (1989) seinalatzen du indizeak hitzen arabera baino adieren arabera antolatzea komeniko litzatekeela, horrela kontzeptu-thesaurus erabat desanbiguatua lortuz.

Urrunago bada ere, gogoan ditugu orobat aplikazio-domeinua hiztegia izan gabe, HIZTSUAREN arkitektura, eta batez ere ezagutzaren errepresentazio-eredua, erabili ahal izango luketen laguntza-sistemak: esate baterako, *programazio-lengoaia baten laguntza-sistema*, non erreferentzi eskuliburua lengoaiaren kontzeptuei buruzko hiztegi moduko bat baita (Hochon *et al.*, 92).

Beste ideia bat litzateke sistema adituen eraikuntzari dagokiona. Hiztegi-sistemak daukan ezagutza, ez al da bada maiz ezagutza orokorra, mundu errearekin zerikusi handikoa dena? Sistema adituentzat osagarri interesgarria litzatekeen hiztegi-sistema adimenduna dugu buruan.

#### **VI.4.1 Lengoaia naturalaren tratamendurako sistemen osagai lexikalak eraikitzeke iturburua.**

Autore anitzek seinalatzen du LNPko sistemen osagai lexikalak eratzeko hiztegi arruntez baliatzeak duen interesa (Alshawi, 89; Byrd, 90; Wilks *et al.*, 90). Autore horiek, oro har, modu horretan eraikitako sistema lexikalak lexiko laburren osagarri bezala ikusten dituzte, munduari buruzko ezagutza orokorra edukiko luketen sistemak bezala, alegia.

Txostenaren II. kapituluari aritu gara hiztegi-definizioetatik, sakon aztertuz gero, eraz daitezkeen ezagutza desberdinez. LNPko programa batek ezagutza hori erabiltzearen aldetik ikusita, argi dago hiztegi-definizioak inola ere gutxietsi ezin den ezagutza-iturri inportantea direla. Ezagutza horien ustiatzea eta erabilera espezifikoko horri begirako funtzionalitatea zehaztea lirateke ikerkizun geratutako beste bide batzuk.





## VII. Ondorio gisa.

Ikerlan honen emaitza gisa hiztegi-sistema urgazle adimendun bat diseinatu eta sortu dugu. Sistemaren adimena domeinuari buruz duen ezagutzan datza batik bat, hots, hizkuntz lexikoari buruzko hiztegi-ezagutzan. Ezagutza horren gainean ezarritako dedukzio-mekanismoei esker, posible bihurtzen da hasiera batean esplizitu ez zen informaziorik lortzea.

Sistemaren gunea den ezagutzaren iturburua hiztegitxo arrunt bat izan da. Hiztegi horretako ezagutzaren erauzketa, euskarri informatikoan ezarritako hiztegi-bertsioaren gainean —hiztegi datu-basea— egindako prozedura erdiautomatikoaren bitartez burutu da. Prozesu horretan alderdirik azpimarragarriena adieren definizio-esaldien analisisa izan da. Ezagutza-erauzte lan hauetarako metodologia bat zehaztu dugu, prozedura estatistiko eta enpirikoetan oinarritua: metodologia horren helburua hiztegiko definizio-eskemak nolakoak diren ezagutzea da, alegia zeintzuk diren definizio horietan ageri diren egitura sintaktikoak, zeintzuk lexikografoak definizioak idazteko garaian erabili ohi dituen formula espezifikokoak —gutxi edo gehiago estereotipatuak—, eta abar. Metodo honek eraman gaitu hiztegiko definizioen analisi sintaktiko-semantiko bat egitera, definizio horien bitartez ezarri kontzeptuarteko erlazioen errepresentazioa lortzera zuzendua.

Hiztegi-ezagutzaren errepresentaziorako sare semantikoaren eredu hautatu dugu, non nodoak —frameak— kontzeptu edo zentzuen adierazpide baititugu, kontzeptuarteko erlazio lexiko-semantikoaren bitartez elkarlotuak. Ezagutza honen errepresentaziorako arkitektura berezia asmatu dugu, hiru ezagutza-azpibasetan banatuz. Arkitektura horri esker, kontzeptuak eta kontzeptuon arteko erlazioak ezagutza-baseko egituretatik eta hitzetatik bereiz jasotzen dira. Sarearen errepresentazioaren aldetik begiratuta, gure ekarria delako konfigurazio edo sintagma-egiturak errepresentatzeko hartu dugun moduan legoke. Izan ere, konfiguraziook mota-kontzeptu bezala tratatzen baititugu, hierarkia kontzeptualean beroriekin integratuz. Gure iritziz, herentzidun errepresentazio hierarkiko hori egokia da, bai kontzeptuen errepresentaziorako eta bai dedukzio-mekanismoen ezarpenerako ere.

Iturburu-hiztegiko azpimultzo bat hartu eta hiztegi ezagutza-basearen prototipo bat eraiki dugu. Prototipo horren gainean dedukzio-mekanismo sail bat ezarri da. Mekanismo horien artean, nabarmenarazi nahi genituzke hasieran sortutako ezagutza-basearen aberaste bat ekarri dutenak eta delako konposaketa lexikala, zeinek aukera ematen baitu ezagutza inplizitua lortzeko era dinamiko batez, erlazio desberdinen arteko konbinaketan oinarrituz. Prototipoa inplementatzeko ezagutzaren ingeniartzarako ingurune batez baliatu gara, hiztegi-ezagutzaren errepresentaziorako zein dedukzio-mekanismoak ezartzeko; mekanismo

## VII. kapitulua

horien ezarpenerako adimen artifizialean ezagun den zenbait teknika usatu dugu, hala nola prozeduren esleipena, produkzio-erregelak, eta beste. Pentsatzen dugu horrelako tresnak egokiak direla ezagutza lexikalaren errepresentazioak eta berorren ustiatze adimendutsuak dituen arazoei aurre egiteko.

Laguntza-sistema adimendunari bereziki dagokionez, halako sistema batentzat arkitektura diseinatu eta erabiltzaile eta sistemaren arteko komunikazioa nolakoa izango litzatekeen idaroki eta zirriboratu dugu. Sistemaren alderdi funtzionalaz berriz, oinarrizko funtzionalitateen multzoa definitu eta zehaztu dugu, hiztegi-ezagutza basean ezarritako dedukzio-mekanismoetan duela euskarria. Hiztegi-erabilerako funtzionalitate horrek ematen dio erabiltzaileari bestela aurkiezina litzaiokeen informaziorik bilatzeko biderik.

Oinarrizko funtzionalitate honen aberasketa kuantitatibo zein kualitatiboak eramango bide gaitu —sistemaren laguntza-sistema bezalako portaera adimendunagoaz lagundurik— proiektu honen motibazio nagusiena izan den *hiztegi adimendunaren* bidean urrats berririk ematera.

# BIBLIOGRAFIA

## Orokorra

- Ageno A., Castellón I., Martí M.A., Rigau G., Ribas F., Rodríguez H., Taulé M., Verdejo F.. SEISD: An environment for extraction of Semantic Information from on-line dictionaries, *Proceedings 3rd. Conference on Applied Natural Language Processing* (Trento, Italia), 253-254 (poster session). 1992.
- Agirre E., Arregi X., Artola X., Bastarrika J.R., Evrard F., Sarasola K.. *Hiztegi elebarkar bateko definizioen analisisa: patro-i-hierarkiatan oinarritutako analizatzailearen inplementazioa eta lortutako emaitzak*. FISS-I-47.1-LSI-91 Barne-txostena (Donostia). 1991.
- Agirre E., Arregi X., Artola X., Díaz de Ilarraza A., Evrard F., Sarasola K.. *Construction of Phrasal Pattern Hierarchies to Produce a Semantic Representation of Dictionary Definitions*, (EACL'93ra bidalia). 1992.
- Ahlsweide T.E.. A tool kit for lexicon building, *Proc. 23rd Annual Meeting of the ACL* (Chicago), 268-276. 1985.
- Ahlsweide T., Evens M., Rossi K., Markowitz J.. Building a Lexical Database by Parsing Webster's Seventh Collegiate Dictionary, *Advances in Lexicology. Proceedings 2nd Annual Conference of the UW Centre for the New Oxford English Dictionary* (Waterloo, Canada), 65-78. 1986.
- Ahlsweide T.E.. New Techniques for Identifying Relational Structures in Dictionary Definitions, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.
- Allen J.F.. *Natural Language Understanding*. Menlo Park (California): Benjamin/Cummings, 1987.

## Bibliografia

- Alshawi H.. Processing dictionary definitions with Phrasal Pattern Hierarchies, *Computational Linguistics* 13, ns. 3-4, 195-202. 1987.
- Alshawi H.. Analysing dictionary definitions, in B. Boguraev, T. Briscoe eds., 153-169, *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989.
- Alterman R.. A dictionary based on concept coherence, *Artificial Intelligence*, 25, 153-186. 1985.
- Amsler R.A., White J.S.. *Development of a Computational Methodology for Deriving Natural Language Semantic Structures via Analysis of Machine-Readable Dictionaries*, LRC-79-7:1 Report. National Science Foundation. 1979.
- Amsler R.A.. *The Structure of The Merriam-Webster Pocket Dictionary*, PhD. Diss. Computer Science, TR-164. University of Texas, Austin. 1980.
- Amsler R.A.. A Taxonomy for English Nouns and Verbs, *Proc. 19th Annual Meeting ACL*, 133-138. 1981.
- Amsler R.A.. Lexical Knowledge Bases, *Proc. COLING'84* (Stanford Univ.), 458-459. 1984.
- Arango G.. Une approche pour amorcer le processus de compréhension et d'utilisation du sens des mots en langage naturel. Thèse de 3<sup>e</sup> cycle (Paris VI). Publications du Groupe de Recherche Claude François Picard. 1983.
- Arregi X., Artola, X., Díaz de Ilarraza A., Evrard F., Sarasola K.. Aproximación funcional a DIAC: Diccionario inteligente de ayuda a la comprensión, *Boletín SEPLN*, 11, 127-138. 1991.
- Artola X., Evrard F.. Dictionnaire intelligent d'aide à la compréhension, *Actas IV Congreso Int. EURALEX '90* (Benalmádena), 45-57. Barcelona: Biblograph, 1992.
- Basili R., Pazienza M.T., Velardi P.. Computational Lexicons: the Neat Examples and the Odd Exemplars, *Proceedings 3rd. Conference on Applied Natural Language Processing* (Trento, Italia), 96-103. 1992.
- Beck H.W.. A Lexicon Design Based on Theories of Categorization, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.
- Bélec Y.. Des règles expertes pour une méthode applicative d'analyse ou de génération du langage naturel. Thèse. Lab. IRIT, Univ. Paul Sabatier (Toulouse). 1990.
- Binot J.L., Jensen K.. A semantic expert using an on-line standard dictionary, *Proc. IJCAI*, 709-714. 1987.
- Bobrow D.G., Winograd T.. An overview of KRL, a knowledge representation language, *Cognitive Science*, 1, 3-46. 1977
- Boguraev B.. Machine-readable Dictionaries and Research in Computational Linguistics, *Workshop "Automating the Lexicon"* (Grosseto). 1986.

- Boguraev B.. The Definitional Power of Words, *Proc. TINLAP-3* (New Mexico), 11-15. 1987.
- Boguraev B., Briscoe T.. Large Lexicons for Natural Language Processing: utilising the grammar coding system of LDOCE, *Computational Linguistics*, vol. 13, n. 3-4, 203-218. 1987.
- Boguraev B., Levin B.. *Machine Readable Dictionaries: a computational linguistics perspective*. Tutorial (ACL Second Conference on Applied Natural Language Processing. Austin). 1988.
- Boguraev B.. *The Lexicon*. Tutorial (4th Conference ACL, European Chapter. Manchester). 1989.
- Boguraev B., Briscoe T. (eds.). *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989a.
- Boguraev B., Byrd R.J., Klavans J.L., Neff M.S.. From Structural Analysis of Lexical Resources to Semantics in a Lexical Knowledge Base, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989b.
- Brachman R.J.. On the Epistemological Status of Semantic Networks, in N.V. Findler ed., 3-50, *Associative Networks. Representation and Use of Knowledge by Computers*. Orlando (Florida): Academic Press, 1979.
- Brachman R.J., Fikes R.E., Levesque H.J.. *KRYPTON: A Functional Approach to Knowledge Representation*, Fairchild Techn. Report, No. 639. 1983.
- Brachman R.J., Schmolze J.G.. An Overview of the KL-ONE Knowledge Representation System, *Cognitive Science*, 9, 171-216. 1985 (reprinted in Mylopoulos, Brodie eds., 207-229, *Readings in Artificial Intelligence and Databases*. S. Mateo (California): Morgan Kauffman, 1989).
- Brachman R.J., Pigman V., Levesque H.J.. An Essential Hybrid Reasoning System: Knowledge and Symbol Level Accounts for Krypton, in Mylopoulos, Brodie eds., 293-300, *Readings in Artificial Intelligence and Databases*. S. Mateo (California): Morgan Kauffman, 1989.
- Breuker J. (ed.). *EUROHELP. Developing Intelligent Help Systems*. Report on the P280 ESPRIT Project EUROHELP. Kopenhagen, 1990.
- Briscoe T., Copestake A., Boguraev B.. Enjoy the Paper: Lexical Semantics via Lexicology. *Proc. COLING'90* (Helsinki Univ.), Vol. 2, 42-47. 1990.
- Brustkern J., Hess K.D.. The BONNLEX lexicon system, in J. Goetschalckx, L. Rolling eds., 33-40, *Lexicography in the electronic age*. Luxembourg: North-Holland, 1982.
- Byrd R.J., Calzolari N., Chodorow M.S., Klavans J.L., Neff M.S., Rizk O.A.. Tools and Methods for Computational Lexicography, *Computational Linguistics*, vol. 13, ns. 3-4, 219-240. 1987.

## Bibliografia

- Byrd R.J.. Computational Lexicology for Building On-line Dictionaries: the WORDSMITH Experience, in L. Fignoni, C. Peters eds., vol. I, 117-137, *Computational Lexicology and Lexicography (Special Issue dedicated to B. Quemada)*. Pisa: Giardini, 1990.
- Calzolari N., Pecchia L., Zampolli A.. Working on the italian machine dictionary: a semantic approach, in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics (Pisa, 1973)*, vol. 37, 49-69. 1980.
- Calzolari N.. Semantics Links and the Dictionary, *Proc. VIth Int. Conf. on Computers and the Humanities (Raleigh, North Carolina)*, 47-50. 1983.
- Calzolari N.. Machine-readable dictionaries, lexical data bases and the lexical system, *Proc. COLING'84 (Stanford Univ.)*, p. 460. 1984a.
- Calzolari N.. Detecting patterns in a lexical data base, *Proc. COLING'84 (Stanford Univ.)*, 170-173. 1984b.
- Calzolari N.. Structure and access in an automated lexicon and related issues, *Workshop "Automating the Lexicon"* (Grosseto). 1986.
- Calzolari, N., Picchi, E.. Acquisition of semantic information from an on-line dictionary, *Proc. COLING'84 (Budapest)*, 87-92. 1988.
- Calzolari N.. Structure and access in an automated lexicon and related issues, in L. Fignoni, C. Peters eds., vol. I, 139-161, *Computational Lexicology and Lexicography (Special Issue dedicated to B. Quemada)*. Pisa: Giardini, 1990a.
- Calzolari N.. Trends in Computational Lexicography and Natural Language Processing, lecture read at the *X Reunión Anual SEPLN (Donostia)*. 1990b.
- Carpenter B.. Typed feature structures: Inheritance (In)equality and Extensionality, *Proceedings of the First Int. Workshop on Inheritance in NLP (Tilburg, The Netherlands)*, 9-18. 1990.
- Chodorow M.S., Byrd R.J., Heidorn G.E.. Extracting semantic hierarchies from a large on-line dictionary, *Proc. 23rd. Annual Meeting ACL*, 299-304. 1985.
- Chodorow M.S., Ravin Y., Sachar H.E.. A tool for investigating the synonymy relation in a sense disambiguated thesaurus, *Proc. 2nd conference on Applied Natural Language Processing (Austin, Texas)*, 144-153. 1988.
- Chouraqui E., Godbert E.. Représentation des descriptions définies dans un réseau sémantique, *Actes 7ème Congrès Reconnaissance des Formes et Intelligence Artificielle (AFCET-INRIA, Paris)*, vol. 2, 855-868. 1989.
- Copestake A.. An approach to building the hierarchical element of a lexical knowledge base from a machine-readable dictionary, *Proceedings of the First Int. Workshop on Inheritance in NLP (Tilburg, The Netherlands)*, 19-29. 1990.

- Copestake A.. The ACQUILEX LKB: representation issues in semi-automatic acquisition of large lexicons, *Proceedings 3rd. Conference on Applied Natural Language Processing* (Trento, Italia), 88-95. 1992.
- Cruse D.A.. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986.
- Dahlgren K.. *Naive Semantics for Natural Language Understanding*. Norwell (Massachussets): Kluwer Academic Pub., 1988.
- Dailey D.P.. The extraction of a minimum set of semantic primitives from a monolingual dictionary is NP-complete, *Computational Linguistics*, vol. 12, no.4, 306-307. 1986.
- Décary M., Lapalme G.. An Editor for the Explanatory and Combinatory Dictionary of Contemporary French (DECFC), *Computational Linguistics*, vol. 16, no. 3, p. 145-154. 1990.
- Doyle J., Patil R.S.. Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services, *Artificial Intelligence*, 48, 261-297. 1991.
- Dubois J., Giacomo M., Guespin L., Marcellesi C., Marcellesi J.B., Mével J.P.. *Dictionnaire de Linguistique*. Paris: Librairie Larousse, 1973.
- Dyer M.G.. Lexical Acquisition through Symbol Recirculation in Distributed Connectionist Networks, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.
- Evens M.W. (ed.). *Relational models of the lexicon. Representing knowledge in semantic networks*. Cambridge: Cambridge Univ. Press, 1988.
- Evrard F., Pascual E., Virbel J.. *Outils personnalisés de modélisation de l'utilisateur dans la gestion électronique de documents*. Rapport interne, IRIT (Toulouse). 1990.
- Ferrari G.. Dictionnaire automatique et dictionnaire-machine: une hypothèse, in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics* (Pisa, 1973), vol. 36, 257-262. 1980.
- Fillmore C.J.. The case for case, in E. Bach, R.T. Harms eds., 1-88, *Universals in Linguistic Theory*. New York: Holt, Rhinehart and Winston, 1968.
- Findler N.V. (ed.). *Associative Networks. Representation and Use of Knowledge by Computers*. Orlando (Florida): Academic Press, 1979.
- Fox E.A., Nutter J.T., Ahlswede T., Evens M., Markowitz J.. Building a Large Thesaurus for Information Retrieval, *Procs. 2nd. Conference on Applied NLP* (ACL, Austin-Texas), 101-108. 1988.
- Fox M.S., Bebel D.J., Parker A.C.. The Automated Dictionary, *IEEE Computer*, 35-48. July, 1980.

## Bibliografia

- Goetschalckx J., Rolling L. (eds.). *Lexicography in the Electronic Age*. Brussels: North-Holland, 1982.
- Gosselin M.. *Le bon usage des dictionnaires*. Sainte-Foy (Québec): Les Éditions La Liberté, 1989.
- Grau B.. *Analyse et représentation d'un texte d'après le thème du discours*. Thèse. Univ. Paris VI. 1983.
- Grimes J. E.. Denormalization and cross referencing in theoretical lexicography, *Proc. COLING'84* (Stanford Univ.), 38-41. 1984.
- Guckler G.. A dictionary as a database, ?, 205-222. 1978.
- Guenther F., Sabatier P.. Sémantique formelle et représentation des connaissances, *Langages*, no. 87, 103-121. 1987.
- Hartley R., Pilkington R.. Software Tools for Supporting Learning: Intelligent On-Line Help Systems, in P. Ercoli, R. Lewis eds., 39-65, *Artificial Intelligence Tools in Education*. North-Holland: Elsevier, 1988.
- Hartley J.R., Smith M.J.. Question answering and explanation giving, in J.A. Self ed., 338-360, *Artificial Intelligence and Human Learning*. London: Chapman and Hall, 1988.
- Hirst G.. *Semantic interpretation and the resolution of ambiguity*. Cambridge: Cambridge Univ. Press, 1987.
- Hochon J.C., Tayrac P., Evrard F.. MAYDAY: Un système d'aide intelligente pour des applications informatiques, *Actes du Congrès Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID'92, Clermont-Ferrand)*, 541-551. 1992.
- van den Hurk I., Meijs W.. The dictionary as a corpus: analyzing LDOCE's definition-language, *Corpus Linguistics II*, 99-125.
- Ingria R.. Lexical Information for Parsing Systems: Points of Convergence and Divergence, *Workshop "Automating the Lexicon"* (Grosseto). 1986.
- Iris M.A., Litowitz B.E., Evens M.. Problems of the part-whole relation, in M.W. Evens ed., c. 12, 261-288, *Relational Models of the Lexicon*. Cambridge: Cambridge Univ. Press, 1988.
- Jacobs P.S.. Making Sense of Lexical Acquisition, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.
- Jensen K., Binot J.L.. Disambiguating Prepositional Phrase Attachments by Using On-Line Dictionary Definitions, *Computational Linguistics*, vol. 13, ns. 3-4, 251-260. 1987.



- Jensen K., Binot J.L.. Dictionary Text Entries as a Source of Knowledge for Syntactic and Other Disambiguations, *Procs. 2nd. Conference on Applied NLP* (ACL, Austin-Texas), 152-159. 1988.
- Johnson Laird P.N.. Mental models of meaning, in Joshi et al. eds., p. 107, *Elements of discourse understanding*. Cambridge University Press, 1981.
- Kay M.. The Dictionary of the Future and the Future of the Dictionary, in A. Zampolli, A. Capelli eds., 161-174, *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries*. Pisa: Giardini, 1983.
- Kay M.. The Dictionary Server, *Proc. COLING'84* (Stanford Univ.), p. 461. 1984.
- Krovetz R.. Lexical Acquisition and Information Retrieval, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.
- Landau S.I.. *Dictionaries. The Art and Craft of Lexicography*. New York: Scribner, 1984.
- Lenders W.. What's in a lexical entry? The contribution of computers to lexicography, in L. Fignoni, C. Peters eds., vol. II, 45-63, *Computational Lexicology and Lexicography (Special Issue dedicated to B. Quemada)*. Pisa: Giardini, 1990.
- Lesk M.. Automatic Sense Disambiguation Using Machine-readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone, *Proc. of the 1986 SIGDOC Conference*, 24-26. 1986.
- Levesque H.J.. (1986) Knowledge Representation and Reasoning, in Mylopoulos, Brodie eds., 35-51, *Readings in Artificial Intelligence and Databases*. S. Mateo (California): Morgan Kauffman, 1989.
- Levin B. (ed.). *Lexical Semantics in Review*, Lexicon Project Working Papers, no. 1. (MIT). 1985.
- Litkowsky K.C.. Models of the semantic structure of dictionaries, *American Journal of Computational Linguistics*, Mf. 81, 25-74. 1978.
- Litkowsky K.C.. Requirements of text processing lexicons, *Proc. 18th Annual Conference of the ACL* (Philadelphia, Pennsylvania), 153-154. 1980.
- Litkowsky K.C.. On the search for semantic primitives, *Computational Linguistics*, vol. 14, no. 1, p. 52. 1988.
- Marchionini G.. Making the transition from print to electronic encyclopaedias: adaption of mental models, *Int. J. of Man-Machine Studies*, no. 30, 591-618. 1989.
- Markowitz J., Ahlswede T., Evens M.. Semantically significant patterns in dictionary definitions, *Proc. 24th Annual Meeting ACL* (New York), 112-119. 1986.
- Martí M.A., Castellón I.. Gramática para el análisis del diccionario VOX. *Boletín SEPLN*, 10, 123-143. 1990.
- Mel'cûk I.A.. A new kind of dictionary and its role as a core component of automatic processing systems, *T.A. Informations*, no. 2, 3-8. 1978a.

## Bibliografia

- Mel'cûk I.A.. Lexical functions in lexicographic descriptions, *Proc. VIIIth Annual Meeting of the Berkeley Linguistic Society*, 427-444. 1982.
- Michiels A., Noël J.. Approaches to thesaurus production, in J. Horecky ed., 227-232, *COLING'82*. Academia: North-Holland, 1982a.
- Michiels A., Mullenders J., Noël J.. The LONGMAN-LIEGE project, in J. Goetschalckx, L. Rolling eds., 201-210, *Lexicography in the electronic age*. Luxembourg: North-Holland, 1982b.
- Miller G.A.. How to misread a dictionary, *Proc. COLING'84* (Stanford Univ.), p. 462. 1984.
- Miller G.A.. Dictionaries of the Mind, *Proc. 23rd Annual Meeting of the ACL* (Chicago), 305-314. 1985.
- Miller G.A., Fellbaum C., Kegl J., Miller K.. The Princeton Lexicon Project: A Report on WordNet. *Procs. BUDALEX'88* (Budapest), 543-558. 1988.
- Mylopolous J., Levesque H.J.. An Overview of Knowledge Representation, in Brodie, Schmidt eds., 3-17, *On Conceptual Modelling. Perspectives from Artificial Intelligence, Databases and Programming Languages*. New York: Springer-Verlag, 1984.
- Nagao M., Tsujii J., Ueda Y., Takiyama M.. An attempt to computerize dictionary data bases, in J. Goetschalckx, L. Rolling eds., 51-73, *Lexicography in the electronic age*. Luxembourg: North-Holland, 1982.
- Neff M.S., Byrd R.J., Rizk O.A.. Creating and querying lexical data bases, *Proc. 2nd Conference on Applied Natural Language Processing* (ACL, Austin), 84-92. 1988.
- Neff M.S., Boguraev B.K.. Dictionaries, dictionary grammars and dictionary entry parsing, *Proc. 27th Annual Meeting of the ACL* (Vancouver), 91-101. 1989.
- Papegaaij B.C., Sadler V., Witkam A.P.M.. Experiments with an MT-directed lexical knowledge bank, *Proc. COLING'86* (Bonn), 432-434. 1986.
- Patel-Schneider P.F.. What's Inheritance got to do with Knowledge Representation, in M. Lenzerini, D. Nardi, M. Simi eds., 1-11, *Inheritance Hierarchies in Knowledge Representation and Programming Languages*. Chichester (West Sussex): Wiley & Sons, 1991.
- Pazienza M.T., Velardi P.. A structured representation of word-senses for semantic analysis, *Proc. 3rd European Conference ACL* (Copenhague), 249-257. 1987.
- Pigamo F.. *Outils de traitement sémantique du langage naturel*. Thèse. Telecom (Paris). 1990.
- Pustejovsky J.. The semantic representation of lexical knowledge, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.

- Pustejovsky J.. The Generative Lexicon, *Computational Linguistics* 17, 4, 409-441. 1991.
- Quillian M.R.. Semantic Memory, in M. Minsky ed., 227-270, *Semantic Information Processing*. Cambridge (Mass.): MIT Press, 1968.
- Quillian M.R.. The Teachable Language Comprehender: A Simulation Program and Theory of Language, *Communications of the ACM*, 12, 459-476. 1969.
- Ramamoorthy C.V., Shekhar S., Garg V.. Software Development Support for AI Programs, *IEEE Computer*, 30-40. January, 1987.
- Rastier F.. Représentation du contenu lexical et formalismes de l'intelligence artificielle, *Langages*, no. 87, 79-102. 1987.
- Reimer U., Hahn U.. A formal approach to the semantics of a frame data model, *Procs. of IJCAI'83*, 337-339, 1983.
- Reiter, R.. On closed world data bases, in H. Gallaire & J. Minker eds., 55-76, *Logic and Databases*. New York: Plenum, 1978.
- Sager N.. *Natural Language Information Processing*. New York: Addison-Wesley, 1981.
- Schank R.C., Colby M. (eds.). *Computer Models of Thought and Language*. S. Francisco: Freeman, 1973.
- Shortliffe E.H.. *Computer-based medical consultation: MYCIN*. Artificial Intelligence Series. North-Holland: Elsevier, 1976.
- Simmons R.F.. Semantic Networks: Their Computation and Use for Understanding English Sentences, in Schank, Colby eds., 63-113, *Computer Models of Thought and Language*. S. Francisco: Freeman, 1973.
- Simpson J.. The New OED Project, *Procs. First Conf. of the UW Centre for the New Oxford English Dictionary* (Waterloo, Canada), 1-6. 1985.
- Sinclair J.M.. *Looking up. An Account of the COBUILD Project in Lexical Computing*. London: Collins, 1987.
- Slator B.M.. Using Context for Sense Preference, in U. Zernik ed. *Proc. First Int. Lexical Acquisition Workshop* (Detroit). 1989.
- Smith R.N., Maxwell E.. An english dictionary for computerized syntactic and semantic processing systems, in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics* (Pisa, 1973), vol. 36, 303-322. 1980.
- Sowa J.F.. *Conceptual Structures. Information processing in mind and machine*. New York: Addison-Wesley, 1984.
- Sowa J.F.. Using a lexicon of canonical graphs in a semantic interpreter, in M.W. Evens ed., c. 5, 113-137, *Relational Models of the Lexicon*. Cambridge: Cambridge Univ. Press, 1988.

## Bibliografia

- Sparck-Jones K.. *Synonymy and Semantic Classification*, (Doctoral Thesis, Univ. of Cambridge, 1964). Michaelson & Wilks eds., Edinburgh : Edinburgh Univ. Press, 1986.
- Stachowitz A.. Beyond the feasibility study: lexicographic progress, in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics* (Pisa, 1973), vol. 36, 323-335. 1980.
- Starosta S.. Les visages du cas, *Langages*, no. 38, 104-128. 1975.
- Teubert W.. Applications of a lexicographical database for German, *Procs. of COLING'84* (Stanford), 34-37. 1984.
- Tsurumaru H., Hitaka T., Yoshida S.. An attempt to automatic thesaurus construction from an ordinary japanese language dictionary, *Proc. COLING'86* (Bonn), 445-447. 1986.
- Vossen P.. The Structure of Lexical Knowledge as Envisaged in the LINKS project, in J. Connolly, S. Dik eds., chap. 4.3, 177-199, *Functional Grammar and the Computer*. Dordrecht: Foris, 1989.
- Vossen P., Meijs W., den Broeder M.. Meaning and structure in dictionary definitions, in B. Boguraev, T. Briscoe eds., 171-192, *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989.
- Vossen P., Serail I.. *DEVIL: A taxonomy-browser for decomposition via the lexicon*. Techn. Report. Faculty of Arts, Univ. of Amsterdam, 1990.
- Walker D.E.. *Machine-readable dictionaries* (Panel Session), *Proc. COLING'84* (Stanford Univ.), p. 457. 1984.
- Warnesson I., Marcotorchino F.. *Recherche algorithmique de la pertinence synonymique par aggregation de similarités*. Étude no. F-053 (Centre Scientifique IBM-France). Octobre, 1982.
- Warnesson I.. *Linguistics and Data Analysis. Semantic Relations and Aggregation of similarities*. Étude no. F-078 (Centre Scientifique IBM-France). Août, 1984.
- Webber H.R.. Machine-readable components in a variety of information-system applications, *Proc. COLING'84* (Stanford Univ.), p. 463. 1984.
- Weiner E.S.C.. Editing the OED in the Electronic Age, *Procs. Fifth Annual Conf. of the UW Centre for the New Oxford English Dictionary* (Oxford), 23-31. 1989.
- Wilks Y.. *Grammar, Meaning and the Machine Analysis of Language*. London: Routledge & Kegan Paul, 1972.
- Wilks Y., Fass D., Cheng-Ming G., McDonald J., Plate T., Slator B.. A Tractable Machine Dictionary as a Resource for Computational Semantics, in B. Boguraev, T. Briscoe eds., chap. 9, 193-228, *Computational Lexicography for Natural Language Processing*. New York: Longman, 1989.

- Wilks Y., Fass D., Cheng-Ming G., McDonald J.E., Plate T., Slator B.M.. Providing Machine Tractable Dictionary Tools, *Machine Translation*, no. 5, 99-154. 1990.
- Winston M.E., Chaffin R., Herrmann D.. A Taxonomy of Part-Whole Relations, *Cognitive Science*, no. 11, 417-444. 1987.
- Woods W.A.. What's in a link? Foundations for semantic networks, in D.G. Bobrow, A.M. Collins eds., 35-82, *Representation and Understanding*. New York: Academic Press, 1975.
- Yokoyama S., Hanakata K.. Conceptual lexicon using an Object-Oriented language, *Proc. COLING'86 (Bonn)*, 226-228. 1986.
- Yoshida S., Tsurumaru H., Hitaka T.. Man-assisted machine construction of a semantic dictionary for natural language processing, in J. Horecky ed., 419-424, *COLING'82*. Academia: North-Holland, 1982.
- Zampolli A.. The Time, Place and Nature of the Conference, in A. Zampolli, N. Calzolari eds. *Proceedings of the International Conference on Computational Linguistics (Pisa, 1973)*, vol. 36, XIX-XXV. 1980.
- Zampolli A.. Perspectives for an Italian Multifunctional Lexical Database, in A. Zampolli ed., 301-341, *Studies in Honour of Roberto Busa S.J.* Pisa: Giardini, 1987.

## Hiztegiak

- Casares J.. *Diccionario ideológico de la lengua española*. Barcelona: Gustavo Gili, 1988.
- DICOLOGIQUE*. Caen: Memodata Editions, 1991.
- Gougenheim G.. *Le Français Fondamental*. Paris: Didier, 1958.
- Larousse de Poche*. Paris: Librairie Larousse, 1979.
- Le Plus Petit Larousse*. Paris: Librairie Larousse, 1980.
- LONGMAN Dictionary of Contemporary English*. London: Longman Editions, 1981.
- Mel'cûk I.A.. *Dictionnaire explicatif et combinatoire du français contemporain*. Montréal: Les Presses de l'Université de Montréal, 1984.
- Moliner M.. *Diccionario de uso del español*. Madrid: Ed. Gredos, 1983.
- Roget P.M.. *Thesaurus of English Words and Phrases*, 1852.
- West M., Gareth Endicott J.. *An International Readers Dictionary* (2nd. ed.). London: Longman, 1977.