



Filologia eta Geografi-Historia Fakultatea

SINTAXI-SEMANTIKA ELKARGUNEA
ZENBAIT TEORIATAN:
EUSKARAREN EZAGUTZA-BASE
LEXIKO-SEMANTIKORANTZ

Ikaslea: ELISABETE POCIELLO IRIGOIEN
Zuzendaria: BEATRIZ FERNANDEZ FERNANDEZ

Ikerkuntza-aldia gainditzeko lana, 2004ko abenduaren 2a.

Gaien aurkibidea

Gaien aurkibidea	3
1 SARRERA ETA AURKEZPEN OROKORRA	5
1.1 Lanaren motibazioa	5
1.2 Lanaren kokapena	6
1.3 EBLSa definitzeko zailtasunak	7
1.4 Lanaren helburuak eta mugak	8
1.5 Aukeratutako teoria eta lanak	9
1.6 Eduki nagusiak	10
2 METODOLOGIA	11
3 HIZKUNTZALARITZA TEORIKOAN OINARRITUTAKO IKERLANAK	13
3.1 Jackendoff (1990)	13
3.1.1 Sarrera	13
3.1.2 Sintaxi-semantika elkargunearen adierazpena	14
3.1.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	16
3.1.4 Erabilera	17
3.2 Levin (1993)	19
3.2.1 Sarrera	19
3.2.2 Sintaxi-semantika elkargunea	19
3.2.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	20
3.2.4 Erabilera	21
3.3 Pustejovsky (1995)	22
3.3.1 Sarrera	22
3.3.2 Sintaxi-semantika elkargunea	24
3.3.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	25
3.3.4 Erabilera	27
4 HIZKUNTZALARITZA TEORIKO ETA KONPUTAZIONALAREN ERDIBIDEAN DAUDEN IKERLANAK	29
4.1 Lexical Functional Grammar (LFG)	29
4.1.1 Sarrera	29
4.1.2 Sintaxi-semantika elkargunea	32
4.1.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	35
4.1.4 Erabilera	36

4.2	Head-Driven Phrase Structure Grammar (HPSG)	36
4.2.1	Sarrera	36
4.2.2	Sintaxi-semantika elkargunea	39
4.2.3	Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	41
4.2.4	Erabilera	42
5	HIZKUNTZALARITZA KONPUTAZIONALEAN OINARRITUTAKO IKERLANAK	43
5.1	FrameNet	43
5.1.1	Sarrera	43
5.1.2	Sintaxi-semantika elkargunea	43
5.1.3	Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	45
5.1.4	Bestelako erlazio semantikoak	48
5.1.5	Erabilera	49
5.2	WordNet eta WordNet-etik abiatutakoak	51
5.2.1	Sarrera	51
5.2.2	Aditza eta informazio sintaktiko-semantikoa	53
5.2.3	Bestelako erlazio semantikoak	56
5.2.4	Erabilera	57
5.2.5	EuroWordNet	58
5.2.6	The Multilingual Central Repository (MCR)	61
5.3	Volem	63
5.3.1	Sarrera	63
5.3.2	Sintaxi-semantika elkargunea	63
5.3.3	Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa	66
5.3.4	Erabilera	67
6	ONDORIO NAGUSIAK ETA ETORKIZUNERAKO LANAK	69
6.1	Proposamenaren arrazoiak	70
6.2	Gabeziak	71
6.3	Beste lanetako informazioa	71
6.4	Etorkizunerako lana	75
	Irudien zerrenda	77
	Bibliografia	79

1 SARRERA ETA AURKEZPEN OROKORRA

1.1 Lanaren motibazioa

Euskal Herriko Unibertsitateko Informatika Fakultateko IXA taldeak hamar urte baino gehiago daramatza Lengoaia Naturalaren Prozesamenduan (aurrerantzean LNP) lanean. Arlo zabal horren barruan, euskararen gaineko ikerketa aplikatua da gure xede nagusia, eta helburu horrekin orain arte, batez ere, morfologia (Agirre *et al.*, 1992; Aduriz *et al.*, 1994, besteak beste) eta sintaxia (Aduriz *et al.*, 1998; Aranzabe *et al.*, 2003; Aldezabal *et al.*, 2001, besteren artean) landu ditugu. Arlo hauetan lan handia egiteke dagoen arren, hurrengo aurrerapauso garrantzitsua semantika jorratzea da.

Semantika beharrezkoa da hainbat ataza konputazionalan aurrera egin ahal izateko (egitura sintaktikoen desanbiguazioan, hitzen adieren desanbiguazioan, anaforaren ebazpenen eta itzulpen automatikoan. . .). Bestalde, semantikaren informazioa lexikoak jasotzen du. Hori dela eta, uneotan IXA taldean ezagutza-base lexiko-semantiko (aurrerantzean EBLS) baten beharra dago, eta arrazoi horregatik, lan honen egitasmoa euskararen informazio lexiko-semantikoa jasotzen duen ezagutza-base bat proposatzea izango da.

IXA taldean, dagoeneko hasiak gara ezagutza lexiko-semantikoaren ikasketan murgiltzen. Lan horietako batzuk jadanik doktorego-tesiak sortu dituzte, eta beste lan batzuk, berriz, egin bidean dauden doktorego-tesiak dira:

- Euskarako aditzen azpikategorizazioaren azterketa, hiztegi elebakar batean (Arriola, 2000; Arriola *et al.*, 1999) edo corpusetan oinarrituta (Aldezabal *et al.*, 2001; Agirre *et al.*, 2004).
- Euskarako aditzen alternantzien eta klase semantikoaren azterketa (Aldezabal, 2004).
- Aditzen hautapen-murritzapenen eskurapen automatikoa WordNet-en eta corpusetan oinarrituta (Agirre eta Martínez, 2001, 2002)
- Aditzen adieraren desanbiguazioa (Agirre *et al.*, 2001).
- Erlazio lexiko-semantikoaren gauzatze sintaktikoa (Agirre eta Lersundi, 2001)

Lan hauei etekin handiagoa aterako litzaieke formalismo beraren arabera antolatutako ezagutza-base batean egongo balira. Azpimarratzekoa da, bestalde, betebeharrak ez direla IXA taldearenak bakarrik, semantika konputazionala edo hizkuntzaren ulermena burutu nahi duen edozein talderenak baizik.

1.2 Lanaren kokapena

Hizkuntzalaritza teorikoak lexikoari buruz zuen ikuspegia aldatu egin da urteotan, sarrera lexikalaren edukari garrantzi gehiago emanez. Hizkuntzalaritzaren ibilbidean sintaxia izan da aztergai nagusia, eta lexikoari ez zaio agian behar beste arretarik jarri, lexikoa, hizkuntza bakoitzak berezkoa duen ataltzat hartu baita eta erregela edo mekanismo linguistikoez orokortzerik ez duena. Gramatika Sortzailearen hasierako erdua (Chomsky, 1965) eta ondorengo Gobernu eta Uztarduraren Teoria dira aipatutakoaren adibideak (Chomsky, 1987).

Teoria hauei egiten zaien gaitzespen azpimarragarrienetako bat da beraien erregelekin hizkuntzaren egitura orokorrenak bakarrik azal daitezkeela. Gabezia hau eta beste batzuk, gainditzeko asmoz, gramatika aztertzeo ikuspuntua aldatu egingo da denboraz, eta sintaxiaren arloan aritzen diren teoriak norabide lexikalistagoa hartuz joango dira (Hale eta Keyser, 1987; Jackendoff, 1990; Levin, 1993; Pustejovsky, 1995, ...).

Asistimos desde hace un cierto tiempo a un razonable aminoramiento —que no es a la aniquilación— de la idea de que no hay explicación gramatical posible si ésta no se puede formular en reglas estrictas y ordenadas. Con ello llega también el renacer de campos que, por razones próximas a ese punto de vista, habían sido relegados durante un largo período. La concepción de las teorías como sistemas modulares y de principios es buena prueba de lo primero; el retorno al estudio de las palabras en cuanto elementos portadores de un significado, de los conjuntos en que se organizan, de sus relaciones y su papel en la gramática, es signo de lo segundo. (Demonte, 1991, 24. orr.)

Egile hauen ustez, lexikoa *salbuespenen zakua* izatetik, egitura konplexua duen atala izatera pasatu da, bertan sintaxi eta semantikaren arteko hartu-emanak islatzen direlarik.

Hizkuntzalaritza konputazionalak teorikoaren antzeko ibilbidea izan du. Hastapenetan, 1950 eta 1960ko hamarkadetan, sistema konputazional gehienek *jostailuzko* lexikoak lantzen zituzten, oso aplikazio-domeinu zehatzei lotuak eta sarrera kopuru murriztekoak. Askotan zerrenda soilak baino ez ziren izaten. B. Boguraev eta T. Briscoe-k esaterako, hau diote:

Knowledge of words underlies these tasks, yet until very recently dictionaries (or lexicons, as linguists usually call them) for natural language processing systems have by and large been the poor sisters of computational linguistic research. (Boguraev eta Briscoe, 1989, 34. orr.)

Oro har, sintaxia eta erregela gramatikaletan jartzen zituzten beren indarrak. 1970 eta 1980ko hamarkadetan, LNPrekiko¹ interesa areagotzeaz gain, azpimarratzekoa da epe horretan hurbilpen-aldaketa gertatu zela: alderdi lexikoan arreta handiagoa jarri zen — *Generalized Phrase Structure Grammar* (Gazdar *et al.*, 1985) adibidez—, eta lexiko konplexuak proposatzen hasi ziren.

Gaur egun joera lexikalistak dirau, bai hizkuntzalaritza teorikoan bai konputazionalen ere. Ondorioz, erregela gramatikalak erraztera jotzen dute, eta lexikoa da muina.

Hizkuntzalaritza konputazionalaren ikuspegitik, zenbat eta lexiko aberatsagoa izan, hainbat eta emaitza hobek egoten dira ataza konputazionalan. Guretzat, eta Pustejovsky-ren (1993) ildo jarraituz, lexikoa aberatsa da, baldin eta:

¹Txosten honetan laburdurak izen berezi bezala deklinatu dira: LNPK, LNPN eta abar. Izan berezi bezala ez badira deklinatu, ohartarazi egingo dugu. Maileguei dagokionez, bukatzen diren hizkiaren (bokale eta kontsonante) arabera deklinatuko ditugu (esaterako, *wine-ren*, *sing-en*).

- Batetik, sarrera lexikalaren edukia oso landuta badago edo sarrera horri dagokion informazio guztia adierazita badu.
- Eta bestetik, lexikoaren antolaketa oso landuta badago edo lexikoa osatzen duten sarrerak beraien artean harremanekin lotuak badaude.

Lehenengoak, makinari sarrera lexikal bati dagokion informazio guztia informatiko-ki ikasten laguntzen dio, hots, makinak ale lexikal batek adiera zehatz batekin daraman informazio sintaktiko-semanticoko guztia ezagutza-basean jasota izango luke. Bigarrenak, berriz, hizkuntza bera ulertzeko behar diren inferentziak eskaintzen dizkigu, hau da, ale lexikalak unitate handiagoak sortuz elkartzen dira, beraz, makinak unitate konplexu horiek aztertu ahal izateko, ezagutza-basean ale lexikalen arteko harremanak definituak izan behar ditu.

Definizio honengatik ere, EBLs² bat definitzen ez da horren erraza, hurrengo atalean (1.3) ikusiko dugun bezala.

1.3 EBLSa definitzeko zailtasunak

EBLS bat lantzeko orduan zenbait zailtasunekin topatuko gara. Batetik, EBLs-ak egiteko eredu edo formalismoen aniztasuna dago. Ondorioz, hizkuntzalaritza teorikoan eredu ugari proposatu izan dira, (Dowty, 1979; Jackendoff, 1990; Talmy, 1985, besteak beste) baina beraien artean ez dago batasunik, eta batzuetan gainera, bata bestearekin kontraesanean daude.

Hizkuntzalaritza konputazionalen ere proposamenak ugariak dira (Bresnan eta Kaplan, 1982; Fillmore eta Baker, 2001a; Miller, 1985; Kipper *et al.*, 2000, eta abar), askotan fenomeno linguistiko zehatz bati mugatutako EBLs-ak direlako.

Bestetik, definitzen zailak diren fenomeno linguistikoak zehaztu behar dira. Esaterako, ale lexikalak definitzen dituen EBLs bat sortzerakoan, ale lexikalak semantikaren ikuspegitik ere definitzen ari garela pentsatu behar da. Hala ere, ez dago batere argi ale lexikal batek izan behar dituen ezaugarri semanticokoak zeintzuk diren. Izan ere, egun oraindik iritzi ezberdinak daude ale lexikalen izaera semanticoko definitzerakoan: ale lexikalak berezko semantika du ala testuinguru sintaktikoaren eraginaren ondorioz jaso du semantika hori? Eta hori horrela izanda, zer ezaugarri dira ale lexikalean berezkoak eta zeintzuk dira testuinguru sintaktikoaren eraginaren ondorioz sortutakoak?

Hala, semantika eta sintaxiaren arteko bereizketa ez da hain argia, eta gaur egun gauza onartua da bi atal hauen artean harremanik izan badela. Dena den, harreman hori nola gauzatzen den oso arazo eztabaidatua da. Bi maila hauen arteko lotura hori bideratzeko *sintaxi-semantika elkarguneaz* hitz egiten da.

In short, we come to see semantics not as *derived* from syntax, but as an independent generative system correlated with syntax through an interface³. (Jackendoff, 2000, 124. orr.)

²EBLS laburdura “e-be-ele-ese” irakurtzen dela suposatuko dugu, eta izen arrunt bat bezala deklinatuko dugu: EBLs, EBLs, EBLs, EBLs eta abar.

³3.1. atalean ikusiko dugun bezala, nahiz eta Jackendoff-ek bere lanean (1990) sintaxi eta semantikaren arteko harremana onartu, berez, semantika sintaxitik banatzen du, hau da, Jackendoff-entzat semantika sintaxiarekin menpekotasunik gabea izango dela defendatzen du.

Horrela bada, EBLSak ale lexikalen izaera semantikoa definitzerakoan zenbait ikuspegi izan ditzake: semantiko hutsa, sintaktikoa edo sintaktiko-semantikoa. Hortaz, EBLSaren ikuspegiaren arabera sarrera lexikala ezaugarri desberdinekin zehaztua etorriko da.

1.4 Lanaren helburuak eta mugak

Lan honen helburua, euskararako EBLS bat egiteko eredu bat proposatzea izango da, eta horretarako, zenbait eredu edo formalismo aztertu dira. Gorago ikusi dugun bezala (1.3 atalean), EBLSen eraikuntzarako ez dago eredu bat bakarrik, ez hizkuntzalaritza teorikoan ezta konputazionalan ere. Proposamen ugari daude, eta hizkuntzalaritza konputazionalaren kasuan, proposamen hauek arloetan zehar sakabanatuak daude. Egoera honen aurrean, eta ikerlan honen ezinbesteko muga dela-eta, azterketaren esparrua murriztu behar izan dugu.

Sintaxia, semantika eta sintaxi-semantika elkargunea hiztegi baten bitartez azaltzen saiatu diren lan batzuk aztertuko ditugu. LNPrean arloan jorratuak izan direnak interesatzen zaizkigu bereziki, baina askotan hauek lan teorikoetan oinarrituak daudenez, garrantzitsua iruditu zaigu lan teoriko hauen ezagutza ere izatea.

Asko izan dira sintaxi-semantika elkargunearen arloan sakontzen saiatu diren hizkuntzalaritza, eta ikerlan honetarako aukeraketa bat egin behar izan dugu, honen barruan bi ikuspegiak formalismoak sartzen saiatu garelarik, hizkuntzalaritza teorikoa eta konputazionala, alegia. Hala ere, formalismo batzuk ezin dira argi eta garbi ikuspegi baten pean kokatu, eta horregatik, azpimultzoak egin ditugu, modu honetara izendatu ditugunak: *Hizkuntzalaritza teorikoan oinarritutako ikerlanak* (3. atala), *Hizkuntzalaritza teoriko eta konputazionalaren erdibidean dauden ikerlanak* (4. atala) eta *Hizkuntzalaritza konputazionalan oinarritutako ikerlanak* (5. atala).

Lan honen helburua gogoan izanda, eredu edo formalismo hauen azterketarako ondoko metodologia landu eta jarraitu dugu (gero 2. atalean azalduko duguna): lan bakoitzeko *fitxa* bat prestatu dugu, bertan euskarako ezagutza-basea antolatzeke beharrezkoa iruditzen zaigun informazio guztiarekin. Horrela, euskarako EBLSa egiteko interesgarri iruditzen zaigun informazioa mugatuko gara. Honenbestez, eredu bat aztertu dugula diogunean, ez dugu adierazi nahi azterketa sakona egin dugunik, baizik eta euskararako baliagarria dena aurkitu dugula.

Azkenik, euskarako EBLSaren diseinua irizpide batzuetara mugatu dugu, hau da, euskarako ezagutza-basea egiteko aukeratuko dugun EBLS ereduak ondorengo baldintzak bete beharko ditu:

- a) Hizkuntza bere osotasunean adierazten duen EBLSa izan behar du, ale lexikal bakoitzari dagokion adiera, klase semantikoa eta informazio sintaktiko-semantikoa (errol-thematikoak, azpikategorizazioa, hautapen-murriztapenak, funtzio gramatikalak, kategoriak. . .) zehaztuta dituen EBLSa. Dagoeneko esan dugun bezala, hizkuntzalaritza konputazionalaren ikuspegitik, zenbat eta lexiko aberatsagoa izan, hainbat eta emaitza hobekotzen direla ataza konputazionalan.
- b) Ahal dela, teoria edo ikerlan bakar bati lotua ez dagoen EBLS eredu izatea, hau da, beste eredu edo formalismo batzuetatik edan dezakeen EBLSa izatea. Behin eta berriz aipatu dugunez, EBLSaren eraikuntzarako ez dago eredu bat bakarrik, ez hizkuntzalaritza teorikoan ezta konputazionalan ere, eta eredu bakarrik jarraitzen

duen EBLSa mugatzea arriskutsua izan daiteke. Beraz, ahalik eta *deskribatzaileena* den eredia interesatzen zaigu.

- c) Konputazionalki inplementa daitekeen EBLSa izatea, hots, LNPn erabilgarria dena.
- d) Aukeratutako eredu horretatik gertu beste lan konputazionalak egotea, gure EBLSa horien informazioarekin ere aberastu ahal izateko.

1.5 Aukeratutako teoria eta lanak

Aztergai izan ditugun ikerlanak hurrengoak dira:

- *Hizkuntzalaritza teorikoan oinarritutako ikerlanak*: Jackendoff (1990), Levin (1993) eta Pustejovsky (1995). Nahiz eta gehienbat LNPrean arloan jorratuak izan diren ikerlanak interesatzen zaizkigun, alderdi teorikoko lanak ere aukeratu ditugu, EBLSaren arloari buruzko oinarri teorikoa izateko. Hiru ikerlan hauek oihartzun handia izan dute (baita gaur egun ere) hizkuntzalaritza konputazionalan eta haue-tatik abiatuta LNPrako lan ugari egin dira (Dorr, 1997, 1993; Fernández *et al.*, 2002; Buitelaar, 1998; Poznanski eta Sanfilippo, 1993; Saint-Dizier, 1996, eta abar).

- Hizkuntzalaritza teoriko eta konputazionalaren erdibidean dauden ikerlanak: *Lexical Functional Grammar* (LFG aurrerantzean) (Bresnan eta Kaplan, 1982), *Generalized Phrase Structure Grammar* (GPSG aurrerantzean) (Gazdar *et al.*, 1985) eta *Head-Driven Phrase Structure Grammar* (HPSG aurrerantzean) (Pollard eta Sag, 1994). Teoria hauek hizkuntzalaritza teoriko eta konputazionalaren artean kokatu ditugu, zeren oinarri teorikoak badarabiltzate ere, erabilpen konputazionalerantz jotzen hasiak dira-eta.

EBLS eredu hauek, sarrera lexikalean informazio sintaktiko-semantiko ugari dakarte. Beraz, sintaxi-semantika elkargunera nola hurbiltzen diren aztertzea interesgarria iruditu zaigu.

HPSG, GPSGren garapena denez, GPSG zaharkitua geratu da. Hala ere, GPSGren azalpen labur bat ematea gomendagarria iruditu zaigu, HPSGren hastapenen berri izateko.

- Hizkuntzalaritza konputazionalan oinarritutako ikerlanak: *FrameNet* (Fillmore eta Baker, 2001a), *WordNet* (Miller, 1985; Fellbaum, 1998), *EuroWordNet* (Vossen, 1998), *The Multilingual Central Repository* (MCR) (Atserias *et al.*, 2004) eta *Volem* (Fernández *et al.*, 2002). Hauen kasuan, iturri desberdinetan oinarrituta sortutako EBLSak dira, hau da, EBLs baten hiztegi eredu bat landu beharrean, besteen eruedetatik abiatuta beraiena sortu dutenak. Gaur egun, LNPn ikertalde gehienek (nahiz eta beraien ikuspegi teorikoa askotan bat ez etorri) EBLs hauek ezagutu eta erabiltzen dituzte.

1.6 Eduki nagusiak

Txosten hau lau atal nagusitan banatu dugu. Sarrera honen ondoren lan honetako metodologia deskribatzen duen atala dator (2. atalean). Ondoren, ikerlanen azterketarekin hasiko gara. Lehendabizi 3. atalean, hizkuntzalaritza teorikoan kokatu ditugun lanak ikusiko ditugu: Jackendoff (1990), Levin (1993) eta Pustejovsky (1995), (3.3), (3.1), (3.2) eta (3.3) atalak hurrenez hurren.

4. atalean, hizkuntzalaritza teoriko eta konputazionalaren artean kokatu ditugun formalismoez mintzatuko gara, LFG (4.1), GPSG (4.2) eta HPSG (4.2) formalismoez, hain zuzen ere.

5. atalean, LNPrekin harremanetan dauden EBLS batzuk deskribatuko ditugu, hala nola, FrameNet (5.1), WordNet (5.2), WordNet-en ildotik jarraitu duten EuroWordNet (5.2.5) eta MCR (5.2.6), eta azkenik, Volem (5.3).

Amaitzeko, 6. atalean lan hauetatik ateratako ondorio nagusiak eta etorkizunerako eginbeharrak azalduko ditugu.

2 METODOLOGIA

1.4 atalean azaldutako helburuak bete ahal izateko, ondoko metodologia jarraituko dugu. Euskarako EBLSa eraikitzeke, informazio zehatza lortuko dugu lehenago aipatutako lan edo formalismoetatik. Horretarako, helburu jakineko irakurketa egingo dugu lanez lan, hau da, bakoitzetik alderdi zehatz batzuk baino ez ditugu aztertuko, beti ere gure helburuaren mesedetan. Hala, gure lanerako oinarri formala izan daitekeena lortu nahi dugularik, lan bakoitzeko *fitxa* bat prestatu dugu bertan EBLs bat antolatzeke beharrezkoa den informazioarekin.

Fitxan bete beharreko informazioa hurrengo izango da:

a) ***Sarrera:***

Aztergai dugun lan edo formalismoari buruzko informazio orokorrarekin osatuta: teoriaren laburpen bat, bere aipamenak, noiz eta zer motibazioarekin sortu zen, . . .

b) ***Sintaxi-semantika elkargunearen adierazpena:***

Lanen alderdi hau aztertu dugu sakonkiago –ez ahaztu gure egitasmo orokorra euskararako EBLs bat osatzea dela eta bertan informazio sintaktiko-semantikoa jaso nahi dugula. Horrela, bada, ondorengo galderak erantzuten saiatu gara:

- Zer ikuspegitik lantzen du sintaxi-semantika elkargunea?
- Nola adierazten du sintaxia eta semantikaren arteko harreman hori?
- Adierak eta klase semantikoak kontuan hartzen al ditu?

c) ***Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa:***¹

Sintaxi-semantika elkargunearen muina aditza da, honek bere baitan jasotzen baitu perpauseko beste elementuen izaerari buruzko informazio guztia, hala nola:

- Azpikategorizazioa
- Hautapen-murritzapenak
- Errol-thematikoak

Sintaxi-semantika elkarguneari buruz hitz egiterakoan, normala da lan eta formalismo gehienetan aditzean oinarritzea. Hala ere, beste kategoriek era bereko informazio sintaktiko-semantikoa izan dezaketenez, lan batzuetan beraien sarrera lexikalak deskribatzen dituzte. Lan hauetan aditzaz gain, informazio sintaktiko-semantikoa definitzeko gainontzeko kategoria hauek erabiltzen badituzte, guk ere aipatu aipatuko ditugu.

¹Kontuan izanda ikerlan honetan aztertuko ditugun lanak ingeleserako eginak daudela, eta adibideak hizkuntza horretatik hartuak daudela nagusiki, preposizioez hitz egingo dugu gehiagotan postposizioez baino.

Dena den, teoria guztiek ez dute informazio hau berdin erabili, eta erabilera horiek zehazten saiatuko gara lanean zehar.

d) ***Bestelako erlazio semantikoak:***

Sarrera lexikalak beraien artean harremanetan egon daitezke, eta batzuetan harreman horiek informazio semantikoa adierazten dute, hau da, sinonimia, antonimia, meronimia, hiperonimia, hiponimia eta antzekoak. Lan batzuetan horrelako harremanak darabiltzatenez, atal hau egitea interesgarria izan zitekeela iruditu zitzaigun. Hala, puntu honetan ale lexikalen arteko erlazio semantikoei eta erlazio hauek adierazten dutenari erreparatu diogu.

e) ***Erabilera:***

Dagoeneko aipatu dugu (1.4 atalean) implementa daitekeen EBLSa nahi dugula. Hori dela eta, formalismo bakoitzaren implementazioa edo erabilera azpimarratzea garrantzitsua iruditzen zaigu. Horrela, lan bakoitzean beheko galderak erantzuten saiatu gara:

- Lexikorik ba al du? Ikusgarri al dago? Zenbat sarrera ditu?
- Erabilera praktikorik izan al du?

Jarraitutako metodologiaren berri eman ondoren, 3. atalean ikerlan honetan landutako teoriak aurkezteari ekingo diogu.

3 HIZKUNTZALARITZA TEORIKOAN OINARRITUTAKO IKERLANAK

3.1 Jackendoff (1990)

3.1.1 Sarrera

Arestian aipatu dugun bezala (1.2 atalean), Gramatika Sortzailean eta Gobernu eta Uztarduraren Teorian, hizkuntzaren gaitasun sortzailea sintaxiari esker gertatzen da hein handi batean, eta semantika eta fonologia sintaxiaren menpe dauden interpretapen mailak baino ez dira. Hau *sintaktozentrismo*tzat har daiteke, eta Jackendoff-ek bere jardunbidean zehar honi aurka egingo dio gogotik.

The organization proposed (...) eliminates this syntactocentrism. It treats the three levels [syntax, phonology and semantics] as equally “creative”; none is derived from the others. Rather, they are autonomous structures, placed in correspondence with each other by further independent rule components. (Jackendoff, 1990, 19. orr.)

Jackendoff-en lanean (Jackendoff, 1990), lexikoa ikuspegi honi lotua egongo denez, bere lexikoa sortzailea, independentea eta sintaxiarekin menpekotasunik gabea izango da.

Hala eta guztiz ere, Jackendoff ez da lehenengo hizkuntzalaria izan sintaktozentrismoa zalantzan jarri duena, Hale eta Keyser (1987, 1993) izan baitira sintaxia eta semantikaren arteko elkarreagina defendatzen lehenetakoak. Hauen iritziz, kategoria sintaktiko bakoi-tza ezaugarri semantiko batekin dago lotuta, *kategoria nozional* (*notional type*) deitzen dutena; hala nola, izena *entity* kategoria nozioanalarekin, adjektiboa *state*-arekin, preposizioa *interrelation*-arekin eta aditza *event*-arekin¹. Horrela bada, kategoria sintaktikoen konbinazio ezberdinekin erlazio semantiko batzuk sortuko dira. Adibidez, *kausari* dago-kion erlazio semantikoa gertatzeko, *event* kategoria nozionala duten bi egitura sintaktiko (bi aditz, alegia) konbinatu behar dira.

Testuinguru honetan kokatu behar dugu, beraz, Jackendoff-en lan hau (1990). Autore honek adierazpen eredu abstraktu bat proposatzen du: *Egitura Lexikal-Kontzeptuala* (aurrerantzean ELK)². Egitura hau batetik, hainbat primitibo semantiko osatzen da —*TO, FROM, TOWARD, AWAY-FROM, CAUSE, GO, VIA...*— eta bestetik ere, hainbat kategoria kontzeptualez —*Thing, Event, State, Action, Place, Path, Property, Amount*, eta

¹Txosten honetan, autoreek errol-thematiko eta bestelako tasun, primitibo edo ezaugarri semantikoentzat darabilten terminologia gorde dugula azpimarratu beharra dago. Kasu batzuetan euskaratzea ez da lan horren zaila baina beste batzuetan, aldiz, euskarako itzulpena eta berezko terminoa guztiz bat ez datozela susmatzen dugunez, eta txosten osoa irizpide berarekin idatzia egotea komeni denez, berezko ingeleseko terminologia dagoen dagoenean uztea erabaki dugu. Hala ere, ez dugu uste lana ulertzeko irizpide hau oztopo izango denik, termino hauetako gehienak oso ezagunak baitira hizkuntzalaritzaren arloan. Dena den, lan hauetan erabilitako bestelako terminologia asko (adierazpen ereduari ematen dioten izena, adierazpen eredu osatzen duten egiturak, eta abar) itzulita dator, beti ere jatorrizko terminoa zein den zehaztuz.

²Jackendoff-ek *Lexical Conceptual Structure (LCS)* deitzen du.

abar. Primitibo semantikoak kategoría kontzeptual orokor horiei dagozkie. Adibidez, *TO*, *FROM*, *TOWARD*, *AWAY-FROM* eta *VIA* primitiboek, *Path* kategoría kontzeptuala osatzen dute; eta *GO*, *STAY*, eta *CAUSE* primitiboek, berriz, *Event* kategoría kontzeptuala.

3.1.2 Sintaxi-semantika elkargunearen adierazpena

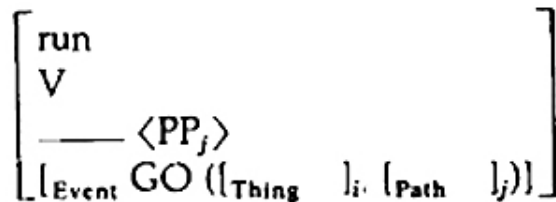
ELK-k³ abantaila handiak ditu sintaxi eta semantikaren arteko harremana adierazterakoan:

- (a) Kategoría sintagmatikoak kategoría kontzeptualekin uztartzen dira, alegia, IS batek *Thing* (*the dog*), *Event* (*the war*) edota *Property* (*redness*) kategoría kontzeptualei erreferentzia egin diezaieke, eta ildo beretik, PS batek, *Place* (*in the house*), *Path* (*to the kitchen*) edota *Property* (*in luck*) kategoría kontzeptualei.⁴
- (b) Primitibo semantikoak aditzaren argumentuei lotzen zaizkie, eta aldi berean, argumentu hauek duten egitura sintaktikoari.

Ikus dezagun adibide batean argiago:

- (1) Egitura sintaktikoa: [P[IS John] AS[ran[PS into[IS the room]]]]
- (2) Egitura kontzeptuala: [Event GO([Thing JOHN][Path TO ([Place IN ([Thing ROOM])]])]

Perpauk honen ELK (*Egitura Lexikal-Kontzeptuala*) perpauseko ale lexikaletatik sortzen da, hots, *John*, *ran*, *into* eta *room*-etik. Perpauseko informazio nagusia aditzak darama, eta honen barruan beste elementuen ELKak egongo direnez, (2) lortzeko nahitaezkoa da aditzaren ELKtik abiatzea (ikus 1. irudia).



1. irudia: *run* aditzaren ELK.

Sarrera lexikal honek *GO* primitiboa darama, eta Jackendoff-ek mugimenduzko egitura kontzeptualak primitibo honekin definitzen ditu⁵. *Run* mugimenduzko aditza izaki, bi argumentu eskatzen ditu: batetik, mugitzen den *gaia* (irudian *Thing*) eta bestetik, mugitzen den horrek egin behar duen *ibilbidea* (irudian *Path*). Lehenengoa *i* azpindize batez

³ELK laburdura “e-le-ka” irakurtzen dela suposatuko dugu eta izen arruntak bezala deklinatuko dugu. Kasu honetan laburdura “k” hizkiak bukatzen denez, eta pluraleko kasuak bereiztearren, horrela deklinatzea erabaki dugu: mugagabeen eta absolutibo singularrean ez du markarik eramango, eta ergatibo singularra, absolutibo plurala eta ergatibo pluralaren artean ezberdintzeko “-k” (ELK-k), “ak” (ELKak) eta “ek” (ELKek) kasu-markak erabili ditugu, hurrenez hurren.

⁴3.1 atalean agertzen diren adibide guztiak Jackendoff-en lanetik (1990) hartuak dira.

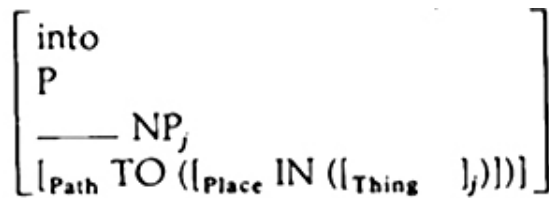
⁵*GO* primitiboa beti egongo da *Event* kategoría kontzeptualean: [EVENT]= [Event GO([Thing],[Path])].

markatuko da (subjektua)⁶ eta bigarrena berriz, j azpindize batez, PSaren (irudian PP) osagarria dela adieraziz. Azken hau, aukerakoa izan arren, lexikoan agertuko da⁷.

Behin *run*-en ELK dugula, 1. iruditik ondoriozta dezakegu:

- (a) Perpaus hau mugimendua adierazten duen perpaus bat dela.
- (b) *John* ale lexikalak perpausaren subjektua izan behar duela, ELKko *GO*ren lehenengo argumentuarekin (*Thing*-ekin, alegia) bat datorrelako.
- (c) Eta azkenik, ELKko *GO*ren bigarren argumentua (*Path*) PS bat (irudian PP) dela.

Path kategoria kontzeptuala preposizioei dagokie. Aditzen ondoren, preposizioa da bere osagarriei buruzko informazio sintaktiko-semantiko gehien daraman egitura sintaktikoa. Hala, bi kategoria hauek azpikategorizazio aberatsenak duten kategoriak dira. Hori dela eta, Jackendoff-ek bereziki aditzak eta preposizioak aztertzen ditu sakonkien. Ikus dezagun *into* preposizioaren ELK:



2. irudia: *into* preposizioaren ELK.

Into-k IS objektu bat eskatzen du bere azpikategorizazioan, eta sarrera honen egitura lexikal-kontzeptualean horixe agertzen da, j azpindizearekin markatua.

(1) eta (2) adibideetara itzuliz, *into*-ren ELKri esker perpausaren bigarren argumentuaren (*into the room*) semantika xeda dezakegu: *Path* kategoria kontzeptualeko *TO* primitiboak *Place* kategoria kontzeptuala hartzen du argumentu gisa, azken hau *IN* primitiboan eta berak hartzen duen *Thing* motako argumentuan (*room*) deskonposatzen delarik.

Esan dezakegu, beraz, lexikoan egitura kontzeptuala eta sintaktikoaren arteko korrespondentzia gauzatzen dela, eta agertu dugun bezala, korrespondentzia hori ale lexikalen sarreretan ageri dela.

Jackendoff (1990) sintaxi-semantika elkargunearen adierazpenaz arduratu zenez, ELKak sortu zituenean arreta berezia jarri zien azpikategorizazioa eskatzen duten ale lexikalei. Beraz, bere ikerlanean ez zuen hizkuntza bere osotasunean aztertu, sintaxi-semantikaren hartu-emanetara mugatu baitzen. Hori dela eta, adiera bigarren mailan dago lan honetan, hots, hitzen anbiguotasuna ez zuen esplizituki kontuan hartu. Hala ere, nahiz eta bere lexikoan adiera ezberdina duten forma berdinetan ez sakondu, inplizituki proposaturiko ELKek forma horiek desanbiguatzeko gaitasuna dute, beraien sarrera lexikalean kategoria kontzeptual, primitibo eta argumentu desberdinak izango dituztelako. Adibidez, *run* aditza *I'll run you home* perpausaren agertzen denean, ez du 1. irudiko adiera

⁶Jackendoff-ek i eta j azpindizeekin subjektu eta objektuen guneak adierazten ditu. (Jackendoff, 1990, 45. orr.)

⁷Txostenean aztertuko ditugun adibide asko lanetatik aukeratutakoak dira. Hauetan autoreek darabilten terminologia agertzen denez, testuan hauen berri emango dugu. Bestalde, kontuan izanda autore hauen lanak ingelesez daudela, hizkuntzalaritzako termino arruntak (kategoriai-eta bezalakoak) adibidean ere ingelesez agertuko dira. Hala, nahiz eta azalpenean euskarako baliokideak erabili, adibideen azalpena ulerkorragoa egin ahal izateko euskarakoaren jarraian, hauen ingeleseko ordaina ere aipatuko dugu.

bera, ezta argumentu-egitura bera ere. Beraz, pentsatzekoa da aditz honek adiera berri honekin, Jackendoff-en lexikoan beste ELK batekin definitua etorriko litzatekeela.

Adierarekin bezala, klase semantikoak ere ez ditu esplizituki lantzen, baina, implizituki ematen du batzuen berri. Esate baterako, zenbait aditz motekin primitibo batzuk zehaztaperen gehiago behar zutela ohartu zen. Hauen artean *ukipen aditzak* (*contact verbs*) aipatzen ditu, ondorengo perpausen irakurketak bereiziz:

(3) Bill ran into the wall (“Billek hormaren kontra jo zuen”)

(4) The cockroach ran into the wall. (“Labezomorroa hormarantz abiatu zen”)

Jackendoff-en ustez, hauen arteko berezitasuna ukipenean dago. Lehenengoak ukipena adierazten du eta bigarrenak ez nahitaez, eta hori *into* preposizioaren *Egitura Lexiko Kontzeptualean* [+/- ukipen] ([+/- *contact*]) batekin adieraziko du.

Dena den, ez du *ukipen aditzek* osatzen duten aditzen zerrenda ematen; Jackendoff-ek klase semantiko horri dagokion aditzen batekin topo eginez gero, nola landu beharko genukeen bakarrik erakusten digu.

3.1.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

Gorago aipatu dugun bezala (3.1.2 atalean), Jackendoff-ek bere lanean (Jackendoff, 1990) bereziki aditzak eta preposizioak aztertzen ditu sakonkien, kategoria hauek baitira azpikategorizazio aberatsenak dutenak.

- Azpikategorizazioa:

Esan bezala, aditzen eta preposizioen azpikategorizazioa da batez ere lantzen duena. Gorago ingeleseko *run* aditza eta *into* preposizioa nola lantzen dituen ikusi dugu (1. eta 2. irudietan), eta bi ELK horietan azpikategorizazioa zehaztua dator. Esaterako, *run* aditzaren azpikategorizazioak bi argumentu eskatzen ditu; bata subjektua izango dena, eta bestea berriz, PS (irudian PP) baten osagarria.

- Hautapen-murritzapenak:

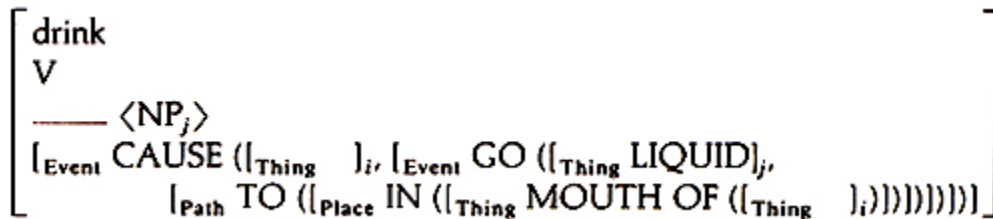
Hautapen-murritzapenak ELKn adierazten ditu, ELKko primitibo eta kategoria kontzeptualetako informazioa baliatuz.

Selectional restrictions evidently are constructed out of a subvocabulary of conceptual structures. That is, the set of possible selectional restrictions is chosen from primitives and principles of combination present in conceptual structure, including not only major conceptual category but also distinctions such as solid versus liquid, human versus animal, and so on. Thus the appropriate linguistic level for stating them is conceptual structure and not syntax. (Jackendoff, 1990, 52. orr.)

Jackendoff-en ustez, hautapen-murritzapenak aditzaren adieraren zati bat dira, eta arrazoi horregatik, aditzaren ELKn adierazita egon behar dute. Hala, aditzaren eta ize-naren ezaugarriak syntaxian bateratzen direnean, ezaugarri hori *betetzen* dela egiaztatzen da.

Har dezagun, Jackendoff-ek berak erabiltzen duen adibidea. *Drink* aditzaren ELK 3. irudian duguna bada, honek *LIQUID* primitiboa eskatzen duela ikusten dugu, hots,

[+liquid] aditz horrek eskatzen duen hautapen-murritzapena da. *Drink* eta *wine* txertatzerakoan (*Harry drank the wine* bezalako perpausetan), primitibo hori desagertu egiten da *wine* izenak ere beregain daramalako [+liquid] informazio hori (hau da, *wine*-k ELKn LIQUID primitiboa darama). Jackendoff-ek erregela sintaktikoen bidez perpausa osatzen denean, ezaugarri hori fusionatu egiten dela esaten du, hots, ezaugarri hori *bete* egiten da aditzaren eta izenaren ezaugarriak syntaxian bateratzen direnean. Horrela, LIQUID hautapen-murritzapenarekin bateragarria ez den ezaugarria daraman ISak ezingo du *drink* aditzarekin joan⁸.



3. irudia: *drink* aditzaren ELK.

Ondorio gisa, esan daiteke hautapen-murritzapena ELKko buruak ematen duen informazio dela (*wine*-ek adibidez), informazio kontzeptual bezala formalizatua.

- Errol-thematikoak:

Jackendoff-ek errol-thematikoak ez ditu primitibo gisa tratatzen. Honen ustez, bakarrik gunea adierazten dute. Esate baterako, ELKn *gaia* (*Thing*) egitura horretako lehenengo argumentua izango da (ikus 3. irudia). Hortaz, Jackendoff-ek errol-thematikoak deitura gisa bakarrik ulertzen ditu.

My claim, therefore, is that the terms *Theme*, *Agent* and so on, are not primitives of semantic theory. Rather, they are relational notions defined structurally over conceptual structure, with a status precisely comparable to that of the notions of Subject and Object in many syntactic theories.
(Jackendoff, 1990, 47. orr.)

3.1.4 Erabilera

Jackendoff-en lexikoaren ezaugarriak (zenbat sarrera dituen, ikusgarri dagoen ala ez, ...) ez ditugu ezagutzen, baina bai ordea, honetatik abiatuta egin diren lexikoena. Esaterako, Dorr-ek, (Dorr, 1993, 1997) Jackendoff-en lanean oinarritutako aditzen EBLS bat sortu zuen eta berarekin tutore sistemak eta itzulpengintza automatikoa landu zituen. Aditzen EBLS horretako aditzak sailkatzeko Levin-en aditz-klaseak (Levin, 1993) erabili zituen eta klase hauek WordNet-eko (Fellbaum, 1998) aditzen adieretara lotuak daude (ikus 3.2.1. atala). Bere txostenetan adierazten denez, erabilitako lexikoak 4432 aditz zituen eta 492 aditz-klase –informazio hau guztia, hurrengo web orrian dago ikusgarri: <http://www.umiacs.umd.edu/bonnie/LCS/Database/Documentation.html>

ELKen inplementazio praktikoa eta konputazional baten adibide bat sakonkiago ikustearren, Dorr-ek itzulpen automatikoa egin duen lana deskribatuko dugu (Dorr, 1993).

⁸Jackendoff-ek *Argument Fusion* erregela erabiltzen du hautapen-murritzapenen eta argumentuen arteko bateragarritasun hau azaltzeko. Argibide gehiago Jackendoff-en lanaren (1990) 53. orrialdean.

Dorr-ek ingelesa, gaztelania eta alemanerako itzulpen-sistema automatikoa egin du (UNI-TRAN) bi oinarri teoriko erabiliz: Gobernu eta Uztardura Teoria eta Jackendoff-en ELKk.

Gobernu eta Uztardurako X marraren bidez itzuli beharreko perpausaren aurkezpen maila adierazten da, hots, azaleko egitura. ELKen bitartez, perpaus horren sakoneko egitura islatzen da, hizkuntzetatik independentea den informazioa, unibertsa (adibidez, “eman” ekintzak, *emaile* bat eskatzen duela). Hala, ELKak hizkuntza guztiek komunean duten informazio jasotzen duenez, hau izango da “jatorrizko hizkuntza” eta “jatorrizko hizkuntzaren itzulpena” lotzen dituen.

Hala ere, itzulpen-sistemak hizkuntza bakoitzak berezkoa dituen informazio eta egiturei ere aurre egin behar die. Horretarako, Gobernu eta Uztarduren hatsarre eta parametroak baliatzen dituzte. Esaterako, Gobernu eta Uztarduraren arabera, ekintza guztiek subjektu kontzeptual bat behar dute, baina hau ez da beti hizkuntza guztietan sintaktikoki islatzen, eta horrelako ezberdintasunak mugatzeko Gobernu eta Uztarduran parametroak darabiltzate, kasu honetan *subjektu isilaren parametroa* deiturikoa. Parametro honen bidez, subjektua nahitaez adierazi behar dutenen eta behar ez dutenen arteko desberdintasuna azaltzen da. Parametroaren mutur batean dauden hizkuntzek (gaztelaniaz, adibidez) ez dute, nahitaez, azaleko egituran subjektu sintaktiko bat eduki behar. Aldiz, parametroaren beste muturrean dauden hizkuntzek (ingelesa eta alemana, esaterako), azaleko egituran, nahitaez, subjektu sintaktiko bat izan behar dute. Beraz, itzulpen-sistema horrelako parametroen arabera definitua dago, hizkuntzen arteko desberdintasunei aurre egin ahal izateko.

Laburbilduz, UNITRAN sistemak ingelesetik gaztelaniara perpaus bat itzultzeko, lehenengo ingeleseko perpausaren analisi sintaktiko automatikoa egiten da (X marraren eremuan), hau ELKko informazioarekin automatikoki parekatzen da⁹ eta baita definitutako parametroen murriztapenak egiaztatu ere. Prozesu automatiko honen emaitza, gaztelaniako itzulpenaren egitura sintaktikoa izango da.

Jackendoff-en ELKen erabilerrari buruz mintzatzean, Volem izeneko proiektua aipatu dezakegu (Fernández *et al.*, 2002). Proiektu honek gaztelaniako, frantseseko eta katalaneko aditz eta preposizioen EBLSa sortu du, aditz eta preposizio bakoitzaren izaera sintaktikoaren deskribapena (azpikategoriazioa, hautapen-murriztapenak eta alternantziak) eta informazio semantikoa (ELK, errol-thematikoak eta aditzen WordNet-eko klase semantiko nagusia) ematen dituelarik. Proiektu honen berri 5.3. atalean emango dugu.

IXA taldean ere ikerlan batzuk egin dira Dorr-en lanetik abiatuta. Agirre eta Lersundiren lanean (2003), interpretapen berdina duten preposizio/postposizio inbentario eleantza sortu dute. Zerrenda honetako preposizio/postposizioak interpretapenaren arabera multzokatuak daude, hau da, hartzen dituzten errol-thematikoen arabera. Gaztelania eta ingeleseko preposizioen inbentarioa eta interpretapenak Dorr-en lanetik hartu dira, eta euskarakoak aldiz, Aldezabalaren ikerlanetik (2004). Dorr-en ELKetak interpretapenak Aldezabalarekin parekatu ondoren, ingeleseko, gaztelaniako eta euskarako preposizio/postposizioen inbentario bakarra lortu dute. Hau oso erabilgarria izan daiteke bai itzulpen automatikorako, bai hizkuntza ezberdinetako preposizio/postposizioen informazio sintaktiko-semantikoari buruz ikasteko.

⁹Parekatze hau egin ahal izateko Jackendoff-en ELKak beraien beharretara eraldatzen dituzte.

3.2 Levin (1993)

3.2.1 Sarrera

Levin-ek bere lanean (Levin, 1993) ingeleseko aditzen sintaxia eta semantika sakonki aztertzen du. Liburuan bertan landutako aditzen zerrenda ematen du, bakoitzaren informazio sintaktiko-semantiko, klase semantiko eta diatesi-alternantziekin batera.¹⁰

Orain arte ikusi ditugun teoretatik pixka bat aldendu egingo da. Levin-ek ez baitu zehazten sarrera lexikalaren itxurak nolakoa izan behar duen; horren ordez, Levin-ek sarrera lexikal hori osatzeko behar diren baliabideak eskaintzen ditu.

This book offers an attempt at delimiting and systematizing the facets of verb behavior. Its contents should help pave the way toward the development of a theory of lexical knowledge. Ideally, such a theory must provide linguistically motivated lexical entries for verbs which incorporate a representation of verb meaning and which allow the meanings of verbs to be properly associated with the syntactic expressions of their arguments. (Levin, 1993, 1go orr.)

Baina, lan hau ez da harremanik gabeko aditzen klase semantiko eta diatesi-alternantzien zerrenda bat bakarrik. 3.2.2 atalean ikusiko dugun bezala, Levin-ek aitzindari den hipotesi bat sortu eta erabili zuen.

3.2.2 Sintaxi-semantika elkargunea

Levin-ek sintaxi-semantika elkargunearen muintzat joko du adiera. Hale eta Keyser; Hale eta Keyser-en lanean (1987; 1993) oinarrituta, aditzaren izaera eta berak hartzen dituen argumentuena, aditzaren semantikari lotuak daudela defendatzen du. Ikuspuntu honetatik abiatuta, Levin-ek bere hipotesia garatuko du: klase semantiko berean dauden aditzek, osagai semantiko berdinak izaki, izaera sintaktiko bera izango dute, eta ondorioz, diatesi-alternantzia berak ere izango dituzte.

Levin-en aburuz, hipotesi honekin esplika daiteke gizakiok dugun gaitasuna edozein aditzi dagokion izaera sintaktikoa esleitzeko, nahiz eta aditz hori lehenengo aldiz entzun. Esate baterako, ingeleseko hiztun batek *sing* aditzak duen semantika ezagutzen badu (*I sing a tune*-ren adibidez), bere izaera sintaktikoa zein izan daitekeen esateko gai izango da:

- (5) You **sing**. [IS + A]
- (6) You **sing** a tune. [IS + A + IS]
- (7) You **sing** me a tune. [IS + A + Izord + IS]
- (8) You **sing** a tune to me. [IS + A + IS + PS]
- (9) You **sing** a tune for me. [IS + A + IS + PS]

Demagun hiztun berak *chant* aditza entzuten duela lehenengo aldiz eta testuinguruari esker *sing*-en adiera antzekoa duela erabakitzen duela. Erabaki horren ondorioz, *sing* eta *chant* klase semantiko berdinean kokatu dituzenez, *sing*-i egokitutako egitura sintaktiko berak erabiliko ditu *chant*-ekin ere:

¹⁰“Diathesis Alternations: alternations in the expressions of arguments, sometimes accompanied by changes of meaning” (Levin, 1993, 2. orr.).

- (10) You **chant**. [IS + A]
 (11) You **chant** a tune. [IS + A + IS]
 (12) You **chant** me a tune. [IS + A + Izord + IS]
 (13) You **chant** a tune to me. [IS + A + IS + PS]
 (14) You **chant** a tune for me. [IS + A + IS + PS]

Hori dela eta, hiztunak aditz bakoitzerako ez du bere izaera sintaktikoa ikasi behar, baizik eta klase semantiko bakoitzerako —kasu honetan *performance verbs* delako klase semantikorako.

Horrela bada, Levin-en teoriaren ardatza adiera eta klase semantikoak dira. Aditz batek bere adiera eta klase semantikoaren baitan definituko ditu bere osagai sintaktikoak. Teoria honen arabera beraz, forma bera baina adiera desberdinak dituen aditz batek, izaera sintaktiko desberdinak izango ditu ere.

Adibidez, ingeleseko *sing* aditza, *performance verbs* klase semantikoari dagokionean, (5), (6), (7), (8) eta (9) adibideetako edozein egitura sintaktikorekin ager daiteke. Aldiz, *sing* aditza, *verbs of sound emission* klase semantikoan dagoenean, beste adiera bat duenez, izan ditzakeen egitura sintaktikoak hurrengoak izango dira:

- (15) A bird **sang** in the trees. [IS + A + PS]
 (16) The trees **sang** with birds. [IS + A + PS]
 (17) In the trees there **sang** the birds. [PS + Adlg + A + IS]
 (18) ...

3.2.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

Levin-ek, bere liburuaren izenburuan ikus daitekeen bezala (*English Verbs Classes and Alternations*), aditzak bakarrik lantzen ditu, izenak eta adjektiboak alde batera utziz. Behin baino gehiagotan aipatu izan dugu, ale lexikalen artean aditzak bereizten direla, hauek sintaxi-semantikaren hartu-emanen duten garrantziarengatik. Arrazoi honexegatik ere, Levin-ek bere lana aditzetan oinarritu du.

- Azpikategoriazioa, hautapen-murritzapenak eta errol-thematikoak:

Nahiz eta aditzak bere baitan jasotzen duen informazioa sakonki landu, informazio hori guztia ez dago adierazpen sintaktiko-semantiko esplizituen bitartez emanda. Hau da, klase semantiko bakoitzean, klase semantikoa osatzen duten aditzen zerrenda ematen du, hauek har ditzaketan egitura sintaktiko guztiekin, eta ondoren, klase horri buruzko informazio sintaktiko-semantikoa idatziz ematen du, adierazpenik erabili gabe. Ikus dezagun liburutik hartutako adibide bat:

LEAVE VERBS:**Class Members:** abandon, desert, leave**Properties:**

- (a) We abandoned the area
- (b) *We abandoned from the area
- (c) Adjectival Passive Participle (some verbs): an abandoned house

Comments: These verbs do not specify a manner of motion; they simply indicate that motion away from a location has taken place. The direct object of these verbs is understood to be the location that has been left. The location cannot be expressed in a prepositional phrase. (Levin, 1993, 264. orr.)

Klase semantiko honen azterketatik deduzi daitezke, *abandon*, *desert*, *leave* aditzek duten azpikategorizazioa (IS + A + IS), hautapen-murriztapenak (objektu zuzenak — adibidean *direct object*— lekua adierazi behar du) eta errol-thematikoak (objektu zuzena *abiapuntua* izango da).

Lehenago esan dugun bezala (3.2.1 atalean), Levin-ek ez du sarrera lexikalaren erre-presentazioa zehaztu beharrean, horretarako behar dituen baliabideak idatziz eskaintzen ditu. Hala ere, baliabide horietatik aditz bakoitzerako informazio hori ateratzea, lan handia eskatzen duen ataza da.

3.2.4 Erabilera

Nahiz eta berez lexikorik ez definitu, liburuan 8 klase semantiko orokor eta 79 diatesi-alternantzia ditugu eskuragarri, eta hauek arrakasta handia izan dute hizkuntzalaritza konputazionalaren arloan, bere sailkapenak aditzen egitura sintaktikoak orokortzen ahalbidetzen baitu.

Levin-en lana erabilia izan da lexiko konputazionalak eraikitzeke, hala nola, *Acquilex* (Poznanski eta Sanfilippo, 1993). Poznanski-k eta Sanfilippo-k ingeleseko diatesi-alternantziak definitu zituzten, ondoren *Acquilex* ezagutza-basean (Boguraev eta Briscoe, 1989) txertatzeko. Azterketa horren abiapuntua Levin-en lana izan zen.

Bestalde, Levin-en lanean oinarrituta itzulpengintza automatikoa ere egin izan da, esate baterako, *UNITRAN* (Dorr, 1993)¹¹. Dorr-ek Levin-en diatesi-alternantzia eta klase semantikoetatik abiatuz, patroi sintaktikoak sortzen ditu, eta horietako patroi bakoitzari Jackendoff-en (1990) ELK bat egokituko zaio gutxienez.

Hauetaz gain, aditzen sailkapen automatikoak lortzeko ere inplementatu da, Saint-Dizier (1996) adibidez. Saint-Dizier-ek Levin-en sailkapen semantikoa frantsesera itzuli eta klase bakoitzerako diatesi-alternantziak definitzen ditu.

IXA taldean ere Levin-en lana erabili da euskal aditzaren azpikategorizazioa jorratzeko (Aldezabal, 2004), nahiz eta lan honetan Levin-en teoriak hutsuneak dituela agerian jarri. Gogora dezagun, Levin-en teoriak dioela diatesi-alternantzia berdina dituzten aditzekin klase semantikoak egin daitezkeela. Baina Aldezabalek teoria honi aurka egiten dizkion adibideak topatu ditu, hau da, Levin-en aditzen klase semantikoak ez dira beti osatzen konpartitzen dituzten alternantzien arabera. Adibidez, Levin-ek *put verbs* eta *remove verbs* klase semantikoak bereizten ditu, beraz, Levin-en teoriaren arabera, klase semantiko bateko eta besteko aditzek diatesi-alternantzia desberdinak izan behar dituzte. Levin-ek, aldiz, bi klase semantiko hauek diatesi-alternantzia berdinekin deskribatzen ditu eta ez

¹¹Ikus 3.1.4 atala.

du azaltzen zergatia. Aldezabalek Levin-en diatesi-alternantzian oinarrituriko teoriaren trinkotasunik eza sakonkiago azaltzen du.

3.3 Pustejovsky (1995)

3.3.1 Sarrera

Pustejovsky-k (1995) *Lexiko Sortzailea* (ingelesez *Generative Lexicon*) proposatzen du. Bere ustetan, teoria honek, aurreko teoriek ez bezala, lexiko semantikaren arazorik nabar-menena saihesten ditu:

- Hizkuntzaren izaera *polimorfikoa* argitzen du.
- Lengoiaren perpausen semantika adierazten du.
- Ale lexikalen izaera sortzailea islatzen du.

Helburua hauek lortzearren, bere teoria funtsekoak diren hiru hatsarretan oinarrituko du:

- Egitura sintaktikoa kontuan hartu gabe, ezinezkoa da semantika lexikalean aurrera egitea. Adiera ezin da bere egituratik banatu.
- Ale lexikalaren adierazpenak errol-thematikoen deskribapena baino zerbait gehiago izan behar du.
- Semantika lexikalak kategoria guztien adierazpen semantikoak landu behar ditu, eta ez aditzena bakarrik.

Hala, egile honen ustez lexikoak sistema kognitiboaren oinarritzko egitura kontzeptuala adierazteko gai izan behar du, eta gainera, ahalmen sortzailea ere izan behar du. Horretarako, deskonposaketan oinarritutako teoria proposatzen du.

I would like to suggest that lexical (and conceptual) decomposition is possible if it is performed *generatively*. Rather than assuming a fixed set of primitives, let us assume a fixed number of *generative devices* that can be seen as constructing semantic expressions. (Pustejovsky, 1991, 417. orr.)

Pustejovsky-ren teorian sarrera lexikalaren deskonposaketa lau adierazpen-mailatan islatzen da. Teorian sakondu ahal izateko, har ditzagun ondorengo adibideak abiapuntu gisa:

(19) Atea irekia dago.

(20) Atea ireki zen.

(21) Jonek atea ireki zuen.

Hiru adibideetako *ireki* aditzaren gutxieneko deskonposaketa *irekita* eta *itxita* aurkakotasuna litzateke. Pustejovsky-k aurkakotasun ezaugarri hau, eta honen antzekoak, **qualia-egituran** (*qualia-structure*) adierazten ditu. Bertan, ale lexikalaren semantika, oinarritzko lau alderditan zehaztu daiteke:

- **Alderdi konstitutiboa (*Constitutive*):** Gauza bat eta bere osagaien arteko harremana adierazten du; adibidez, *gorputz* eta *eskuren* artekoa.
- **Alderdi formala (*Formal*):** Gauzak beraien eremutik ezberdintzen dituen harremana da. Adibidez, badakigu *emakume* eta *gizon* izenak *gizaki* eremuaren barruan kokatzen direla, eta eremu honetan ezberdintzen dituen harremana generoa dela.
- **Alderdi telikoa (*Telic*):** Gauzaren funtzioa eta helburua adierazten dituen harremana da. Adibidez, *ate* ale lexikalaren alderdi telikoa *ireki/itxi* aditzekin zehaztuko litzateke.
- **Alderdi agentiboa (*Agentive*):** Gauza “sorrarazteko” behar diren jatorrizko faktoreak zehazten ditu. Beste hitz batzuekin esanda, adierazitako ale lexikoaren sorrera (demagun *liburu*) ekintza bati (*idatzi*) lotua egongo da.¹².

Esan beharra dago, qualia-egiturako alderdi hauek guztiak sarrera lexikalean beti ez datozela guztiak adierazita.

(19), (20) eta (21) adibideekin jarraituaz, beraien qualia-egiturak ematen dizkigun desberdintasun semantikoen esanahia areagotu egiten da aditzaren aspektua kontuan izanez gero. Esaterako, lehenengoan (19), egoera bat adierazten da (atea irekia dago eta ixtetik irekitzera dagoen prozesuaren berri ez dugu); beste bietan, berriz, itxita/irekita aurkakotasun nabariagoa da, hots, ixtetik irekitzera prozesu bat egon da. Hala ere, prozesu hori ekintza desberdinetan gertatzen da, hau da, (21)ean Jonek irekitzen du atea, eta (20)an atea bakarrik irekitzen da.

Pustejovsky-k ale lexikalen aspektuari buruzko informazio hau guztia ***gertaera-egituran*** (*event structure* delakoan) formalizatzen du. Alderdi honetan, aditzaren zenbait gertaera-mota islatzen dira, eta qualia-egitura bezala, motak bereiz daitezke: **prozesuak** (*process*), **egoerak** (*state*), **trantsizioak** (*transition*).

Pustejovsky-ren aburuz, (19) adibidea egoera bat litzateke, eta (20) eta (21), aldiz, prozesu bat. Adibide hauek, nahiz eta aditz bera izan, ekintza desberdinak adierazten dituzte, eta are gehiago, beraien argumentu-egitura ere desberdintzen da.

Hortaz, ale lexikalaren zati bat sintaxia eta azpikategorizazioari lotua dago. Eta ***argumentu-egitura*** (*argument structure*) izango da honetaz arduratuko dena, hain zuzen ere. Egitura honek ale lexikalaren semantika zehaztuko du; aditzen kasuan adibidez, argumentu-kopurua, argumentu logiko-motak eta beraien gauzapen sintaktikoak.

Azpimarratzekoa da, autorea “argumentu” kontzeptua finkatzen saiatzen dela, tradizionalki izan dituen nahasteak argitzeko helburuarekin. Horrela bada, lau argumentu-mota bereizten ditu¹³:

- **Egiazko argumentuak (*True Arguments*):** sintaktikoki gauzaturiko argumentuak. Adibidez, *John arrived late*.
- **Besterik ezean dauden argumentuak (*Default Arguments*):** qualia-egiturako adierazpen logikoan agertzen diren argumentuak, baina derrigorrez sintaktikoki gauzatu behar ez direnak. Adibidez, *John built the house out of bricks*.

¹²Alderdi hauetako bakoitza ezaugarri gehiagoz osatua dago, operatzaile gisa adieraziak daudenak (ikus adibidez, 24. orriko irudia.).

¹³Adibide hauek guztiak Pustejovsky (1995) lanetik hartuak daude.

- **Ezkatuko argumentuak (*Shadow Arguments*):** lexikoari erantsita dauden argumentuak dira; diskurtsoan, zerbaiti buruz xehetasun gehiago eman nahi denean bakarrik agertzen dira. Adibidez, *Mary buttered her toast with an expensive butter.*
- **Egiazko adjuntuak (*True Adjuncts*):** ale lexikoaren adierazpen logikoan lagun dezaketen argumentuak, baina berez egoeraren adierazle direnak (espazioa eta denbora zehazten dutenak) eta ez dutenak inongo ale lexikaletan parte hartzen. Adibidez, *Mary drove down to New York on Tuesday.*

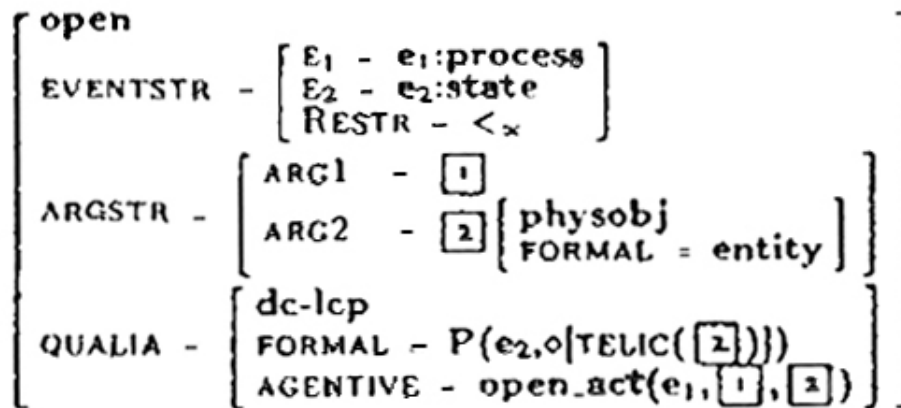
Ale lexikalaren adierazpenak ezin du isolatua egon, baizik eta lexiko osoarekin harremanetan egon behar du; are gehiago sortzailea izan nahi badu. Hori adierazi ahal izateko, *herentzia lexikalaren egitura* (*lexical inheritance structure*) darabil. Qualia-n definitutako operatzaileekin sare bat osatuko du, eta horren bidez, mota eta azpimotak erlazionatzen dira (adibidez, *atxilotu* izenak *ihes egin* aditzarekin *jan* aditzarekin baino zerikusi gehiago duela adierazteko gai da). Horrela, informazioa heredatzeko bidea zabaltzen da, lexikoari egitura antolatua bat emanez, eta ale lexikalaren informazioa sinplifikatuz.

Honenbestez, maila haueetatik abiatuta, Pustejovsky-k hatsarre-multzo bat proposatzen du, transformazio semantikoen bidez qualia-egiturako semantika, beste mailetako informazioarekin murrizteko.

3.3.2 Sintaxi-semantika elkargunea

Lehenago adierazi dugun, bezala, Pustejovsky-rentzat adiera ezin da bere egituratik banatu, hau da, bere aburuz, egitura sintaktikoa kontuan hartu gabe ezinezkoa da ale lexikalaren adierazpena egitea. Nahiz eta autore honen ustez ale lexikalaren adieraren muina qualia-egituran egon, beste egituretako informazioak mugatu egiten du.

Sintaxi-semantika elkargunea nola lantzen duen ulertzeko, bere sarrera lexikal bat ikusiko dugu, (19), (20) eta (21) adibideetan erabili dugun aditzarena (ingelesez) hain zuzen ere (ikus 4. irudia).



4. irudia: *open* aditzaren sarrera lexikoa Pustejovsky-ren teorian.

Ingeleseko aditz honek bi argumentu eskatzen ditu (1 eta 2 zenbakiekin markatuak), eta hauek zehaztuak datoz egitura bakoitzean. Qualia-egiturako (irudian *QUALIA*) *AGENTIVE* ezaugarriak adierazten digu lehenengo argumentuak bigarrena “irekitzen” duela

(irudiko *open act*), eta ekintza hori telikoa (*telic*) dela. Bestetik, argumentu-egituran (irudian *ARGSTR*) bigarren argumentua gauza fisikoa (*physobj*) eta entitate bat (*entity*) dela zehazten zaigu. Eta azkenik, gertaera-egiturari (irudian *EVENTSTR*) esker dakigu bi argumentuek jasaten duten ekintza prozesu bat dela (*process*), eta honen emaitza egoera (*state*) bat izango dela (atea irekia egotea, alegia).

Beraz, argumentu-egitura eta sintaxia erlazioatuak daude; batetik, azpikategorizatu-tako argumentuak aukerazkoak diren ala ez adierazten delako, eta bestetik, argumentuak beste maila guztiekin lotuak daudelako (zenbakien bidez). Honenbestez, teoria honetan sintaxi-semantika elkarguneak oso deskribapen aberatsa du.

Bestalde, nahiz eta autore honentzat adiera oso garrantzitsua izan bere teoriaren antolakuntzarako, ez du erabiltzen klase semantikoak eraikitzeko. Qualia-ako operatzaileekin osatutako sare semantikotik lor daitezke klase semantikoak, baina orokorregiak dira.

3.3.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

Pustejovsky-k hasieratik garbi uzten du bere teorian kategoria guztien adierazpen semantikoak landuko dituela, eta ez aditzena bakarrik.

- Azpikategorizazioa, hautapen-murriztapenak eta errol-thematikoak:

Informazio sintaktiko-semantikoaren adierazpena oso aberatsa da, eta azpikategorizazioa, hautapen-murriztapenak eta errol-thematikoak argi asko agertzen ditu. Ez dugu azpikategorizazioari buruz azalpen gehiago emango, 4. irudian esplikatzerakoan ikusi baitugu nola lantzen duen.

Argumentu-egituran, erabiltzen dituen etiketek aditzak eskatzen dituen tasunen berri ematen dute. Beraz, tasun hauek izango dira aditzaren hautapen-murriztapenak. Esate baterako, 4. iruditik jakin dezakegu *open* aditzak objektu gisa gauza fisikoa eskatzen duela (irudian *physobj* etiketarekin adierazita).

Errol-thematikoei dagokienez, hasieran azaldu dugu Pustejovsky-rentzat errol-thematikoak ez direla aski ale lexikalaren adierazpenerako. Berak errolak erabiltzen ditu baina ez dator bat aurreko teoriak proposaturiko errol-thematikoekin, hauek orokorregiak baitira perpaus baten interpretapen semantikoa hornitzeko.

Autoren honen errolak qualia-egitura gauzatzen dira, egitura hau osatzen duten lau alderdietan. Alderdi bakoitzak bere errolak ditu, hau da, alderdi konstitutiboak *errol konstitutiboa* (*constitutive role*), alderdi formalak *errol formalak* (*formal role*), alderdi agentiboak *errol agentiboa* (*agentive role*) eta alderdi telikoak *errol telikoa* (*telic role*). Bestalde hauetako errol bakoitza balio batzuekin zehaztu daiteke, esate batera, *errol agentiboak artifact* (gizakia sortua) edo *natural kind* (naturalki sortua) bezalako balioak ditu. Pustejovsky-en lanean (1991) ondorengo errol-zerrenda ematen da:

- **Errol konstitutiboa** (***Constitutive Role***): Gauza bat eta bere osagaien arteko harremana adierazten duen errola. Bere balioak hurrengoak dira:
 - (a) Materiala den ala ez (*Material*)
 - (b) Pisua duen ala ez (*Weight*)
 - (c) Zatiak dituen ala ez (*Parts and component elements*)

- **Errol formala (*Formal Role*):** Gauzak beraien eremutik ezberdintzen dituen errola da. Balioak:

- (a) Orientazioa (*Orientation*)
- (b) Magnitudea (*Magnitude*)
- (c) Forma (*Shape*)
- (d) Dimentsioa (*Dimensionality*)
- (e) Kolorea (*Color*)
- (f) Posizioa (*Position*)

- **Errol telikoa (*Telic Role*):** Gauzaren funtzioa eta helburua adierazten dituen harremana da. Bere balioak ondorengoak dira:

- (a) *Egileak* ekintza horrekin egiteko duen asmoa (*Purpose that an agent has in performing an act*)
- (b) Ekintzaren helburua (*Aim that specifies certain activities*)

- **Errol agentiboa (*Agentive Role*):** Gauza “sorrarazteko” behar diren jatorrizko faktoreak zehazten dituen errola.

- (a) Sortzailea (*Creator*)
- (b) Tresna edo gizakiak sortutakoa izatea (*Artifact*)
- (c) Naturalki sortutakoa izatea (*Natural Kind*)
- (d) Katean sortutakoa izatea (*Causal Chain*)

(22) adibidean errol hauen erabilera ikus dezakegu. Bertan *novel* izenaren adierazpena dugu.

- (22) novel (x)
Const: narrative
Form: book (x)
Telic: read (T,y,x)
Agentive: artifact(x), write(T,y,x)

Adierazpen honi esker jakin dezakegu *novel* izena material narratiboa dela (*narrative*), liburu (*book*) itxura duela, *egileak* ekintza horrekin irakurtzeko asmoa duela (*read*) eta idaztearen ondorioz sortutako tresna dela (*artifact, write*).

Azkenik aipatu dezagun, Pustejovsky-k errolak izen eta adjektiboen semantika antolatzeke erabiltzen dituela. Hala, hauek beste egitura sintaktikoen adierazpenetan txertatzean (aditza eta preposizioetan) azken hauen adierak zehazten dituzte, anbiguotasuna murriztuz. Hortaz, izen eta adjektiboen qualia-egiturako ezaugarri hauek, aditz eta preposizioen qualia-egituretan egongo dira ere. Honenbestez, 4. irudiko *open* aditzaren adierazpenean, aditz honek hartzen dituen objektuen qualia-egiturako informazioa ere jasoko du. Beraz, ingeleseko *door* izenaren informazioa *open* aditzean ere egongo da (ikus 4. eta 5. irudiak).

5. irudiak esplizitu egiten du ale lexikal hau gauza fisiko bat dela (irudian *physobj*) eta “ireki” (*aperture*) ekintza batekin agertuko dela testuinguruan. Bestalde, *open* aditzaren adierazpenean, 4. irudian, informazio hau guztia nolabait errepikatua azalduko zaigu: qualia-egiturako *AGENTIVE* ezaugarriaren bitartez lehenengo argumentuak bigarrena “irekitzen” duela, ekintza hori telikoa dela, argumentu-egiturako bigarren argumentua gauza fisikoa dela, eta abar.

Azken finean, hautapen-murritzapenak, errol-thematikoak eta azpikategorizazioa adierazten dituzten etiketa hauek guztiak kategoria guztien sarrera lexikaletan agertuko dira.

$$\left[\begin{array}{l} \text{door} \\ \text{QUALIA} \cdot \left[\begin{array}{l} \text{physobj aperture-lcp} \\ \text{TELIC} = \text{walk_through}(c,z,y) \end{array} \right] \end{array} \right]$$

5. irudia: *door* izenaren sarrera lexikoa Pustejovsky-ren teorian.

3.3.4 Erabilera

Pustejovsky-k hizkuntzalaritza konputazional eta teorikoaren elkar beharra aldarrikatzen du, eta ondorioz, aplikazio konputazionalak izan ditzakeen teoria egiten saiatu zen. Hala ere, implementazioari dagokionez, Pustejovsky-ren lanean oinarritutako ikerlan konputazional gutxi ezagutzen dugu, eta bere teoritik abiatuta lexiko errealik ez dago.

Ezagutzen den inplementazio garrantzitsuenetako bat Buitelaar-ena da (Buitelaar, 1998); bere lanaren ondorioz izenen EBLSa sortu da (CORELEX), 126 klase semantiko eta 40.000 izen inguru dituena. Klase semantiko hauen adierazpena Pustejovsky-ren qualia-egituran oinarritua dago. EBLs honen bidez, eta WordNet-eko (Miller, 1985) adieretatik abiatuta, hainbat “klase polisemiko sistematiko” (*systematic polysemy classes*) definitzen dira, beraien “azpiespezifikazioarekin” (*underspecification*). Hau da, CORELEXen ale lexikal bakoitza “klase polisemiko sistematiko” bati lotzen zaio, eta aldi berean, klase horren “azpiespezifikazioari” ere. Sistematikoki polisemiko diren izenak homonimoak dira, hots, beste hitz batekiko, berdin ahoskatu edo idatzi arren esanahi desberdina dutenak. Hala ere, homonimian gertatzen ez den bezala, forma berdineko adiera desberdin hauek beraien artean harremanetan dauden interpretapenak dira, eta arrazoi horregatik, interpretapen hauek antzekotasunaren arabera multzoka daitezke “klase polisemiko sistematikoak” osatuz.

Adibidez, demagun *banku* izena dugula. Izen honek homonimo ugari ditu, forma berearekin adiera desberdinak adierazten dituelako (‘banketxea’, ‘eserlekua’, ‘lanerako mahaia’ eta abar) eta testuinguruaren arabera adiera bat ala bestea ulertuko dugulako. Baina adiera hauek beraien artean oso desberdinak dira, ez dute inolako antzekotasunik.

Har dezagun orain adibide gisa *liburu* izena. Ale lexikal honen adiera bat ‘gauza fisikoa’ izan daiteke (“Liburua erosi dut” perpausean adibidez), eta beste bat, aldiz, ‘edukia’ (esate baterako, “Liburua interesgarria da” diogunean). Eta *banku* izenarekin ez bezala, adiera hauek beraien artean harremanetan daude, eta horregatik “klase polisemiko sistematiko” bat osatzen dutela esaten da.

CORELEXen *liburu* sarreran bi adiera hauek “klase polisemiko sistematiko” berean multzokatuak daude, hots, *liburu* ale lexikala adiera hauekin “azpiespezifikatua” dator.

Banku izenaren kasuan, aldiz, ale lexikal honek aipatutako adiera bakoitzeko sarrera bat izango luke, sarrera bakoitza bere adierarekin “azpiespezifikatua” egongo delarik. Buitelaar-ek Pustejovsky-ren qualia-egituretan funtsatzen da “azpiespezifikazio” hauek egiterakoan.

CORELEX hurrengo web orrian dago kontsultagarri: <http://www.cs.brandeis.edu/paulb/CoreLex/corelex.html>

4 HIZKUNTZALARITZA TEORIKO ETA KONPUTAZIONALAREN ERDIBIDEAN DAUDEN IKERLANAK

4.1 Lexical Functional Grammar (LFG)

4.1.1 Sarrera

Lexical Functional Grammar (Bresnan eta Kaplan, 1982) (LFG aurrerantzean) formalismo lexikala da eta Gobernu eta Uztardura Teoriaren atalkako egituran¹ oinarritzen da. Dena den, teoria honek Gobernu eta Uztardura Teoriakiko diferentzia nabarmen bat du: ez du mugimendu edo transformaziorik.

The theory is *lexical* and not transformational: it states relations (...) in the lexicon rather than by means of syntactic transformations. (Dalrymple, 2001, 3. orr.)

Teoria honen izenak adierazten duen bezala (*Lexical Functional Grammar* alegia), bere ikerlana subjektu, objektu eta antzeko funtzio gramatikaletan oinarritzen du, hots, hizkuntzen ezaugarri sintaktiko-semanticoen deskribapena hauen bitartez egiten du. Bestalde, eskaintzen dituzten deskribapen hauek unibertsalak direla defendatzen dute. Horrela bada, LFG funtzio gramatikaletan oinarritu eta transformaziorik ez duen teoria lexikalista da. Lexikalismoan egin ohi den moduan, LFG perpausean ager daitezkeen egitura sintaktiko guztiak lexikoan zehazten saiatzen da. Ale lexikalak, besteak beste, ondoko informazioa izango du: funtzio gramatikala, kategoria sintagmatikoak, eduki semantikoa, azpikategorizazioa, errol-thematikoak eta hautapen-murriztapenak.

Horretaz gainera, harremanetan dauden bi adierazpen-sistema erabiliko dira: perpauseko osagai-egitura (*configurational structure*), eta egitura funtzionala (*functional structure*).

Egitura funtzionalak, perpauseko elementuen arteko erlazioen berri ematen du (subjektu eta objektu erlazioak bezalakoak), eta egitura honetan funtzio gramatikalak primitibo bezala tratatzen dira.

LFGk unibertsalak diren funtzio gramatikalen zerrenda ematen du: subjektua (*SUBJ*), objektua (*OBJ*), errol-thematikoa jasotzen duten objektua (*OBJ_θ*) eta zehar objektua (*OBL_θ*), konplementatzailea (*COMP*), subjektua konplementatzaile-sintagmatik kanpo duen konplementatzailea (*XCOMP*), adjunktua (*ADJ*) eta subjektua adjunktua-sintagmatik kanpo duen adjunktua (*XADJ*)².

¹Gobernu eta Uztarduraren Teoria ez da erregela-multzo batez osatutako sistema, baizik eta hatsarre batzuen arabera parametrizatu daitekeen atalkako egitura, hots, gramatika atalka antolatua dago eta hauek hatsarre unibertsalez osatuak daude (Demonte, 1995, 10. orr.).

²Parentesi artean dauden laburdurak LFG teorian erabiltzen direnak dira. Hauek mantendu ditugu geroago azalduko diren adibideetako terminologiarekin bat etortzeko.

Egitura funtzionaleko informazioa matematikako funtzioen antzera adierazten dira, hau da, matematikako matritzetan ezaugarrietatik baliotara doan funtzioa bezala:

$$(23) \quad \begin{array}{l} \text{EZAUGARRIA1} \rightarrow \text{BALIOA1} \\ \text{EZAUGARRIA2} \rightarrow \text{BALIOA2} \end{array}$$

(23) adibideko adierazpen matematikoetan oinarrituta, LFGn 6. irudiaren antzeko adierazpenak egiten dira. Irudi horretan bertan, *David yawned* perpausaren egitura funtzionalaren adierazpena dugu eta honetan hiru ezaugarri ditu: adierari dagokiona (*PRED*), denborari dagokiona (*TENSE*) eta subjektuari dagokiona (*SUBJ*). Eta ondoren bakoitzaren balioa dator zehazturik: ‘*YAWN<SUBJ>*’, *PAST*, eta subjektuaren informazioa adierazten duen beste azpimatriz bat, hurrenez hurren. Azpimatriz honetan bi atributu daude: adierari dagokiona (*PRED*) eta numeroari dagokiona (*NUM*), eta berriro ere, bakoitzaren balioen zehaztapena: ‘*DAVID*’ eta *SG*, hurrenez hurren³.

$$g \left[\begin{array}{ll} \text{PRED} & \text{'YAWN<SUBJ>'} \\ \text{TENSE} & \text{PAST} \\ \text{SUBJ} & f \left[\begin{array}{ll} \text{PRED} & \text{'DAVID'} \\ \text{NUM} & \text{SG} \end{array} \right] \end{array} \right]$$

6. irudia: *David yawned* perpausaren adierazpena LFGn.

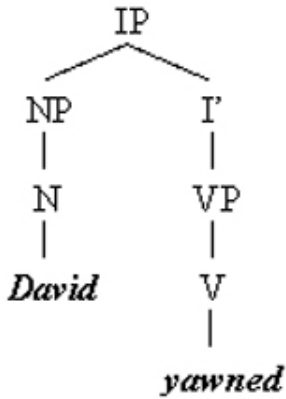
Beste hitz batzuekin esanda, perpausa osatzen duten ale lexikalen deskribapena egiten da. Batetik *yawned* ale lexikalarena: bere adiera *PRED*en adierazita du, eta bertan subjektu baten beharra duela eta iraganean dagoela adierazten zaigu. Eta bestetik, *David* ale lexikalarena: hau subjektuaren informazioa adierazten duen azpimatrizean dator, bere adiera eta numeroarekin.

Osagai-egiturak, perpausaren egitura sintagmatikoa deskribatzen du, hau da, ale lexikalekin sintagmak osatzeko beharrezkoa den antolaketa hierarkiko eta lineala adierazten duen ageriko maila da. Honetarako, X marraren formalismoan oinarritzen dira. Beraz, 7. irudian ikus daitekeen bezala, egitura hau X marran egindako zuhaitz estrukturalak osatzen dute.

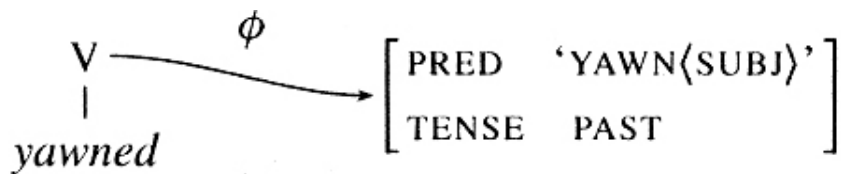
LFGk osagai-egitura eta egitura funtzionalaren informazioa hiztegian uztartu egiten ditu, hau da, osagai-egiturako subjektua osatzen zuten ale lexikalak, eta egitura funtzionaleko *SUBJ* ezaugarriean dagoen informazioa bateratu egiten dute. Osagai-egitura eta egitura funtzionalaren arteko uztardura hau, funtzio batek ahalbidetzen du, \emptyset funtzioak, hain zuzen ere.

8. irudian ikus daitekeen bezala, *yawned* ale lexikala osagai-egiturako aditz bat da (irudian *V* batez adierazita), eta honek egitura funtzionaleko zenbait informazio jasotzen

³Atal honetako adierazpenak Dalrymple (2001) lanetik hartutakoak dira. Bestalde, irudietako laburdurak eta terminologia LFG teoriaran erabiltzen diren bezala mantendu ditugu.



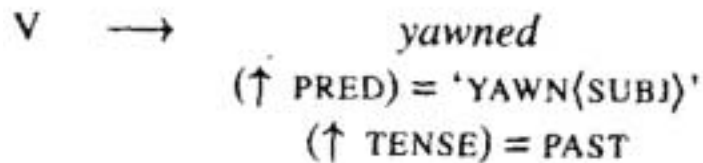
7. irudia: *David yawned* perpausaren osagai-egitura LFGn.



8. irudia: Osagai-egitura eta egitura funtzionalaren arteko uztardura LFGn.

du: bere adiera (irudian *PRED* ezaugarriaren bidez adierazia), subjektu baten beharra duela (irudian *SUBJ*) eta iraganean dagoela (irudian *PAST*).

Uztartutako informazio hau hiztegi-sarreran bertan adierazten dute \uparrow eta \downarrow ikurrekin. Egitura funtzionalari osagai-egituraren informazioa gehitzen diote ikur hauen bitartez, hauek ale lexikalak osatzen duen zuhaitz estrukturalaren sinplifikazioa baitira. \uparrow ikurrak ale lexikala menderatzen duen adabegia adierazten du, eta \downarrow ikurrak, aldiz, ale lexikal horri zer egitura funtzional dagokion. Ingeleseko *yawned* ale lexikalaren adierazpena 9. irudikoa litzateke:



9. irudia: *yawned* ale lexikalaren adierazpena LFGn.

Yawned ale lexikala menderatzen duen lehen adabegia aditza da (*V*) da, eta honi dagokion egitura funtzionalak *PRED* eta *TENSE* ezaugarriak ditu bakoitzak bere balioekin.

Hiztegian definitutako ale lexikoekin, eta ikur hauekin (\uparrow eta \downarrow), perpaus osoen adierazpenak egin daitezke. Horrela, sintagma-egiturak \uparrow eta \downarrow ikurrekin definitzen dituzte.

10. irudia azaltzeko har dezagun irudiko inflexio sintagma (adibideko *IP*) adibide gisa. *IP* sintagma *NP* eta *I'* batez osatzen da. Sintagma-egitura honek informazio gehiago du:

$$\begin{array}{l}
 \mathbf{IP} \longrightarrow \left(\begin{array}{c} \mathbf{NP} \\ (\uparrow \text{SUBJ}) = \downarrow \end{array} \right) \left(\begin{array}{c} \mathbf{I}' \\ \uparrow = \downarrow \end{array} \right) \\
 \mathbf{I}' \longrightarrow \left(\begin{array}{c} \mathbf{I} \\ \uparrow = \downarrow \end{array} \right) \left(\begin{array}{c} \mathbf{VP} \\ \uparrow = \downarrow \end{array} \right) \\
 \mathbf{VP} \longrightarrow \left(\begin{array}{c} \mathbf{V} \\ \uparrow = \downarrow \end{array} \right) \\
 \mathbf{NP} \longrightarrow \left(\begin{array}{c} \mathbf{N} \\ \uparrow = \downarrow \end{array} \right)
 \end{array}$$

10. irudia: Sintagmen adierazpena LFGn.

$\uparrow = \downarrow$ ikurra; honen bidez ikur hau daramanak (I' eta NP kasu honetan) eta hau menderatzen duenak (gure kasuan IP), egitura funtzional bera dutela adierazi nahi da. Hortaz, I' k duen egitura funtzionala IP aren bera izango da, eta NP ak duen egitura funtzionala ere IP aren bera izango da. Honetaz gain, IP ak hartzen duen NP a subjektua dela adierazten dute.

4.1.2 Sintaxi-semantika elkargunea

Orain arte, LFGren alderdi sintaktikoenez mintzatu gara, egitura sintaktikoei erreparatzen dien alderdiaz, alegia. Baina teoria honek argumentu-egituraren informazioa ere landu egiten du. Are gehiago, sintaxiarekin duen harremana zehaztu egiten du errol-thematikoak funtzio gramatikalekin lotuaz.

Bresnan eta Kaplan-ek (1982) sintaxi-semantika elkargunearen aurkezpena ondorengo irudian dugu ikusgarri:

$$\begin{array}{c}
 \text{SUBJ} \quad \quad \quad \text{OBJ} \quad \quad \quad \text{OBLGOAL} \\
 \text{'give} \langle \quad \text{—} \quad , \quad \text{—} \quad , \quad \text{—} \quad \rangle \\
 \text{AGENT} \quad \quad \quad \text{THEME} \quad \quad \quad \text{GOAL}
 \end{array}$$

11. irudia: Sintaxi-semantika elkargunea LFGn (Bresnan eta Kaplan, 1982).

11. irudian ikus daitekeen bezala, *give* aditzak hiru argumentu ditu, eta bakoitzaren errol-thematikoak adierazita datoz. Bestalde, errol-thematiko hauei funtzio gramatikalak esleitzen zaizkie: *egileari* subjektua, *gaiari* objektua eta *helburuari* zehar objektua. Hortaz, Kaplan eta Bresnan-ek funtzio gramatikalak eta errol-thematikoen arteko hartumana egitura funtzionaleko *PRED* ezaugarrian eransten dute.

Aditz batzuen argumentu-egituretan, errol-thematiko bati funtzio gramatikal bat baino gehiago esleitu dakioke, esate baterako, alternantzia pasiboa onartzen duten aditz guztien

kasuan. Alternantzia pasiboa onartzen duen aditzak argumentu-egitura desberdina izango du aditz hori bertsio aktiboan dagoenean eta bertsio pasiboan dagoenean: batetik, bertsio aktiboko objektua bertsio pasiboko subjektua bilakatzen da, eta bestetik, bertsio aktiboko subjektua bertsio pasiboko zehar objektua. Hala, alternantzia pasiboa hartzen duten aditzen errol-thematikoak bi funtzio gramatikalez adieraziak etor daitezke, aditzaren bertsioaren arabera. Horrelako desorekak, erregela lexikoekin erabiltzen dira⁴, hala nola, (24) adibideko erregela lexikoak, alternantzia pasiboa onartzen duten aditzen argumentu-egitura (bertsio aktiborako eta bertsio pasiborako) zehazten du.

(24) $SUBJ \rightarrow OBL_{agent}$
 $OBJ \rightarrow SUBJ$

Bresnan eta Kanerva-k geroagoko lanetan (1989) sintaxi-semantika elkargunearen teoria orokor eta oso bat egiteko erregela lexikoak nahikoak ez direla aldarrikatzen dute, eta elkargune hau lantzeko ikuspuntu berri bat proposatzen dute⁵. Beraien ustetan, $SUBJ$, OBJ , OBJ_{θ} eta OBL_{θ} ⁶, funtzio gramatikalak berezkoak dituzten bi tasunetan deskonposatu daitezke: [+/- murriztu] eta [+/- objektibo] (Bresnan eta Kanerva-k *restricted* eta *objective* deitzen dituzte eta beraien terminologia jarraituaz [+/-R] eta [+/-O] bezala laburtuko ditugu hemendik aurrera).

- [-R] tasuna: edozein errol-thematiko har dezakeen funtzio gramatikala, hots, $SUBJ$ eta OBJ funtzioak.
- [+R] tasuna: errol-thematiko batzuetara mugatuak dauden funtzio gramatikalak, OBJ_{θ} eta OBL_{θ} funtzioak, alegia.
- [-O] tasuna: objektua ez diren funtzio gramatikalak, hots, $SUBJ$ eta OBL_{θ} funtzioak.
- [+O] tasuna: objektuak diren funtzio gramatikalak, hots, OBJ eta OBJ_{θ} funtzioak.

Esandakoaren arabera, 12. irudiko taula osa dezakegu.

	-R	+R
-O	SUBJ	OBL_{θ}
+O	OBJ	OBJ_{θ}

12. irudia: Funtzio gramatikalen tasunak LFGn (Bresnan eta Kanerva, 1989).

⁴Egitura sintaktiko desberdinak erregela lexikal modura adierazten dira, eta horrela, hainbat fenomeno linguistiko deskribatzen dituzte, esaterako, alternantzia pasiboa, flexio erregularreko prozedurak, eratorpena, eta abar. Honenbestez, erregela hauekin orokortasun linguistikoekin berri ematen saiatzen dira.

⁵Teoria honek LFG teorian eragin handia izan zuen arren, formalismo honetan bertan, sintaxi-semantikaren arteko elkarreraginari buruz proposamen ugari egon dira: Zaenen (1993) eta Falk (2001), besteak beste.

⁶Gogora dezagun laburdura hauen esanahia: subjektua ($SUBJ$), objektua (OBJ), errol-thematikoa jasotzen duten objektua (OBJ_{θ}) eta zehar objektua (OBL_{θ}).

Bresnan eta Kanerva-k argumentuen eta funtzio gramatikalen arteko harremana errol-thematikoari [+/-R] eta [+/-O] tasunak esleituaz egiten dute. Adibidez, *egile* errol-thematikoari [-O] tasuna dagokio, hau da, *egile* errola ez da inoiz objektua adierazten duen funtzio sintaktiko batekin adierazia agertuko. Hortaz, eta taulan [-O] tasunaren lerroan adierazia datorren bezala, errol-thematiko hau *SUBJ* edo *OBL_θ* funtzioekin egongo da harremanetan.

Autore hauek *gaia* errol-thematikoari [-R] tasuna esleitzen diote, hots, *gaia* edozein errol-thematiko har dezaken funtzioarekin agertu behar duen errol-thematikoa da, *SUBJ* edo *OBJ*, alegia.

Berezkoak diren tasun hauen sailkapenarekin batera, Bresnan eta Kanerva-k hierarkia thematikoan oinarritutako sailkapena proposatzen dute:

(25) *Egilea* => *Onuraduna* => *Esperimentatzailea* => *Bitartekoa* => *Gaia* => *Lokatiboa*

Aditzaren errol-thematiko bat bestea baino altuago badago hierarkian, zuzenean, [-R] tasuna esleitzen zaio. Aditz horrekin agertzen diren beste errol guztiak, berriz, [+R] bezala sailkatzen dira.

Sailkapen hau berezkoak diren tasunekin batera erabiltzen dira, baina beraien artean ezadostasuna balego, hierarkiako tasuna ez litzateke kontuan hartuko.

Bi sailkapen motak kontuan hartuta, azter dezagun adibide bat:

(26) David selected Chris.

Ingeleseko aditz honek (*select*) bi argumentu hartzen ditu (*egilea* eta *gaia*), beraien berezko tasunak [-O] eta [-R] direlarik, hurrenez hurren. Hortaz, *egileak* *SUBJ* edo *OBLagent* izan beharko du eta *gaia* *SUBJ* edo *OBJ*.

Hierarkian duten lekuari begiratuz gero, *egilea* da altuen dagoena, beraz, honek hartuko du [-R] tasuna berezko tasunarekin bateragarria bada. Tasunekin arazorik ez dagoenez (ikus 12. irudian), ondorioz, *select* aditzaren *egile* errol-thematikoak [-O] eta [-R] tasunak ditu.

Gaia hierarkian beherago dagoenez [+R] bezala sailkatuko da, baldin eta berezko tasunarekin arazorik ez duen sortzen. Kasu honetan, *gaiak* berezko tasun bezala [-R] du eta hau ez denez [+R] tasunarekin bateragarria, azken hau ez zaio esleituko.

Bi sailkapen hauen konbinazioaren emaitza hurrengoa da:

- a) Ingeleseko *select* aditzaren *egile* errol-thematikoa [-O] eta [-R] tasunak dituen funtzio gramatikal batekin adieraziko da: *SUBJ*arekin alegia.
- b) Ingeleseko *select* aditzaren *gaia* [-R] tasuna daraman funtzio gramatikalak adieraziko du, *SUBJ* edo *OBJ* funtzio gramatikalak.

Emaitza hau LFG teoriaren beste bi baldintzek berrikusten dute; bata, perpausak subjektua duela ziurtatzen da (LFGn, *The Subject Condition* deitzen dute), eta besteak funtzio gramatikal berean bi errol-thematiko egotea eragozten du (*Function-Argument Biuniqueness* bezala ezagutzen dena). Horrela, perpausoko bi errol-thematikoak ezin dutenez funtzio sintaktiko berarekin (kasu honetan *SUBJ*arekin) adieraziak egon, bigarren baldintza honek *gaiari* objektu funtzioa esleituko dio.

Horrela bada, Bresnan eta Kanerva-ren teoriak (26) adibideko perpausaren azterketa sintaktiko-semanticoa eskaintzen digu: *egilea* den argumentua (*David*) aditzaren subjektua da, eta *gaia* (*Chris*), berriz, objektua⁷.

Adiera eta klase semantikoei dagokionez, ez dituzte aipatzen. Hala ere, pentsatzekoa da aditz batek, adibidez, beste adiera bat duenean, beste ezaugarri batzuk izango dituela, eta baita beste adierazpen bat ere. Ondorioz, LFGren hiztegian beste sarrera berri bat litzateke. Hala ere, honen inguruko argibiderik ez dugu aurkitu.

4.1.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

LFGko informazio sintaktiko-semanticoekin perpaus osoak adieraz daitezke. Beraz, perpausaren ager daitekeen edozein kategoria sintagmatikoaren informazioa aurkezten dute. Orain arte, aditzaren adibideekin jardun gara batik bat. Beherago beste kategoria sintagmatiko batzuen sarrera lexikalak ditugu ikusgarri:

(27) *student* N (↑ PRED)='STUDENT'
Swedish ADJ [PRED ='SWEDISH']

- Azpikategorizazioa:

Hiztegi-sarreraren muina *PRED* ezaugarria da. Bertan definitzen da sarreraren adiera, baina aditzaren kasuan honen azpikategorizazioaren informazioa ere zehazten da. Noski, informazio hau perpaus osoen zuhaitzak egin ahal izateko ezinbestekoa da. Ez dugu azpikategorizazioari buruz azalpen gehiago emango, sintaxi-semantika elkargunea esplikatzerakoan ikusi baitugu nola lantzen duten.

- Errol-thematikoak:

Sintaxi-semantika elkarguneari buruz hitz egiterakoan, Bresnan eta Kaplan-en (1982) ereduari erreparatu diogu. LFGn egindako geroagoko lanek, beste eredu batzuk hartu dituzte errol-thematikoak deskribatzerakoan, hala nola, autore batzuk (Zaenen, 1993; Alsina, 1993, besteak beste) Dowty-ren proto-errolak (1991) hartzen dituzte abiapuntu gisa (ikus 13. irudia); beste batzuek (Butt, 1996; Broadwell, 1998, eta abar), aldiz, Jackendoff-enak (1990)⁸.

$$\left[\text{PRED 'GIVE} \langle [P-A] [(P-P)] [P-P] \rangle \right]$$

13. irudia: *give* aditzaren adierazpena LFGn proto-errolak erabilia.

13. irudiko LFG sarrera Dowty-ren proto-errolak ditu. Ingeleseko *give* (euskarako *eman*) aditzak hiru argumentu ditu: *proto-agentea* (ematen duena) eta bi *proto-paziente* (jasotzen duena eta ematen den gauza).

⁷Errol-thematikoen hierarkian ez dituzte *helburua*, *iturria* eta beste errol batzuk aipatzen.

⁸Jackendoff-en errol-thematikoak ikerlan honetan azaldu dira, 3.1. atalean.

4.1.4 Erabilera

LFGk inplementazio batzuk izan ditu. Hemen horietako batzuk aipatuko ditugu. Alde batetik, LFG formalismoko egitura funtzionalak erabilia corpus etiketatuak daude, esate baterako Cahill *et al.*-ek (2002) egitura funtzionaleko informazioarekin ingeleseko 100.000 ale lexikal eta 50.000 esaldiko corpusa etiketatu dute erdi-automatikoki. King *et al.*-ek (2003) ere ingeleseko corpus etiketatu bat egin dute, LFG analizatzaile sintaktiko (LNPn *parser* edo *gramatika* deritzoguna) bat erabilia eta ale lexikalen dependentziak ere islatzen dituen: PARC 700 Dependency Bank-a (PARC 700 DEPBANK). Beste corpus bateko 700 perpaus hartuta, beraien analisi sintaktikoa egin dute LFGko *parser* bat erabilia. *Parser* honen irteerak egitura funtzionala eta osagai-egiturako informazioa darama. Ondoren, analisi hauek DEPBANK formatura pasa dituzte automatikoki, hots, perpausoko ale lexikalen dependentziak adierazten dituen formatura (aditza eta honen argumentu edo adjunktuen arteko harremana adierazten duen formatura, alegia). Formatu honek egitura funtzionaleko informazioa bakarrik islatzen du. Azkenik, eskuzko orrazketa eta zuzenketak egin dituzte. Corpus hau beste corpus batzuk etiketatzeko baliatu izan da, baita *parser* eta adieraren desanbiguzio sistemak garatuagoak eraikitze⁹.

LFG ordena libreko hizkuntzak kontuan hartzen dituenek, eta beraz, euskara bezalako hizkuntzetarako prentsaturako formalismoa denez, euskarari aplikatzea sinpleagoa izan da. Horren froga Ruiz *et al.*-ek (1991) LFGn oinarrituta euskararako egindako *parser*-a da.

Horrelako *parser*-ak erabilia itzulpen automatikorako saiakerak ere egin dira, Way (2003) adibidez.

4.2 Head-Driven Phrase Structure Grammar (HPSG)

4.2.1 Sarrera

Head-Driven Phrase Structure Grammar (HPSG aurrerantzean) hizkuntzaren sintaxia eta semantika deskribatzeko formalismoa da. Formalismo honek, *Lexical Functional Grammar* (LFG) eta *Generalized Phrase Structure Grammar* (aurrerantzean GPSG) teoriaren eragin handia jaso du, hortaz, hauetatik abiatutako teoria da. 4.1. atalean LFG formalismoari buruz mintzatu gara eta ondorengo lerrotan, GPSGren aurkezpen labur bat egingo dugu.

GPSG Gerald Gazdar-en lanetik (Gazdar *et al.*, 1985) abiatutako teoria da. GPSG teoriak, oro har, Gobernu eta Uztarduraren X marra teoriaren jarraitzen du, eta LFGn bezalaxe (ikus 4.1. atala), lexikoari ematen zaion garrantzia dela-eta, ez dago transformaziorik. Hala, azaleko egitura adierazteko, maila bakarra proposatzen da.

Así, la GPSG arguye que no sólo es posible contar con un único nivel de representación sintáctica, sino también que al proceder así puede resolver ciertos problemas inveterados cuya descripción requería estipulaciones marginales en el marco de la Gramática transformacional. (Sells, 1989, 76. orr.)

Gobernu eta Uztarduraren eredia jarraituz, kategoriak, ezaugarrien arabera definitzen dira (aditza, esaterako, [-I] eta [+A] ezaugarrien bidez), baina kategoria adierazteko metodologia ezberdina da. Kategoria sintaktikoak ezaugarri-egituren multzoak dira. Ezaugarri-egituraren formatua <ezaugarria, ezaugarri-balio> da, esaterako, IS kategoriaren

⁹PARC 700 Dependency Bank-a <http://www2.parc.com/istl/groups/nlft/fsbank/default.html> web orrian dago eskuragarri.

adierazpena <I, +>, <A, +>, <MARRA, 2> litzateke. Kategoria hau hiru ezaugarri-egituraz osatzen da; lehenengoak izena dela adierazten du, bigarrenak aditza ez dela, eta azkenak X marra eskeman bigarren mailan dagoela. Izena era berdinean definitzen da, baina X marraren maila (MARRA)¹⁰ zehazten duen balioa 0 izango da.

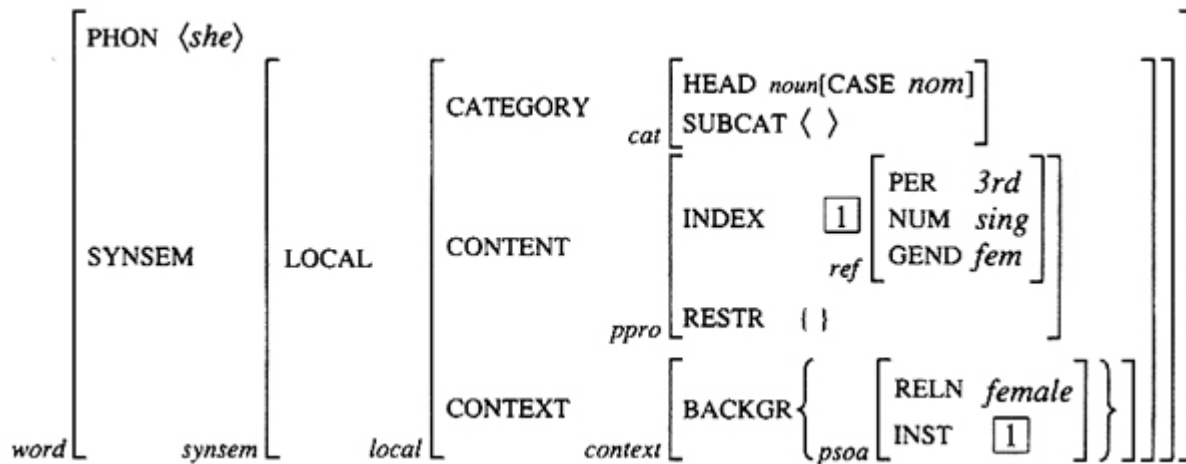
Hauetatik abiatuz GPSG hainbat erregelez osatzen da, batetik ezaugarri hauen konbinazio posibleak mugatzeko, egitura konplexuagoak (sintagmatikoak adibidez) sortzeko, eta azkenik, unibertsalak diren hainbat fenomeno linguistiko (egitura pasiboak adibidez) adierazteko.

Semantikari dagokionez, GPSG Montague-ren ereduari (1974) oinarritzen da: egitura sintaktiko orok interpretapen semantiko bat du. Hala ere, GPSGn sintaxia eta semantikaren arteko harremana estuagoa da Montague-ren teoriaraino baino.

HPSGn (LFGn eta GPSGn bezala) ez dago mugimendu edo transformaziorik; azaleko egitura adierazteko, maila bakarra proposatzen da. Hala ere, ezin da HPSG aurreko bi formalismoekin parekatu, hau aurrekoen garapena baita; alde batetik, hiztegi aberatsagoa du, eta bestetik, aldarrikapen unibertsalagoak lortzen ditu.

HPSGren adierazpenaren muina “zeinuan” (*sign*) datza. Zeinua informazio fonologikoa, sintaktikoa eta semantikoa jasotzen duen unitatea da. Zeinu hauek matematikako antzeko matrizeekin adierazten dira (*attribute-value matrix* deiturikoak) non ezaugarri bakoitzak bere balioa duen. Bestalde, zeinuak ale lexikalak edo sintagmak izan daitezke.

Azter ditzagun matrize hauek astiroago¹¹:



14. irudia: *she* izenordainaren adierazpena HPSGn.

Kasu honetan zeinua ale lexikal bat da (*word*), eta honek zehazten du matrize osoaren izaera, hots, hiztegi-sarrera hau ale lexikal bati dagokio (*she* izenordainari hain zuzen ere). *Word* matrize mota hau bi ezaugarritan banatzen da, *PHON* eta *SYNSEM*¹² deiturikoak. Lehenengoak informazio fonologikoa jasotzen du eta bigarrenak, informazio sintaktiko-semantikoa. *PHON* ezaugarriaren balioa <she> da eta *SYNSEM* ezaugarriarena, berriz, bere izen bera daraman beste matrize konplexu bat: *sysem*¹³. Matrize mota honen balioa

¹⁰GPSGn *BAR* labela erabiltzen dute hau adierazteko.

¹¹Adierazpen guztiak Pollard eta Sag (1994) lanetik hartuak daude.

¹²HPSG teoriarekin zerikusia duten ingeleseko terminologia eta laburdurak errespetatu egin ditugu.

¹³Matrize motak txikiz idazten dira. Beraz, *SYNSEM* ezaugarri bat da eta *sysem* matrize mota bat.

LOCAL da. Balio hau (eta bere aurkakoa *NONLOCAL*) NZ-mugida eta antzeko beste mugimenduekin¹⁴ harremanetan dago. HPSGn horrelako mugimenduek *NONLOCAL* ezaugarria daramate. *LOCAL* ezaugarriaren balioa, berriro ere, matrize konplexu bat da (*local*), eta hau hiru ezaugarritan banatzen da¹⁵: *CATEGORY*, *CONTENT* eta *CONTEXT*. *CATEGORY* ezaugarriak, hitzaren kategoria adierazteaz gain, honek eskatzen dituen argumentuak ere zehazten ditu. *CONTENT* eta *CONTEXT* ezaugarriak hitzaren interpretapen semantikoa zehazten dute. Hauek bere baitan beste hainbat balio eta ezaugarri hartzen dituztenez, banan-banan aztertuko ditugu.

CATEGORY ezaugarriaren balioa *cat* matrize mota da, eta bertan, bi ezaugarri daude: *HEAD* eta *SUBCAT*. *HEAD* ezaugarriak hitzaren kategoria adierazten du, hortaz, bere balioa kategoria bat izan beharko du. Adibide honetan *she* izena da (adibidean *noun*). Kategorien arabera, zenbait matrize mota osatu dira, bakoitzak bere ezaugarri eta balioekin: izenek *CASE* ezaugarria, postposizioek *PREPOSITION-FORM* ezaugarria, besteak beste¹⁶. 14. irudian, *HEAD* nominatiboan (*nom*) dagoen izena (*noun*) dela adierazten da.

Bestalde, *SUBCAT* ezaugarriaren bitartez ale lexikalaren azpikategorizazioa zehazten da. Hitzak ez badu azpikategorizaziorik eskatzen (aztertzen ari garen adibidean bezala), ezaugarri honen balioa hutsik uzten da¹⁷.

CONTENT ezaugarrian ale lexikalaren irakurketa semantikoa zehazten da, batez ere erreferentziaren ikuspuntutik. Hau da, hemen jasoko da ale lexikalak adierazten duen “egoera” perpaus osoaren osotasunetik ikusita.

HPSGko semantika *Situation Semantics* teorian oinarritua dago (Barwise eta Perry, 1983). Teoria honen ideia nagusia ondorengoa da: gure pentsamendu eta ekintzetan hainbat gauza eta ezaugarri darabiltzagu errealitatearen zati eta alderdiak (“egoerak” edo *situations*) zeharka ala zuzenean sailkatzeko. *Situation Semantics* teorian semantika aztertzen da “egoera” moten (*situation types*) arteko harremana balitz bezala. Har dezagun beheko perpausa adibide gisa:

(28) X izpietako aparatuak Mirenek eskua hautsi duela erakutsi du.

Perpaus osoa “egoera” konplexu bat da. Alde batetik, “egoera” bat dago X izpietako aparatuari zenbait ezaugarri esleitzen dizkiona. “Egoera” hau Mirenen eskuarekin lotua dago: Mirenen eskuari egin diote erradiografia. X izpietako aparatua pertsona baten eskuarekin erabiltzen denean eta X izpietako aparatuak dituen ezaugarriak agertzen direnean, pertsona batek eskua hautsia duen “egoera” islatzen da¹⁸. Hortaz, HPSGko *CONTENT* ezaugarria *Situation Semantics* teoriaren ikuspuntuaren ildotik sortutako ezaugarria da.

Ezaugarri honetan, berriro ere, kategorien arabera, matrize mota desberdinak osatzen dira, bakoitzak bere ezaugarri eta balioekin, esate baterako, aditzen kasuan errol-tematikoa hemene zehazten dira, eta izenen kasuan izenordainak eta izenordainak ez direnatan banatzen dira, eta hauen barruan sailkapen gehiago daude (pertsona-izenordainak,

¹⁴Gobernu eta Uztarduraren Teorian KonpS baten espezifikatzailera egiten diren mugimenduak.

¹⁵*NONLOCAL* ezaugarriak ez du horrelako azpimatrizerik.

¹⁶Ezaugarrien kopurua handia denez, batzuk bakarrik aipatuko ditugu. Argibide gehiagorako jo Pollard eta Sag (1994) lanera.

¹⁷Azpikategorizazioari buruz sakonduko dugu 4.2.2 atalean.

¹⁸Argibide gehiago Barwise eta Perry (1983) liburuan.

bihurkariak, anaforak eta abar). Irudiko adibidean *ppro* matrize mota dugu, hots, pertsona-izenordaina. Bestalde, *ppro* matrize mota honek beste bi ezaugarri ditu: *INDEX* eta *RESTRICTION* (*RESTR*).

INDEX izenen ezaugarria da¹⁹, erreferentzia marka bezala erabiltzen dute, “egoera” bat aztertzeo ikuspuntua markatzen duena²⁰. *INDEX* ezaugarriak matrize mota batzuk har ditzake: batetik, izen erreferentzialak (*ref*) eta bestetik, ingeleseko *there* eta *it* bezalako espletiboak. Adibideko *she* izenordaina *ref* motakoa da, are gehiago, singularreko 3. pertsona femeninoa dela zehazten zaigu.

Beste ezaugarria *ppro* matrize motan *RESTRICTION* (*RESTR*) da. *INDEX* ezaugarriko semantikaren murrizketa bat erakusten du; adibidez, entitate linguistiko bat erreferentziadun espresioa bada (demagun *liburu*)²¹, *INDEX*eko azpindizearen erreferentzia entitatea bera izan behar da (liburua bera, izaki bezala), eta erreferente hau *RESTRICTION* ezaugarrian adierazten da beste azpindize baten bidez. Ezaugarri hau izen arruntekin agertzen da, horregatik, hutsik dago aztergai dugun adibidean.

CATEGORY eta *CONTENT* ezaugarriak aztertuta, azkenik *CONTEXT* ezaugarria azaltzea geratzen zaigu. Hemen hitzaren interpretapen semantikoa zehazten da, baina ez *CONTENT*ean egiten zen bezala:

CONTENT values represent contributions to literal (truth-conditional) meaning, while CONTEXT values represent conditions (...) that correspond to presuppositions or conventional implicatures. (Pollard eta Sag, 1994, 27. orr.)

CONTENT ezaugarria azaltzerakoan, erreferentzia errealitateko izaki batekin lotua egon behar duela esan dugu. *CONTEXT* ezaugarria, aldiz, auresuposizioetan oinarritzen da. Adibidez, 14. irudiko adibidera itzuliz, *CONTEXT*ean auresuposatzen da ingeleseko *she* izenordainari dagokion erreferentzia genero femeninoa duela. Frantseseko *elle* izenordainak, ordea, beste *CONTEXT* ezaugarri bat izango du, frantseseko *elle* izenordainaren erreferentzia ez baita nahitaez genero femeninokoa izan behar.

Izenaren adibide bat hartuta, HPSGko zeinua nola osatzen den azaltzen saiatu gara. Noski, hemen ikusitako ezaugarriak ez dira formalismo honek erabiltzen dituen guztiak, baina teoria honek darabilen adierazpenaren ikuspegi orokor bat izateko nahikoa izan daiteke.

Zeinu hauek zenbait hatsarre eta erregelarekin daude hartu-emanetan hauetatik sintagma eta perpaus berriak osatzeko. 15. irudian sintagma baten sarrera lexikala dugu.

4.2.2 Sintaxi-semantika elkargunea

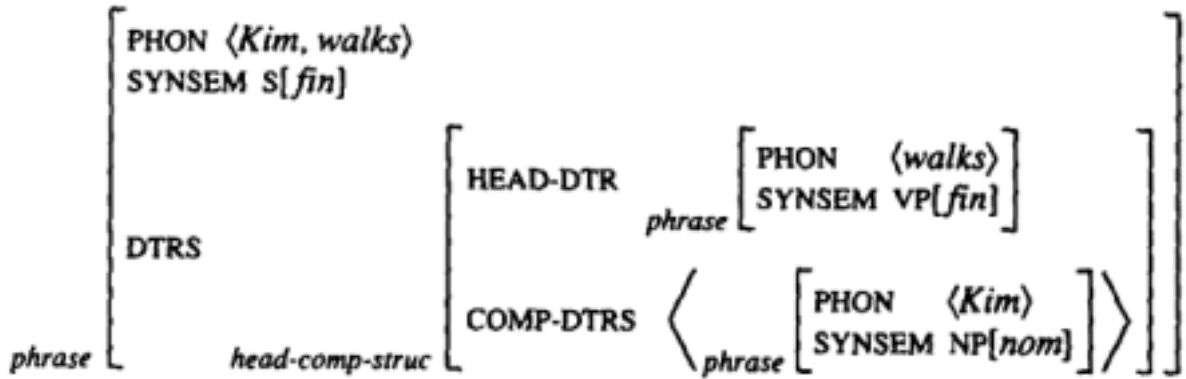
Sintaxi-semantikaren elkargunea azaltzeko aditzen *SUBCAT* ezaugarrian (azpikategorizazioari dagokiona) oinarrituko gara.

Gogoratu ezaugarri honek ale lexikalaren azpikategorizazioa zehazten duela eta *cat* matrize motaren barruan kokatzen dela, *HEAD* ezaugarriarekin batera (ikus 14. eta 16. irudiak).

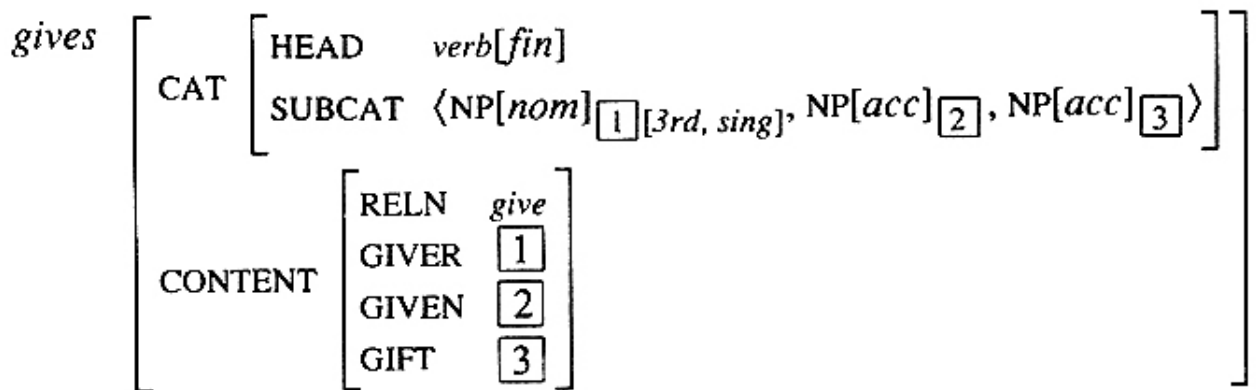
¹⁹Adjektibo eta preposizioen adierazpenek badute *INDEX* ezaugarria, baina honek modifikatzen duten izenari egiten dio erreferentzia.

²⁰Erreferentzia adierazteko azpindize bezala ere uler daiteke.

²¹Erreferentziadun espresioak dira, munduko izaki bati edo besteri erreferentzia egiten diotenak: *liburu*, *Jon*, *bertsolari*. . . (Artiagoitia, 2000).



15. irudia: Sintagmen adierazpena HPSGn.

16. irudia: *gives* aditzaren adierazpena HPSGn.

16. irudian²² *gives* aditzaren sarrera lexikala dugu; *CATEGORY* edo *CAT* ezaugarrian hurrengo informazioa zehazten da: aditz burutua dela (sarrera lexikalean *fin* (*finite*) balioarekin adierazita baitago), eta hiru argumentu hartzen dituela: 3. pertsonan dagoen IS nominatibo bat (irudian *NP[nom1[3rd,sing]]*) eta bi IS akusatibo (irudian *NP[acc]2* eta *NP[acc]3*).

CONTENT ezaugarriaren bitartez adierazten zaigu:

- Ingeleseko *gives* aditza *give* erlazioarekin harremanetan dagoela, honen errolak *GIVER*, *GIVEN* eta *GIFT* direlarik.
- GIVER*, *GIVEN* eta *GIFT* errolak 3. pertsonan dagoen IS nominatiboari (*NP[nom1[3rd,sing]]*) eta bi IS akusatiboari (*NP[acc]2* eta *NP[acc]3*) dagozkie, hurrenez hurren.

Hortaz, azpikategorizazioan dagoen osagarri bakoitza errol batekin lotuak egon behar du, eta lotura hau azpindize berdinekin dator adierazita²³.

²²Sarrera lexikal hauek matrize osoen laburpen bat dira, hau da, *word*, *synsem* eta *local* matrize motak ez dira adierazten, hasierako sailkapen hori irudi hauetan ikusten dugun informazioarekin berreskura baitaiteke.

²³Errol-thematikoak *Situation Semantics* teoriako “egoera” horren ikuspegi desberdinak lirateke.

HPSG, (...), is in principle a theory of *signs*, and is therefore concerned with providing a direct account of the *connection* between syntactic phenomena, e.g. sub-categorization, and semantic phenomena, e.g. roles in the described situation. (Pollard eta Sag, 1987, 99. orr)

Nahiz eta sarrera lexikalean ez ikusi, funtzio gramatikalak ere zehaztu egiten dira. Argumentuen funtzioa azpikategorizazioan (*SUBCAT*en) duten ordenak adierazten du. Hau da, lehenengo argumentua subjektua izango da, bigarrena objektu zuzena eta azkena zehar objektua. 16. irudian, *give* aditzaren argumentuen ordena hurrengoa da:

(29) $\langle NP[nom1[3rd,sing]], NP[acc]2, NP[acc]3 \rangle$

Ordena honen arabera:

- a) Subjektua 3. pertsonan dagoen IS nominatiboa da ($NP[nom1[3rd,sing]]$), *GIVER* errol-thematikoa daramana.
- b) Objektu zuzena bigarren IS nominatiboa da ($NP[acc]2$), *GIVEN* errol-thematikoa daramana.
- c) Zehar objektua azken IS nominatiboa da ($NP[acc]3$), *GIFT* errol-thematikoa duena.

Hortaz, sintaxi-semantika elkargunearen aurkezpena informazio sintaktiko-semantiko osoarekin osatua dago: errol-thematikoak, kategoria sintagmatikoak, kasua eta funtzio sintaktikoak.

Amaitzeko, aipatu beharra dago, adierak eta klase semantikoak ez dituela lantzen.

4.2.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

Izenen eta aditzen sarrera lexikalak aztertu ditugu, baina HPSGn beste kategoriak ere adierazten dituzte. Esate baterako, 17. irudian preposizio baten (ingeleseko *to*) sarrera lexikala dugu ikusgarri.

$$to \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } prep \\ \text{SUBCAT } \langle NP[acc] \boxed{1} \rangle \end{array} \right] \\ \text{CONTENT | NUCLEUS } \left[\begin{array}{l} \text{transfer } \left[\text{GOAL } \boxed{1} \right] \end{array} \right] \end{array} \right]$$

17. irudia: *to* preposizioaren adierazpena HPSGn.

Azpikategorizazioan agertzen den bezala (*SUBCAT*en alegia) preposizio honek IS bat (irudian *NP*) hartzen du. Bestalde, preposizio honek lekualdatze egoera bat adierazten du (irudian *transfer* bezala adierazita) eta egoera honen *helburua* (irudian *GOAL*) IS hori da, errol-thematikoak eta ISak azpindize bera daramatelako.

Atal honetan ez dugu azpikategorizazioaz hitz egingo, gorago (4.2.2 atalean) eman baititugu honi buruzko azalpenak.

- Hautapen-murriztapenak:

HPSGko sarrera lexikalak zehaztasun handiz definituta daude, baita hautapen-murriztapenak ere. Hau oso argi ikus daiteke aditzen sarreretan (ikus 16. irudia), non adierazten zaigun ale lexikal horrek 3. pertsonako IS nominatiboa eskatzen duela subjektu gisa.

- Errol-thematikoak:

Esan dugun bezala, *CONTENT* ezaugarrian ale lexikalaren irakurketa semantikoa zehazten da, beraz, hemen egongo dira errol-thematikoak adierazita. *Situation Semantics* teoriaren eredu jarraituta, HPSGn errol-zerrenda oso zabala dago, aditz bakoitzak bere errolak baititu (*give* aditzak, *GIVER*, *GIVEN* eta *GIFT*; *see* aditzak *SEER* eta *SEEN*; *walk* aditzak *WALKER*...).

In situation semantics (...) each relation is equipped with its own inventory of roles. (...) for our purposes it is enough to keep the roles of any given relation distinct, without worrying about how to classify roles across different relations²⁴. (Pollard eta Sag, 1987, 85-6. orr)

4.2.4 Erabilera

HPSG inplementazio handia duen formalismoa da, eta honen erabileraren berri <http://hpsg.stanford.edu/> web orrian ematen da. Atal honetan erabilera ezagunenak bakarrik aipatuko ditugu.

Batetik, HPSGk analisi sintaktikoak automatikoki egiten dituen *parser*-ak ditu (Minnen, 1999; Nishida *et al.*, 1999; Popowich eta Vogel, 1990; Copestake eta Flickinger, 2000). Esate baterako, Copestake eta Flickinger-ek (2000) ingeleserako *parser* bat egin dute, eta honen aplikazioetako bat itzulpen automatikoa izan da. Proiektu horretan bileren egitaraua eta bidaia-erreserbak ziren itzuli beharreko gaiak edo domeinuak. Hori egin ahal izateko lehenengo corpuseko perpausen adierazpena lortu behar zen eta Copestake eta Flickinger-en ingeleserako *parser*-a (2000) hori egiteko erabili zen.

HPSG formalismoak corpus etiketatua ere baditu, ingeleserako (Oepen *et al.*, 2002, edo LinGO Redwoods deiturikoa) eta beste hizkuntza batzuetarako ere, hala nola bulgariararako (Osenova eta Simov, 2003). Oepen *et al.*-en corpora (2002) egiteko etiketatzaileek *parser*-ak perpausaren analisi ezberdinetatik zuzena zein den aukeratu baino ez du egin behar.

²⁴HPSGn izenak ez diren beste zeinuak (aditzak, preposizioak eta adjektiboak) erlaziotzat hartzen dira (*relations*). Erlazio hauek errealitatearen hainbat ezaugarri (kolore bat), ekintza (irakurtzea) eta egoera (haserre egotea) adierazten dituzte. Errol-thematikoek erlazio hauetan parte nola hartzen duten definitzen dute.

5 HIZKUNTZALARITZA KONPUTAZIONALEAN OINARRITUTAKO IKERLANAK

5.1 FrameNet

5.1.1 Sarrera

FrameNet proiektua (Fillmore eta Baker, 2001a) ingeleserako baliabide lexikografikoa erakitzen ari da, *Frame Semantics* (Fillmore, 1985) teorian oinarritua eta corpus errealeko datuekin lagunduta. *Frame Semantics*-ak aldarrikatzen dituen printzipio nagusienak dira:

- Ale lexikalen semantika eta funtzio gramatikala *frame*-tatik¹ (egitura kontzeptual aberatsetatik) dator.
- Kontzeptualki erlazionatuak dauden ale lexikalak, *frame* (egitura kontzeptual aberats) bereko alderdi desberdinak erakus ditzakete.

Bi printzipio hauetan oinarrituaz, FrameNet-en ale lexikal bakoitza beraiek sortutako *frame*-etan sailkatzen dute, batetik ale honen semantika eta sintaxia definitzeko, eta bestetik, *frame*-ko beste osagaiekin duen harremana zehazteko.

Hurrengo ataletan emango dugu proiektu honen berri, xehetasun gehiagorekin.

5.1.2 Sintaxi-semantika elkargunea

Teoria honetan sakontzearren har dezagun (30) adibidea oinarri gisa:

(30) Hook tries to avenge himself on Peter Pan by becoming a second and better father.

Perpaus hau, *avenge* aditzaren eraginez, *Mendekuaren* esparruari dagokiola esango genuke, hots, *Revenge frame*-ari:

REVENGE FRAME:

An Agent performs a **Response action** on an **Victim** as a punishment for an earlier action, the **Injury**, that was inflicted on an **Injured party**. **The Agent** need not be identical to the **Injured party** but needs to consider the prior action of the **Victim** a wrong. Importantly, the punishment of the **Victim** by the **Agent** is seen as justified by individual or group opinion rather than by law.

Frame hau onartzen duten ale lexikalak hurrengoak dira: *avenge*, *avenger*, *get even*, *retaliate*, *retaliation*, *retribution*, *retributive*, *retributory*, *revenge*, *vengeful*, *revenger*, *vengeance*, *vengeful* eta *vindictive*.

¹Maileguak deklinatzerako idatzizko azken hizkian oinarritu gara: *Frame Semantics*-eAN, *frame*-AN, eta abar.

Revenge frame-aren alderdiak (*Agent*, *Injured party*, *Victim*, ...) ale lexikal desberdinez egongo dira adierazita (ikus 31. adibidea beheraxeago), eta ondorioz, ale lexikal hauek *frame* horren alderdiak agertzen dituzte.

- (31) [Hook **Agent**] tries to avenge [himself **Injured party**] [on Peter Pan **Victim**] [by becoming a second and better father **Response action**].

Hortaz, FrameNet proiektuan egiten dutena hurrengo da: ale lexikal bakoitza bere adieraren arabera sailkatu honi dagokion *frame*-an. Hala, ale lexikala polisemikoa bada, adiera bakoitza *frame* desberdinean egongo da.

Frame-en funtsa adieran dago, ale lexikal beraren adierak *frame* ezberdin batean egongo dira.

It isn't that every word has its own frame, but every sense of every word has its own frame. (www.icsi.berkeley.edu/framenet/book/book.html)

Frame-etan oinarritzeak orokortzeko aukera ematen du, hau da, *frame* bera osatzen duten ale lexikalek klase semantiko bat osatzen dute, eta hori dela eta, *frame*-a definitzen duten ezaugarri kontzeptualak klase semantiko osoari egokitzen zaizkie: *Revenge frame*-aren kasuan, *avenge*, *avenger*, *get even*, *retaliate*, *retaliation*, *retribution*, *retributive*, *retributory*, *revenge*, *vengeful*, *revenger*, *vengeance*, *vengeful* eta *vindictive* ale lexikalei. Klase semantikoak, beraz, beti dator zehaztua berau onartzen duten ale lexikalen zerrendarekin.

Lan hau egiteko etiketatze semantikoa baliatzen dute. Perpaus bakoitzaren etiketatzea *target*-en (perpauko ale lexikal baten) ikuspuntutik eginda dago, hau da, perpauko ale lexikal baten *frame*-a oinarri hartuta², perpauko beste elementuak *frame* horren alderdiei lotuko zaizkie. Esaterako, (31) perpauaren etiketatzean *avenge* aditza izan da etiketatzeko ardatza (*target*-a). Beraz, perpauko beste ale lexikalak *avenge*-ri dagokion *frame*-aren alderdiekin etiketatu dira.

Alderdi hauek *frame*-aren partehartzaileak dira —*frame elements* (FE³ hemendik aurrera) deiturikoak.

Alderdi semantikoarekin batera, osagaien funtzio eta kategoria sintagmatikoak ere etiketatzen dira, eta *target*-aren ikuspuntutik egingo denez, perpauko ale guztiek berarekin duten lotura sintaktikoa adieraziko dute.

Ondorioz, perpauen etiketatze semantikoaren emaitza izango da, perpauko ale lexikal bakoitza etiketatua egotea FE batekin (*Victim*), funtzio sintaktiko batekin (objektua), eta kategoria sintagmatiko batekin (IS). Hala bada, perpauko ale guztiek *target*-arekiko duten lotura sintaktiko-semantikoa adieraziko dute.

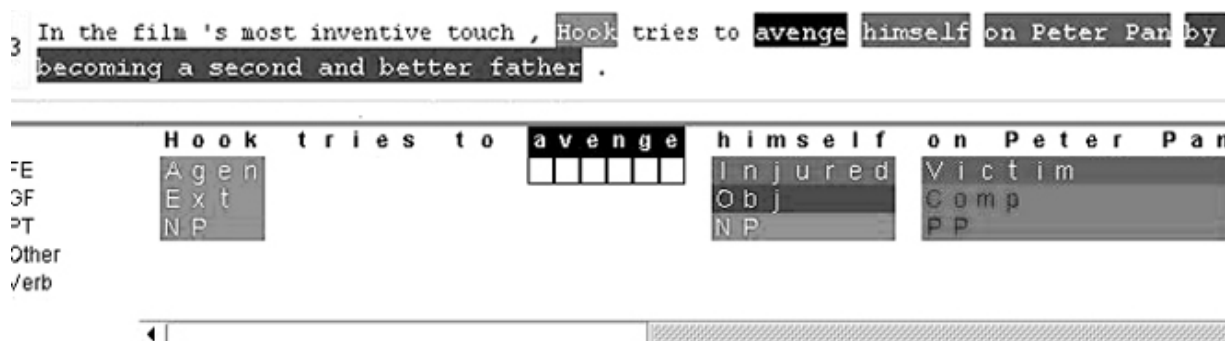
Azaldu bezala, *frame*-ek sintaxi eta semantikaren arteko harremana formalizatzeko aukera ematen dute. FrameNet-ek harreman honen adierazpenerako interfaze⁴ informatiko bat darabil, corpusaren agerpen etiketatuak kontsultatzeko.

18. irudia (31)eko adibide bera da baina oraingoan osatua; FEz gain (*Agent*, *Injured* eta *Victim*), funtzio gramatikalak (*Ext.*, *Obj.* eta *Comp.*), eta kategoria sintagmatikoak (*NP*, *NP* eta *PP*) adierazita datozelako.

²Ale lexikal hauek aditzak, objektuak edo adjektiboak izango dira, hots, gobernatzaileak izan daitezkeen ale lexikalak.

³FE laburdura “efe-e” irakurtzen dela suposatuko dugu, eta izen arrunt bat bezala deklinatuko dugu: FEa, FEan, FEak eta abar.

⁴Informatikaz ari garenean, ez dugu *elkargune* terminoa erabiltzen, arlo honetan *interfaze* terminoa hedatua baitago.



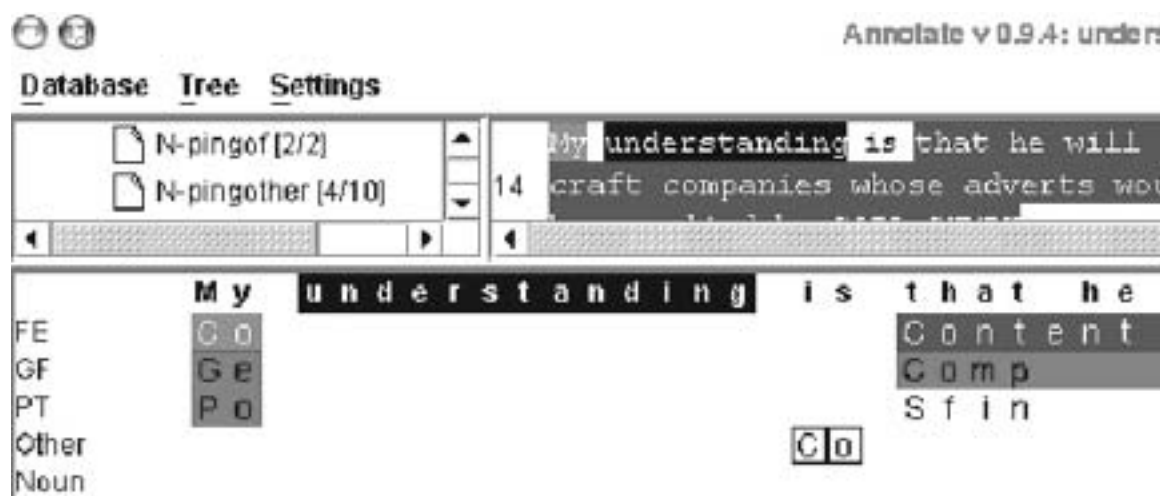
18. irudia: FrameNet-eko interfazearen adibide bat.

5.1.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

Dagoeneko esan dugun bezala (5.1.2 atalean), perpaus bakoitzaren etiketatzea *target* gobernatzailetan oinarritzen da, aditz, objektu edo adjektiboetan, alegia. Edo bestela esanda, *frame*-ak, predikatuetarako bakarrik egiten dituzte.

Preposizio eta adberbioak ez daude *frame* bat onartzen duten ale lexikalen artean. Esaterako, *Revenge frame*-a adibide gisa hartuz gero, *frame* hau onartzen duten ale lexikalak aditzak (*get even* adibidez), izenak (*vengeance* adibidez) eta adjektiboak (*revengeful* adibidez) dira; baina preposizioak eta adberbioak (*revengefully* adibidez) ez dira inoiz hauen artean agertuko. Dena den, preposizio eta adberbioek *frame*-aren FE bat har dezakete. 18. irudian, adibidez, *Victim* FEa *on Peter* PSari (irudian PP) egokitu zaio. Arrazoi honengatik, esan dezakegu teoria honetan sintaxi-semantikaren azterketa aditz, izen eta adjektibo kategorietan oinarritzen dutela.

Aurreko adibideak aditzari zegozkion, eta oraingoan izenari erreparatuko diogu (ikus 19. irudia). Ikus daitekeen bezala, aditzaren prozedura bera darabilte, baita adjektiboekin ere. Horregatik, azpikategorizazio, hautapen-murriztapen eta errol-thematikoez mintzatzen garenean, orokorrean arituko gara.

19. irudia: FrameNet-eko *understanding* izenaren etiketatzea FrameNet-en.

- Azpikategorizazioa:

Target-en deskribapen osoa corpusean egin ahal izateko, gauzatzen diren testuinguru sintaktikoaren berri ematea beharrezkoa da. *Frame*-eko zer ezaugarri —zer FE alegia— agertu behar diren nahitaez (*core* deritzotena), zenbat dauden eta zer funtzio sintaktikori lotzen zaizkion adierazten saiatzen dira.

Horretarako, corpus erreal bat etiketatetik lortzen dituzten datuak erabiltzen dituzte patroi sintaktikoak proposatzeko, hots, *target* bakoitzaren azpikategorizazioa definitzeko. Esaterako, corpuseko agerpenetan oinarrituaz *Revenge frame*-ean dagoen *avenge* aditzaren azpikategorizazioa 20. irudikoa litzateke.

Number Annotated	Patterns				
	Avenger	Injured Party	Injury	Offender	Punishment
2 TOTAL					
1	NP Ext	NP Obj	PP[for] Comp	-- --	PPing[by] Comp
1	NP Ext	NP Obj	PP[of] Comp	-- --	PPing[by] Comp
11 TOTAL	Avenger	Injured Party	Offender	Punishment	
2	-- --	NP Ext	-- --	-- --	
1	-- --	NP Ext	PP[on] Comp	-- --	
6	NP Ext	NP Obj	-- --	-- --	
1	NP Ext	NP Obj	-- --	PPing[by] Comp	
1	NP Ext	NP Obj	PP[on] Comp	PPing[by] Comp	
19 TOTAL	Avenger	Injury	Offender	Punishment	
3	-- --	NP Ext	-- --	-- --	
1	-- --	NP Ext	-- --	PP[by] Comp	
10	NP Ext	NP Obj	-- --	-- --	
2	NP Ext	NP Obj	-- --	PP[with] Comp	
2	NP Ext	NP Obj	-- --	PPing[by] Comp	
1	Poss Ext	-- --	PP[against] Comp	-- --	

20. irudia: *avenge* aditzaren azpikategorizazioa corpuseko agerpenetan oinarrituta.

- Errol-thematikoak:

Beste teorietan bezala, FrameNet-ek ohiko errol-thematikoen zerrenda ez dute nahikoa ikusten hizkuntzan gertatzen diren harreman semantiko guztiak adierazteko. Horregatik, proposatzen dituzte FEak.

FEekin *frame* bakoitzaren parte-hartzaile eta bestelako ezaugarriak adierazten dira.

These [*frames*] are schematic representations of situations involving various participants, props, and other conceptual roles, each of which is a frame element (FE). The semantic arguments of a predicating word correspond to the FEs of the frame or frames associated with that word. (Johnson eta Fillmore, 2000, 56. orr.)

Hala bada, azpikategorizazioan aipatu bezala, FrameNet-en etiketatze semantikora-ko FEak erabiltzen dituzte perpauseko ale lexikalak (argumentuak, modifikatzaileak eta lexikografikoki aipagarri diren bestelako aleak) etiketatzeko.

Hauen zerrenda oso zabala da, *frame* bakoitzak bere FEak ditu, nahiz eta batzuk ohiko errol-thematikoekin bat datozen (*Agent*-eta bezalakoak, noski). Eta gainera, *frame* bateko FEak kategoria sintagmatiko eta funtzio gramatikal batzuekin ager daitezke *frame* horretako ale lexikalen arabera. Johnson eta Fillmore-ek (2000) fenomeno hau adibide argi batekin agertzen dute:

(32) Buyer **bought** Goods from Seller for Money.

(33) Seller **sold** goods to Buyer for Money.

(34) Seller **sold** Buyer Goods for Money.

Adibide hauek guztiak, *Commercial Transaction* delako *frame*-ari dagozkio, non FEak *Buyer*, *Seller*, *Goods* eta *Money* diren. Salerosketa adierazten duten ale lexikalek (letra lodiz idatziak) FE hauei kategoria sintagmatiko eta funtzio gramatikal batzuk esleitzen dizkiete. Hala nola, *buy* aditzak *Buyer* eta *Seller* FEak IS-subjektu eta PS bezala adierazten ditu hurrenez hurren; *sell* aditzak, berriz, *Buyer* FEa IS-objektu edo PS bati esleitzen dio eta *Seller* FEa IS-subjektu bati.

FrameNet-en lan egiten dutenek proposatzen duten semantika, munduaren ezaguerarekin bat datorrela defendatzen dute. Hitzunari ez baitzaio arrotza egiten ikuspegi aldaketa hau, gure munduaren ezaguerarekin bat datorrelako.

Beraz, *frame* baten semantika FEekin adierazten da, baina FE hauen agerpena alda daiteke, *frame*-aren ikuspegia alda daitekeen bezala.

- Hautapen-murriztapenak:

Nahiz eta ez dituzten esplizituki adierazten, inplizituki ere uler daitezke. Hala, esan beharra dago FrameNet-en errol-thematikoak eta hautapen-murriztapenak ez dituztela bereizten. Adibidez, esan daiteke *avenge* aditzak, *Revenge frame*-ean dagoenean, [+Injured party] tasuna (hautapen-murriztapena) duen objektua eskatzen duela (ikus 18. eta 20. irudiak), baina errol-thematikoez ari bagara, *avenge* aditzak hartzen duen objektuaren FEa edo errol-thematikoa *Injured party* dela esaten digute.

5.1.4 Bestelako erlazio semantikoak

Orain arte, FrameNet ezagutza-basean harreman sintaktiko-semantikoaren adierazpena nola gauzatzen den aztertu dugu, baina badaude ezagutza-basean beste harreman semantiko mota batzuk. Garrantzitsuena *Frame-to-frame relations* deiturikoak dira, zeinek *frame*-en arteko harreman semantikoa erakusten duten. *Frame-to-frame relations* harreman mota ugari dauden arren⁵, ez ditugu denak aztertuko.

Ezagutza-baseak *frame*-ak gordetzean beraien sailkapena ere egiten du. Hau da, *frame* konplexuagoek zehatzagoak direnak barnean hartzen dituzte. *Frame* hauek *SubFrame* erlazioaren bitartez daude harremanetan. Horrela, *frame* konplexu horren FEak lotuak egongo dira *frame* zehatzagoaren FEekin. Ikus dezagun azalpen hau argituko duen adibide bat:

CRIMINAL PROCESS FRAME:

A Suspect is arrested by an Authority on certain Charges, then is arraigned as a Defendant. If at any time the Defendant pleads guilty, then the Defendant is sentenced, otherwise the Defendant first goes to trial. If the Verdict after the trial is guilty, then the Defendant is sentenced. In the end, the Defendant is either released or is given a Sentence by a Judge at the sentencing.

Criminal Process frame-a konplexua da, *frame*-ak aditzera ematen duen prozesuan beste *frame* batzuk adierazten direlako (*Arrest frame*, *Arraignment frame*, *Trial frame*, *Sentencing frame*, ...). Ezagutza-basean *frame* hauek guztiak *Criminal Process frame*-arekin lotuak daude *SubFrame* harremanaren bitartez:

(35) Subframe of: *Crime scenario*

Has Subframes: *Arraignment*, *Arrest*, *Sentencing*, *Trial*

Frame-to-frame relations-aren beste erlazioa mota bat *Frame Inheritance* da. *Frame* batzuk beste batzuk baino zehatzagoak dira (hiperonimia-hiponimia erlazioaren antzera), eta batzuetan FEak berak dira bietan, zehatzean zein orokorrean; eta besteetan, gerta daiteke orokorrean zehaztapenaren FEak eta *subframe*-ak jasotzea. Adibidez, *Perception active frame*-a *Perception frame*-aren zehaztapen bat da, eta arrazoi horregatik, pareko ezaugarri semantikoak dituzte. Hala ere, lehenengoan *agentive* tasuna gaineratu dute, ekintza hori nahitaezkoa dela adierazteko.

	Perception Active	Perception
(36)	Ground	Ground
	Perceiver agentive	Perceiver
	Phenomenon	Phenomenon

⁵*Frame-to-frame relations* harremanetako batzuk *SubFrames*, *Frame Inheritance* eta *SeeAlso* dira. Hala ere, gaur egun harreman hauen kopurua oraindik ez dago zehaztua.

Interfaze informatikoan, berriz, sarrera lexikoan bertan agertuko da bi hauen arteko harremana:

(37) **PERCEPTION ACTIVE**

Inherits From: *Perception*

Is Inherited By:

(38) **PERCEPTION**

Inherits From:

Is Inherited By: *Perception active, Perception experience, ...*

Horrela bada, FrameNet —izenak berak dioen bezala— *frame*-ekin eratutako sare semantikoa dugu.

5.1.5 Erabilera

FrameNet-ek gutxi gorabehera, 450 *frame*, 6.000 ale lexikal eta 130.000 esaldi etiketatu ditu eta handitzen jarraitzen du. Ezagutza-basea irakur daiteke www.icsi.berkeley.edu/frame-net web orrian.

FrameNet erabilerak era askotakoak izan daitezke. Alde batetik, hiztegi eta thesaurus gisa erabil daiteke: ale lexikalen definizioak daude eta FEen gauzatze sintaktikoa zehaztuta dator tauletan eta corpuseko adibideetan. Bestalde, LNPren zenbait arloen ikasketarako oso baliagarria:

- **Adieraren desanbiguazioan:** FrameNet-ek adieraren desanbiguazioan asko lagun dezake, adieraka antolatua dagoelako eta adiera bakoitzak bere informazio sintaktiko-semantikoa daramalako. Esate baterako, *parser*-a ale lexikalaren ezaugarri sintaktiko-semantikoen konbinazioetan oinarritu daiteke ale lexikalaren adiera mugatzeko. Adibidez, eta *Revenge frame*-aren adibidearekin jarraitzeagatik, har dezagun ingeleseko *revenge* ale lexikala. Hau anbigua izan daiteke, ale berak bi kategoria har ditzakeelako: izena eta aditza. Beheko adibideetan bi aukera hauek dituen perpau-sak ditugu — (39)an *revenge* izena da eta (40)an aditza—, hauen *frame*-aren FEak ere adieraziak daudelarik.

(39) [*Avenger*> Many of my friends] will take **revenge** [*Offender*> on me] [*Injury*> for making that remark].

(40) To **revenge** [*Injured Party*> him], [*Avenger*> his widow] joined the independent French section.

Kasu honetan *revenge* ale lexikalaren bi aukerak *Revenge frame*-ari dagozkienez, *parser*-ak zuzenean *Revenge frame*-an kokatuko ditu⁶. Nahiz eta *frame* berean egon, kategoria anbigutasuna ebatzi egin behar da; horretarako, *parser*-ak adiera bakoitzak hartzen dituen FEak kontuan hartuko ditu, hots, bakoitzaren azpikategorizazioa (ikus 21 eta 22. irudiak)⁷.

⁶Ale lexikala anbigua denean eta izan dezakeen adiera bakoitza *frame* ezberdin bati dagokienez (demagun *Revenge frame*-a eta *Perception frame*-a direla), *parser*-ak ale lexikala desanbiguatzeko du adiera bakoitzak dituen FEen arabera (batak *Avenger*, *Injured Party*, *Injury*-eta bezalakoak izango ditu eta besteak, berriz, *Ground*, *Perceiver agentive*, *Phenomenon* bezalakoak).

⁷*Revenge* izenaren azpikategorizazioak corpusean kasuistika altua du eta 21. irudian ez dago bere azpikategorizazio osoa. Patroi sintaktikoak adierazteko nahikoa den informazioa bakarrik aukeratu dugu.

Number Annotated	Patterns				
5 TOTAL	Avenger	Avenger	Injury	Offender	
5	NP Ext	Poss Gen	PP[on] Comp	PP[on] Comp	
1 TOTAL	Avenger	Avenger	Injury	Offender	Punishment
1	NP Ext	Poss Gen	-- --	-- --	Swh Comp
1 TOTAL	Avenger	Injury	Injury	Offender	
1	NP Ext	-- --	PP[on] Comp	PP[on] Comp	
12 TOTAL	Avenger	Injury	Offender		
---	NP Ext	PPing[for] Comp	PP[against] Comp		
8 TOTAL	Avenger	Injury	Offender	Punishment	
---	Poss Gen	PP[for] Comp	-- --	NP Ext	
3 TOTAL	Injury	Offender	Punishment		
---	PP[for] Comp	-- --	NP Ext		

21. irudia: *revenge* izenaren azpikategorizazioa FrameNet-en.

FE antzekoak hartzen dituzten arren, izenak ez du inoiz hartzen aditzak hartzen duen FE bat (*Injured Party*). Horretan oinarrituz gero, (39) eta (40) perpausen aurrean, *revenge* ale lexikalaren desanbiguazioa egin daiteke automatikoki.

- **Azaleko analisi semantikoaren garapenean:** *frame*-ak erabilia eta etiketatutako corpusetik ikasiaz, testu librean erroak identifikatzen dituzten algoritmoak eraiki dira (Gildea eta Jurafsky, 2002).
- **Informazio erauzketan:** FrameNet-ekin etiketatutako corpora oso lagungarria gerta daiteke, corpusetik informazioa ikasteko (azpikategorizazio patroiak, adibidez). Esate baterako, desanbiguaziorako esplikaturako prozesu hori guztia, corpusen zehaztua etorriko da, makinak zuzenean hortik ikas dezakeelarik.
- **Itzulpen automatikoan:** Beste hizkuntzetako FrameNet-ak eraikiz gero, adieren arteko ezberdintasun semantiko eta sintaktikoak ezagutza-basean lotuta egoteko aukera izango genuke. Adibidez, ikusi dugun bezala FrameNet-en ingeleseko *revenge* ale lexikalak bi sarrera lexikaletan agertzen da (bata izena eta bestea aditza adierazteko). Euskarako FrameNet-a izanez gero, sarrera lexikal horietan *mendeku* eta *mendekatu* izango genituzke, hizkuntza batekin bestera automatikoki itzultzeko prozesua erraztuko lukeena.

Number Annotated	Patterns				
2 TOTAL	Avenger	Injured Party	Injury	Offender	Punishment
1	--	NP	--	--	--
	--	Ext	--	--	--
1	NP	NP	PP[ing[for]	PP[on]	--
	Ext	Obj	Comp	Comp	--
3 TOTAL	Avenger	Injured Party	Offender	Punishment	
1	--	NP	--	--	
	--	Ext	--	--	
1	NP	NP	--	--	
	Ext	Obj	--	--	
1	NP	NP	PP[on]	--	
	Ext	Obj	Comp	--	

22. irudia: *revenge* aditzaren azpikategorizazioa FrameNet-en.

Egun, zenbait FrameNet ari dira garatzen hizkuntza ezberdinetan: German FrameNet (Boas, 2002), Spanish FrameNet (Subirats-Rüggeberg eta Petruck, 2003) eta Japanese FrameNet (Ohara *et al.*), hain zuzen ere.

- **Galdera-erantzunean:** *frame*-ek galdera bati dagozkion erantzunak ezagutzea ahalbidetzen dute.

Horrela, nahiz eta FrameNet bertan LNPko aplikaziorik ez egon, bertako informazioa LNPko inplementazioak egiteko (Fillmore *et al.*, 2002; Fillmore eta Baker, 2001b, besteak beste) oso erabilgarria da.

5.2 WordNet eta WordNet-etik abiatutakoak

5.2.1 Sarrera

WordNet (Miller, 1985; Fellbaum, 1998) teoria psikolinguistikoetan oinarritua dagoen ingeleseko ezagutza-base lexikala da. Princeton-eko Unibertsitatean garatzen ari da — Cognitive Science Laboratory delakoan— George A. Miller-en ardurapean.

Ingeleseko izen, aditz, adjektibo eta adberbioak *synonym set* edo *synset*-etan (sinonimo multzotan) antolatuak daude, hauetako bakoitza kontzeptu lexikal bati dagokiolarik. Esaterako, ingeleseko *tree* izenak WordNet-en bi *synset* (adiera)⁸ ditu:

- (41) The noun “tree” has 2 senses:
1. tree (a tall perennial woody plant having a main trunk and branches...)
 2. tree, tree diagram (a figure that branches from a single root; “genealogical tree”)

Lehenengoa ‘landare’ (*plant*) *synset*-ari dagokio eta bigarrena, berriz, ‘diagrama’ (*diagram*) *synset*-ari. Lehenengo *synset*-a ale lexikal bakar batez osatua dago (*tree*), hots, *tree*

⁸Aurrerantzean *synset* terminoa erabiliko dugu.

izenak *synset* horretan ez du beste sinonimorik. Bigarrenak, ordea, *synset*-ean *tree* ale lexikalaz gain, beste ale bat ere badu (*tree diagram*), horrela, bi ale lexikal horiek (*tree* eta *tree diagram*) sinonimoak dira.

Ildo honetatik, WordNet-eko erlazio semantiko garrantzitsu bat **sinonimia** da; eza-gutza-basearen oinarria ale lexikalaren adieran baitago, eta adiera hori ale lexikal batek baino gehiago duenean, ale lexikalak multzokatu egiten dituztelako.

WordNet-eko sinonimiaz hitz egiterakoan, kontuan izan behar da ez dela gauza bera sinonimia eta hitzak bata besteaz elkar trukatzeari. Hau da, WordNet-eko *synset*-a osatzen duten sinonimoak beraien artean truka daitezke, baina **testuinguru batzuetan** bakarrik.

The more modest claim is that WordNet synonyms can be interchanged in some contexts. To be careful, therefore, one should speak of synonymy relative to a context. (Fellbaum, 1998, 24. orr)

WordNet ez da *synset* zerrenda hutsa; *synset*-ak erlazio semantikoen bidez antolatuta daude. Esan dugun bezala, sinonimia da erlazio semantiko garrantzitsuenetakoa, baina honekin batera, WordNet-ek beste hainbat erlazio landu ditu, hala nola, **hiperonimia-hiponimia** erlazioa.

Hiperonimia-hiponimia erlazioak *synset* orokorrenak *synset* zehatzagoekin lotzen ditu⁹. (42) eta (43) adibideetan (41)en hiperonimoak eta hiponimoak ikus ditzakegu hurrenez hurren¹⁰:

(42) Sense 1

tree (a tall perennial woody plant having a main trunk and branches...)
 => woody plant, ligneous plant – (a plant having hard lignified tissues...)
 => vascular plant, tracheophyte – (green plant having a vascular system...)
 => plant, flora, plant life – (a living organism lacking the power of locomotion)
 => life form, organism, being, living thing – (any living entity)
 => entity, something – (anything having existence (living or nonliving))

Sense 2

tree, tree diagram (a figure that branches from a single root; "genealogical tree")
 => plane figure, two-dimensional figure (a 2-dimensional shape)
 => figure (a combination of points and lines and planes that form a visible palpable shape)
 => shape, form (the spatial arrangement of something as distinct from its substance)
 => attribute (an abstraction belonging to or characteristic of an entity)
 => abstraction (a general concept formed by extracting common features...)

(43) Sense 1

tree (a tall perennial woody plant having a main trunk and branches...)
 => yellowwood, yellowwood tree (any of various trees having yellowish wood...)
 => lancewood, lancewood tree, *Oxandra lanceolata* (source of most of the lancewood of commerce)
 => Guinea pepper, negro pepper, *Xylopi aethiopica* (tropical west African evergreen tree...)
 => anise tree (any of several evergreen shrubs and small trees of the genus *Illicium*)
 => winter's bark, winter's bark tree, *Drimys winteri* (South American evergreen tree...)
 => zebra wood, zebra wood tree (any of various trees or shrubs having mottled or striped wood)
 => granadilla tree, granadillo, *Brya ebenus* (West Indian tree yielding a fine grade of green ebony)
 => acacia (any of various spiny trees or shrubs of the genus *Acacia*)
 => ...

Sense 2

tree, tree diagram (a figure that branches from a single root; "genealogical tree")
 => cladogram (a tree diagram used to illustrate phylogenetic relationships)

⁹Ingeleseaz *IS-A relation* bezala ere ezagutzen da, hots, *x is a kind of y*.

¹⁰Adierazpen guztiak WordNet 2.0 bertsiotik hartu ditugu (<http://www.cogsci.princeton.edu/cgi-bin/webwn>).

(42) adibidean *tree* izenaren hiperonimoak ditugu. Lehenengo *synset*-a ('landare') kontuan hartuz gero, *woody plant* mota bat bezala definitzen da; *woody plant vascular plant* mota bat bezala; *vascular plant*, aldi berean, *organims* mota bat bezala, eta azkenik, *organims entity* mota bat bezala. Ondorioz, *tree*, bere lehenengo *synset*-ean, *entity*, *organims*, *vascular plant*, eta *woody plant* bat izango da.

Tree-ren beste *synset*-aren ('diagrama') sailkapenarekin berdin-berdin gertatzen da, baina bere hiperonimoak 'diagrama' adierari lotuak egongo dira.

Hiponimoak hiperonimoen zehaztapenak dira. Hortaz, (43) adibidean, *tree* izenaren lehenengo adieraren zehaztapen gisa zuhaitz motak agertzen dira (*yellowood*, *acacia*...), eta bigarren adieran, aldiz, diagrama motak (kasu honetan bakarra, *cladogram*). Horrela bada, WordNet ontologia edo hierarkia bat da, eta hiperonimia-hiponimia harreman semantikoarekin hierarkian gora eta behera egiteko aukera dugu. Ontologia hau kategoriaka banatua dago, eta kategoria bakoitzak bere hierarkia du; hau da, kategoria bakoitzaren hierarkia erlazioa semantiko nagusi baten arabera antolatzen da. Izen eta aditzen kasuan erlazio semantiko nagusia hiperonimia-hiponimia da¹¹. Adjektibo eta adberbioek, berriz, sinonimia-antonimia dute ardatz gisa beraien antolakuntzan. (44) adibidean, *properly* adberbioaren antonimoa ikus dezakegu (*improperly*):

- (44) Sense 1
 properly , decently, decent, in good order, right, the right way (in the right manner)
 => improperly (in an improper way)

WordNet-eko sailkapena, beraz, *synset*-etan eta beraiek harremanetan jartzen dituzten erlazio semantikoetan datza. Erlazio semantiko hauen bidez, *synset*-ak multzokatzen dira, edo beste era batera esanda, klase semantikoak osatzen dira. *Synset* orokorragoren azpian (adabegi horren azpian) bere zehaztapenak multzokatzen dira. Esaterako, zuhaitz mota desberdinak *synset* baten azpian jasota daude (*tree* izenaren 'landare' *synset*-ean, alegia), hortaz, zuhaitzen motak jasotzen dituen klase semantikoa *synset* horren bitartez adieraz daiteke.

Ondoren (5.2.2 atalean), ikuspegi sintaktiko-semantikoan sakontzearen, aditzaren azterketan murgilduko gara¹².

5.2.2 Aditza eta informazio sintaktiko-semantikoa

Askotan aipatu dugun bezala, sintaxi-semantika elkargunearen muina aditza da, perpau-saren antolakuntzak bere baitan hartzen baitu. Arrazoi honengatik, WordNet-en jasota dagoen informazio sintaktiko-semantikoa aditzari lotua dago.

WordNet-en aditzen *synset*-ak, irizpide semantikoan oinarrituz, 14 klase semantikotan (fitxategietan) banatuak daude (*motion; perception; contact; communication; competition; change; cognition; consumption; creation; emotion; perception; possession; bodily care and functions; verbs referring to social behaviour and interaction*). Bestetik, 14 klase semantiko horietan lekurik ez duten aditzen multzoa dago (*verbs denoting states* delakoan). Hauek egoera adierazten dute (*be, belong, resemble*...).

¹¹Aditzen kasuan, eta gero 5.2.2 atalean ikusiko dugun bezala, hiperonimia-troponimia erlazioz hitz egiten da.

¹²Ezagutza-base hau esplikatu ahal izateko, kapituluaren egitura zertxobait aldatu behar izan dugu: azpiatal batzuk bildu ditugu, bestela informazioa asko errepikatuko genukeelako.

Klase semantiko hauetako bakoitzean aditz horien antolaketaren hastapena dago, *unique beginner* deritzona. Hau izango da klase semantiko horren hierarkian mailarik altuena eta orokorra. Esaterako, *communication* klase semantikoak *unique beginner* bezala *communicate synset*-a du eta honetatik hasten da klase semantiko honetako aditzen sailkapena¹³.

Klase semantiko hauek aditzen sailkapenerako aproposak izan arren, klase semantikoaren arteko muga ez da guztiz hertsia. Hori dela eta, aditz batzuk klase semantiko bat baino gehiagotan egon daitezke; adibidez, ingeleseko *The bullet whistled past him*¹⁴ perpausean, *whistle* aditza *communication* edo *motion* bezala uler daiteke.

Gorago azaldu dugun bezala (5.1.1), WordNet *synset*-en arabera dago antolatua, eta berez, *synset*-a osatzen duten sinonimoak testuinguru batzuetan beraien artean truka daitezke. Aditzen kasuan trukatzeko hau bideratzea zaila gertatzen da. Batzuetan aditzek —*end/terminate*, *rise/ascend* bezalako anglosaxoi/greko-latindar hitz pareek adibidez— adiera bera izan arren, erregistro ezberdina eskatzen dute, adibidez, anglosaxoi/greko-latindar hitz pareen kasuan greko-latindarrak besteak baino erabilera jasoagoa dute.

Beste batzuetan, ordea, aditzen arteko adiera aldaketa hautapen-murritzapen ezberdinekin azaleratzen da. Esaterako, ingeleseko *rise* eta *fall* aditzek entitate abstraktuak (*temperature*, *prices*...) har ditzakete argumentu gisa; aurrekoen adieraren oso antzekoa duten *ascend* eta *descend* aditzek, berriz, ezin dute argumentu mota honekin agertu. Horrelako kasuetan, WordNet-en irizpide nagusia aditzak *synset* desberdinetan banatzea da, hau da, *rise* eta *ascend* bi *synset*-etan kokatzea.

Hortaz, hautapen-murritzapenak kontuan hartzen dituzte, baina ontologian oraindik ez dago adierazita zeintzuk diren aditzak hartzen dituen hautapen-murritzapenak, hau da, WordNet-eko interfaze informatikoak ez du eskaintzen *rise* eta entitate abstraktuak (WordNet-en *abstraction* ale lexikala daraman *synset*-aren bitartez adierazten dena) hautapen-murritzapen gisa lotzeko biderik.

Hautapen-murritzapenekin bezala, ale lexikal baten *synset*-ak ezberdintzerakoan azpikategorizazioa kontuan hartzen da. Horretaz gain, aditz bakoitzaren azpikategorizazioa zehazten dute errol-thematikoak aipatu gabe, eta aditzaren adiera bakoitzeko, azpikategorizazioaren antzeko patroia bat proposatzen dute:

- (45) 4 senses of “descend”
- Sense 1
 - descend , fall, go down, come down (move downward but not necessarily all the way)
 - EX: The airplane is sure to descend
 - Sense 2
 - derive , come, descend (come from; be connected by a relationship of blood, for example)
 - Something is —ing PP
 - Somebody —s PP
 - Sense 3
 - condescend , deign, descend (do something that one considers to be below one’s dignity)
 - Somebody —s to INFINITIVE
 - Sense 4
 - stoop, descend (to sink in status or dignity, or worsen in condition)
 - Somebody —s PP

¹³Gerora hierarkian, *communicate* beste bi hiponimo nagusietan banatzen da: “ahozko komunikazioa/idatzizko komunikazioa”.

¹⁴Adibidea Fellbaum-en lanetik (1998) hartua da.

Horrela bada, WordNet aurretik ikusi ditugun lanen eredutik banandu egiten da, *semantika deskonposatzailea* jarraitzen dutenetatik alegia. Jackendoff-ek bere lanean (1990) adibidez, aditzen azterketa primitiboak baliatuta egiten du (*TO, FROM, TOWARD, AWAY-FROM, CAUSE, GO, VIA...*). WordNet-en ale lexikalak ez daude unitate txikiagoetan deskonposatuak. WordNet-ek *loturazko semantikaren (relational semantics)* ildotik jorratzen ditu aditzak, hortaz, *synset*-ak hitzekin osatzen dira eta ez tasun edo primitiboekin. Hala eta guztiz ere, *synset*-en arteko harreman semantikoek deskonposaketaren alderdi batzuk ere eskain ditzakete. Nahiz eta WordNet-ek primitiboak edo antzeko tasun txikiagoak ez erabili, hauetako batzuk agerian geratzen dira harreman semantikoen bidez. Adibidez, *semantika deskonposatzailean* oihartzun gehien duen tasunetako bat *kau-sa* da (Jackendoff-ek (1990) *CAUSE* primitiboa deitzen duena). WordNet-en informazio hau erlazio semantiko batekin ikus dezakegu, eta bere bitartez *learn* aditza *teach* aditzaren ondorioa dela jakin dezakegu:

(46) 1 of 6 senses of “learn”

Sense 5

teach , learn, instruct (impart skills or knowledge to; ”I taught them French”)

=> learn, larn – (acquire or gain knowledge or skills; ”I learned Sanskrit”)

Bestetik, mugimendua adierazten duen tasunak Jackendoff-ek (1990) *GO* deitzen duena) hierarkiaren hastapen diren *unique beginners*-ek adieraz ditzakete. *Run* aditza adibide gisa hartuz gero, bere hiperonimo garaiena —*motion* klase semantikoaren *unique beginner*-a denak—, [*go, move, ravel, locomote*] *synset*-az osatzen da¹⁵, eta honek erakusten digu *run* mugimenduzko aditza dela.

(47) Sense 1

run (move fast by using one’s feet, with one foot off the ground at any given time)

=> travel rapidly, speed, hurry, zip (move very fast)

=> travel, go, move, locomote (change location; move, travel, or proceed)

Amaitzeko, aditzen moduaren berri hierarkian bertan dugu. Arestian hitz egin dugu hiperonimia-hiponimia erlazio semantikoaz. Aditzek erlazio hau bera dute baina hiponimiaren ordeztan **troponimia** erabiltzen dute. Honen arrazoia da aditzak ezin direla *IS-A* motakoak izan, baizik eta *to x is to y in some particular manner* motakoak. Hortaz, aditz hiperonimo baten (*walk*) troponimoak aditz hiperonimoak adierazten duena egiteko moduak izango dira (*trot, march...*). Hala, WordNet-ek ezagutza sintaktiko-semantikoaren berri emateko beste baliabide semantikoak erabiltzen ditu. Ezagutza-baseko sarrera lexikal bakoitza ez dator zehaztuta tasun zerrenda batekin; zehaztuta etorri beharrean, bere zehaztapena hierarkiatik jasotzen dituen tasunetatik dator.

¹⁵ *Motion* klase semantikoak bi *unique beginner* ditu, bata [*go, move, travel, locomote (change location)*], eta bestea, [*move, displace (cause to move)*]; lehenengoan norbait/zerbait mugitzen da, bigarrenean norbait/zerbait norbait/zerbait mugitzen du.

5.2.3 Bestelako erlazio semantikoak

Sinonimia eta hiperonimia-hiponimia/troponimia erlazio semantikoetaz gain, WordNet-ek beste asko landu ditu. Hemen batzuen aipamen laburra egingo dugu¹⁶.

Izenak lotuak egon daitezke ondorengo erlazio semantikoen bidez:

- *Part-whole relations*:

Zatia eta osotasuna harremanetan jartzen dituen erlazioak dira. Batetik, **meronimia** dago, *x is part of y* definizioari jarraitzen diona; *finger hand*-en zati bat da eta *hand* aldi berean *arm*-ena:

- (48) 1 of 2 senses of “finger”
Sense 1
finger (any of the terminal members of the hand; “her fingers were long and thin”)
PART OF: hand, manus, hook, mauler, mitt, paw (the extremity of the superior limb)
PART OF: arm (technically the part of the superior limb between the shoulder and...)

Eta bestetik, *x has a y (as a part)* definizioarekin bat badator, orduan, **holonimia** erlazioaren bidez lotzen dira; adibidez, *wheeled vehicle* batek *wheel*-ak dauzka:

- (49) 2 of 6 senses of “wheel”
Sense 1
wheel (a simple machine consisting of a circular frame with...)
PART OF: wheeled vehicle (a vehicle that moves on wheels...)

- *Antonimia*:

Izen batzuek antonimoak dituzte eta erlazio semantiko honek lotzen ditu:

- (50) 1 sense of “victory”
Sense 1
victory, triumph (a successful ending of a struggle or contest)
=defeat, licking (an unsuccessful ending)

Aditzen hierarkian erlazio semantiko nabarmenetako bat *entailment* deritzona da (*V1 logically entails V2* edota *snore entails sleeping*).

Adjektibo eta adberbioen kasuan, erlazio semantiko gutxiago daude. Adjektibo batzuk (adibidez *nice*) adiera berdineko izenekin (*niceness*) lotu egiten dira:

- (51) 1 of 6 senses of “nice”
Sense 1
nice (vs. nasty) (pleasant or pleasing or agreeable in nature or appearance; “nice manners”)
=> niceness (the quality of nice)

Esan bezala, erlazio semantiko batzuk baino ez ditugu aipatu. WordNet-en gehiago daude eta hauen kopurua handituz joan da.

¹⁶Argibide gehiago (Fellbaum, 1998) eta (Miller, 1985) lanetan.

5.2.4 Erabilera

WordNet-ek 115.424 *synset* ditu (79.689 izen, 135.08 aditz, 18.563 adjektibo eta 3.664 adberbio)¹⁷.

WordNet-en erabilera era askotakoak izan dira. Alde batetik, hiztegi eta thesaurus gisa erabili izan da. Hiztegi tradizioaletan bezala, WordNet-ek *synset* bakoitzeko definizio bat du, gehienetan adibide eta guzti. Gainera, *synset* bakoitzean ale lexikal bat baino gehiago egon daitezkeenez, thesaurus bezala balia daiteke, adiera berdina adierazteko sinonimo desberdinak ditugulako.

Honenbestez, LNPr-i begira, WordNet-ek erabilera ugari izan ditu. WordNet-eko web orrian agertzen den bibliografian¹⁸ hau erakusten duten 2.000 artikulu inguru daude. Guk arlo bakoitzetik garrantzitsuenak baino ez ditugu aipatuko:

- **Adieraren desanbiguazioan:** WordNet adieran oinarritutako ontologia denez, desanbiguazioan asko lagun dezake. Bestalde, adierak hierarkikoki antolatuta egoteak desanbiguazioaren atazan lagundu egiten du. Arlo honetan esperimentu ugari egin dira (Miller *et al.*, 1994; Banerjee eta Pedersen, 2002; Agirre eta Martínez, 2000; Matwin *et al.*, 1995, besteak beste).
 - **Itzulpen automatikoan** aipagarria da Dorr-ek (Dorr, 1993, 1997) Jackendoff-en teoriaran eta WordNet-eko adieretan oinarrituta egindako lana (ikus 3.1.4 atala). Honetaz gain, ikerlan gehiago ere egin dira: Rigau *et al.* (1995), Knight (1993), Moon eta Kim (1995), Siddharthan (2002) eta abar.
 - **Informazio erauzketan:** WordNet lagungarria izan daiteke erabiltzaileak bilaketa bat egiterakoan, eta beharrezkoa zaion edukia duen dokumentua lortzeko. Bilaketan erabilitako hitzak dokumentuan daudenekin bat etorri behar dute, emaitza egokia lortzeko. WordNet-ek horretan lagun dezake. Zenbait saiakera egin dira, esaterako, Magnini eta Strapparava (2001), Mandala *et al.* (1998), Milhacea eta Moldovan (2001), besteak beste.
- Bestalde, WordNet-ekin etiketatutako corpusa oso lagungarria gerta daiteke, ordenagailuak corpusetik informazioa ikasteko. Honen froga ingelesez etiketatutako corpusa dugu: SemCor (Miller *et al.*, 1994; Fellbaum *et al.*, 2001).
- **Galdera-erantzunean:** *synset*-en arteko harremanek galdera bati dagozkion erantzunak ezagutzea laguntzen dute (Pasca eta Harabagiu, 2001; Harabagiu eta Moldovan, 1996; Mann, 2002, eta abar).

WordNet abiapuntutzat hartuta, beste ezagutza-base batzuk egin dira: EuroWordNet (Vossen, 1998) eta *Multilingual Central Repository* (MCR)(Atserias *et al.*, 2004). Oinarri bera erabili arren, bere gainean aberasketa batzuk egin dituzte. Hurrengo ataletan (5.2.1 eta 5.2.2) oso laburki hauen berri emango dugu, batez ere, sintaxi-semantika elkarguneari erreparatuz.

¹⁷WordNet-en azkeneko bertsioaz ari gara, 2.0 bertsioaz, alegia.

¹⁸Ikus <http://enr.smu.edu/rada/wnb/> web orrian.

5.2.5 EuroWordNet

EuroWordNet proiektua (Vossen, 1998) 1996an hasi eta 1999 urteraino luzatuko den proiektu europarra da. Ezagutza-base eleanitza da, Europako zortzi hizkuntzataraz zabaltzen dena (ingelesez, daniera, italiar, gaztelania, alemana, frantsesa, txekiera eta estoniera).

EuroWordNet-ek Princeton-eko WordNet-aren eredu jarraitzen du (ikus 5.2 atala), hots, Princeton-en ingeleserako egindako WordNet-aren *synset*, harreman semantiko eta hierarkian oinarritu dira beraien WordNet-a sortzeko.

Nahiz eta EuroWordNet-en hizkuntza bakoitzak WordNet “independente” bat izan, EuroWordNet-en helburua WordNet desberdin hauek guztiak ezagutza-base eleanitz bakarrean elkartzea da. Horretarako, hizkuntza guztien WordNet guztiak elkargune bat dute, *Inter-Lingual-Index*-a (hemendik aurrera ILI) deritzana, aldi berean, Princeton-eko WordNet 1.5 bertsioari lotua dagoena. ILI honen bitartez, hizkuntza guztietako WordNet-ak lotuak daude, eta ingeleseko *synset*-a EuroWordNet-a osatzen duten hizkuntza guztietan ikusgarri egongo da. Beste hitz batzuetan esanda, *synset* bera ingelesez, danieraz, italiaraz, gaztelaniaz, alemanez, frantsesez, txekieraz eta estonieraz agertzen da.

WordNet-en egitura, harreman eta *synset*-etan oinarritu arren, WordNet-ek ez zituen ezaugarri batzuk EuroWordNet-en gaineratu dira. Aldaketarik aipagarrienak hurrengoak dira¹⁹:

- Erlazio semantikoen aberasketa:

WordNet-eko erlazio semantiko batzuk findu egin dituzte eta beste erlazio semantiko batzuk aberastu. EuroWordNet-ek hizkuntza barneko erlazio mota gehiago ditu, eta batez ere, morfologikoki aldatzen diren kategoria ezberdinen arteko erlazioak ugaritu dituzte (*nice* eta *niceness* bezalakoak, alegia).

Bestalde, EuroWordNet-ek ez du WordNet-en interfaze informatikoa; EuroWordNet-ena interfaze berria da, hizkuntza bakoitzeko WordNet-ak erlazio berriak gehitzeko aukera duelarik.

- Hierarkiaren aberasketa:

WordNet-ek ez zuen hierarkiari Domeinu-ontologia (*domain ontology*) eta Goi-ontologia bat (*Top ontology*) gehitu dizkiote.

Lehenbizikoak, *synset*-ak domeinuen arabera antolatzen ditu: *free time*, *restaurant*, *traffic*, eta abar. Esate baterako, *jokatu* aditzak kirola adierazten duenean (*futboleko jokatu* diogunean, adibidez), *synset* horrek *free time* domeinuaren marka eramango du; *zuzen jokatu* esan nahi dugunean, ordea, adiera horri dagokion *synset*-ak *psychology* marka izango du²⁰.

Bigarrenak, WordNet ezberdinen goi aldeko *synset*-ak oinarritzko ezaugarri semantikoen arabera sailkatzea ahalbidetzen du²¹, eta nolabait esateko, EuroWordNet-eko domeinuen papera jokatzen du, nahiz eta motibazio linguistiko sakonagoak hartu diren kontuan. Hau da, tasun linguistikoak ([+/- bizidun], [+/- egile] adibidez) kontuan hartzen dituen

¹⁹Argibide gehiago Vossen-en lanean (1998).

²⁰Domeinuen sailkapena ez da EuroWordNet-eko interfazean ikusten, beste fitxategi batzuetan daude.

²¹Goi-ontologiak goi aldeko *synset*-ak sailkatu arren, hauen azpian dauden *synset*-ak ere sailkapen hori mantentzen dute, beraien hiperonimoen ezaugarriak heredatzen dituztelako.

ontologia da eta WordNet tasun hauen arabera eraikitzen da. Hortaz, ale lexikal bat [+biziduna] bada Goi-ontologiaren [+biziduna] adabegiaren azpian kokatuko da eta [-biziduna] bada, aldiz, [-biziduna] ezaugarriaren behean.

Oinarrizko ezaugarri semantikoak definitzerakoan, EuroWordNet-en sortzaileak hizkuntzalaritzan egon diren zenbait sailkapen semantikoen eredutan oinarritu dira: Vendler (1967), Verkuyl (1972), Dowty (1979), Pustejovsky (1991), Levin (1993), Lyons (1977), eta Pustejovsky-ren ereduari (1995), besteak beste.

Guztira, 63 ezaugarri semantikok osatzen dute Goi-ontologia hau, eta Lyons-en lanari (1977) jarraituz hiru maila bereizi dituzte:

- **Lehenengo mailako entitateak (*1st Order Entity*):** Zentzuen bidez antzeman daitekeen eta denboran/lekuan antzeman daitekeen entitateak dira (*animalia, objektu, substantzia* eta abar bezalako ale lexikalak).
- **Bigarren mailako entitateak (*2nd Order Entity*):** Edozein egoera estatiko edo dinamiko, zentzuen bidez objektu fisiko bezala ezagutu ezin daitekeena. Denboran koka daitezke eta “gertatu” egiten dira “existitu” baino gehiago (*gertatu, hasi, jarraitu, izan, eduki, amaitu* bezalako ale lexikalak). Beraz, ekintzak, prozesuak eta egoerak adierazten dituzten ale lexikalak maila honen azpian egongo dira (*gertatu, hasi, jarraitu, izan, eduki, amaitu* bezalakoak).
- **Hirugarren mailako entitateak (*3rd Order Entity*):** Ikus ezin daitezkeen proposizioak dira, denbora eta lekuan koka ezin daitezkeenak. Proposizioak direnez, egia edo gezurra bezala uler daitezke, erreal edo irrealizat baino (*ideia, pentsamendu, informazio, teoria, plana* bezalakoak).

Goi-ontologiako maila hauen arteko desberdintasuna ageriagoa da hauek adierazteko erabiltzen diren kategoria sintaktikoei erreparatzen badiegu:

- **Lehenengo mailako entitateak (*1st Order Entity*):** izen konkretuak
- **Bigarren mailako entitateak (*2nd Order Entity*):** izenak, aditzak eta adjektiboak
- **Hirugarren mailako entitateak (*3rd Order Entity*):** izen abstraktuak

23. irudian Goi-ontologiaren hierarkia guztia zerrendatua ikus dezakegu.

EuroWordNet WordNet-en oinarritutako ezagutza-basea denez, informazio sintaktiko-semantikoa, WordNet-en parekoa da (ikus 5.2.2 atala). Hala eta guztiz ere, EuroWordNet-eko Goi- eta Domeinu-ontologiari esker, informazio sintaktiko-semantikoa aberatsagoa du. Hau da, WordNet-en *synset* batek bere tasun sintaktiko-semantikoak hierarkiatik jasotzen ditu; EuroWordNet-ek *synset* hauek guztiak ditu, eta gainera Goi- eta Domeinu-ontologiatik datozkionak.

5.2.2. atalean *run* aditza hartu dugu adibide gisa, WordNet-en dagokion *unique beginner*-ak (*[travel, go, move, locomote]* delako *synset*-ak) mugimendu tasuna ematen diola ikusteko. EuroWordNet-en *run*-en *synset* honek berak, tasun hori izango du, baina horretaz gain, interfazean bertan Goi-ontologiako *dynamic* eta *location* tasunak ere ikusten ditugu²².

²²Aditz honek Goi-ontologiako bi adabegietan du hastapena.

Top ⁹	
1stOrderEntity ¹	2ndOrderEntity ⁹
Origin⁹ Natural ²¹ Living ³⁰ Plant ¹⁰ Human ¹⁰⁴ Creature ⁷ Anima ²³ Form⁸ Artifact ¹⁴⁴ Substance ³² Solid ⁴³ Liquid ¹³ Gas ¹ Object ⁴² Composition⁸ Part ²⁴ Group ⁴³ Function¹¹ Vehicle ⁵ Representation ¹² MoneyRepresentation ¹⁸ LanguageRepresentation ¹⁴ ImageRepresentation ⁹ Software ⁴ Place ⁴⁷ Occupation ²³ Instrument ¹⁰ Garment ³ Furniture ⁴ Covering ⁵ Container ¹² Comestible ³² Building ¹³	SituationType⁴ Dynamic ¹³⁴ BoundedEvent ¹⁰³ UnboundedEvent ⁴⁰ Static ²⁵ Property ⁴¹ Relation ³⁰ SituationComponent⁶ Cause ⁴⁷ Agentive ¹⁷⁰ Phenomenal ¹⁷ Stimulating ²⁵ Communication ⁷⁰ Condition ⁴² Existence ²⁷ Experience ⁴³ Location ⁷⁴ Manner ²¹ Mental ⁶⁹ Modal ¹⁰ Physical ¹⁴⁰ Possession ²³ Purpose ¹³⁷ Quantity ³⁹ Social ¹⁰² Time ²⁴ Usage ⁵
3rdOrderEntity ²³	

23. irudia: EuroWordNet-eko Goi-ontologia.

24. irudian *Dynamic* eta *Location* dira Goi-ontologiako markak. Nahiz eta Goi-ontologiako tasunak *run* aditzaren *synset*-ean bertan ez egon, bere hiperonimoetatik jaso egiten ditu. EuroWordNet-en tasun hauek ez dituzte *synset*-ez *synset* adierazten, defendatzen dutelako hierarkiari esker herentziaz jaso daitezkeela.

Azalduriko ezaugarriek —eleaniztasunak eta ikerkuntzarako erabilgarria izateak, alegia— oso egoki bihurtu dute ezagutza-base hau LNPrean barnean erabiltzeko, batik bat, informazio erauzketa elebakar eta elebidunerako (Cuyper *et al.*, 1997; Gilarranz *et al.*,

lock 23
01097341v **move_by_running_1** **run_1** move fast by using one's feet, with one
 motion lock 11 **correr_5** foot off the ground at any given time
 lock 11 **córrer_2**

•

lock 35 **hurry_1** **zip_2** **travel_rapidly_1** **speed_3**
01175685v lock 14 **acelerar_4** **darse_presa_1** **apresurarse_3**
 motion lock 14 **acelerar_4** **donar-se_pressa_2** **cuitar_5** **apressar_4**
afanyar-se_3

◦

lock 761 **go_1**
01046072v **travel_1** **move_1** change location; move, travel, or
 base **locomote_1** proceed; "How fast does your new car
 concept lock 417 **desplazarse_1** go?" "We travelled from Rome to
 motion **moverse_4** **ir_2** Naples by bus"; "The policemen went
 Dynamic lock 416 **anar_1** from door to door looking for the
 Location **moure's_3** suspect
desplaçar-se_1

24. irudia: *Run* aditzaren *synset* bat eta bere hiperonimoak EuroWordNet-eko interfazean.

1996; Vossen, 1997). Eta arrazoi horregatik, gaur egun, hainbat WordNet berri sortzen ari dira (katalana, portugesa, grekoa, suediarra, errumaniarra, bulgariarra, norvegiarra, lituaniarra, errusiarra. . .) EuroWordNet-en ezagutza-basean oinarrituta. IXA taldean ere euskararako WordNet-a garatzen hasi gara (Agirre *et al.*, 2002). EuroWordNet <http://nipadio.lsi.upc.es/cgi-bin/public/weil.consult.perl> web orrian dago eskuragarri.

5.2.6 The Multilingual Central Repository (MCR)

The Multilingual Central Repository (MCR) interfaze eleanitza da, non Europa Batzordeko "MEANING: Developing Multilingual Web-Scale Language Technologies" (IST-2001-34460) proiektuan (Rigau *et al.*, 2003) aztertu den informazio guztia integratzen den. Ezagutza-base honek EuroWordNet-en eredia jarraitzen du.

MCRk bost hizkuntzetako WordNet-ekin egiten du lan: euskara, katalana, ingelesa, italiara eta gaztelania. MCR bost hizkuntza horien izen, aditz, adjektibo eta adberbioen adieren inbentarioa da, eta EuroWordNet-en eredia jarraitzen duenez, hizkuntza guztiak lotuta daude. Horregatik, hizkuntza bateko *synset* batekin beste hizkuntzetakoa ere ikusgarri dago.

EuroWordNet WordNet-en garapena denez, MCR EuroWordNet-en bertsio aurreratuagoa ere da. MCR WordNet eta EuroWordNet-en informazioaz baliatzen da, eta honetaz

gain, informazio berria dakar:

- Domeinu-ontologiaren bertsio aberatsago bat:

EuroWordNet-eko domeinuak ugaritu eta orraztu dituzte²³, hierarkian egon zitezkeen irregulartasunak gainditzeko. Bestalde, entitate edo izen bereziei domeinuak esleitu dizkiete, eta horren ondorioz, egitasmo horren emaitza domeinuka antolatutako izen berezi eta entitateen ezagutza-base bat da.

- Hautapen-murritzapenak:

Hautapen-murritzapenen azterketa automatikoa egin da (Agirre *et al.*, 2003; Carroll *et al.*, 2003), hau da, teknika konputazional desberdinak erabiliaz zenbait corpusetatik (*Egunkaria*, *British National Corpus* eta *SemCor*-etik hain zuzen ere) aditzen hautapen-murritzapen usuenak zeintzuk diren ikasi dira, ondoren, MCRn erlazio semantiko gisa hauen berri izateko.

Hortaz, MCR ezagutza-basean aditzen hautapen-murritzapenak kontsulta daitezke *Role* erlazio semantikoa erabilita. Erlazio semantiko honek, zenbait aukera eskaintzen ditu, begiratu nahi dugun erlazioaren arabera:

- *Role agent*
- *Role direction*
- *Role instrument*
- *Role location*
- *Role patient*
- *Role source location*
- *Role target direction*

Corpusetako datuetan oinarrituz, erlazio semantiko hauen guztien bitartez aditz batekin ager daitezkeen ale lexikoak eta har ditzaketen errol-thematikoak bereizteko gai dira. Ondorioz, MCRn aditzaren errol-thematikoen berri ematen duen erlazio semantikoa dago.

MCRn, ale lexikalak ontologian kategoriaka antolatuta daudenez (WordNet eta EuroWordNet-en bezala) *Role* erlazioak inplizituki ere azpikategorizazioaren berri ere eman dezake. Hortaz, *Role* erlazioak *edan* aditzaren *Role patient* erlazioaren bidez aditz horri buruz esango liguke bere hautapen-murritzapena *edari* dela²⁴, izena kategoria duela eta bere errol-thematikoa *gaia (patient)* dela. Hurrengo urratsetan, sintaxi-semantikari buruzko informazio gehiago, hala nola, funtzio gramatikalak, txertatzeko asmoa dago.

MCRren kontsultarako interfazea ondorengo web orrian dago: <http://nipadio.lsi.upc.es/cgi-bin/mcrWei/public/wei.consult.perl>.

Horrela bada, WordNet eta EuroWordNet ezagutza-basearen ildotik jarraituz, MCRk erakutsi du hasieran egitasmo semantiko eta psikolinguistiko soilekin burutu zen ezagutza-basea, informazio sintaktiko-semantikoa jasotzeko ere baliagarria izan daitekeela. Proiektu

²³EuroWordNet-eko hainbat domeinu gehiago zehaztu dituzte, “azpidomeinuak” sortuaz. Esate baterako, *jokatu* aditzak kirol adiera duenean, EuroWordNet-eko *free time* domeinua, *sport* jarrita zehaztu dute.

²⁴*Edan* aditzaren hautapen-murritzapena *edari* eta honen hiponimo guztiak ere badira.

honen hurrengo urratsetan MCR informazio sintaktiko-semantiko gehiagorekin (azpikategorizazioa, diatesi-alternantziak, Dorr-en ELKak (1997), erlazio semantiko konplexuagoak eta abar) osatzeko asmoa dago.

5.3 Volem

5.3.1 Sarrera

Volem proiektuaren (Fernández *et al.*, 2002) helburua zera da: Pirinio inguruko hizkuntza batzuetako (gaztelania, katalana eta frantsesa) aditz eta preposizioen ezaugarriekin EBLIS bat eraikitzea.

Proiektuan parte hartzen duten ikertaldeen —GRIAL eta Institut de Recherche en Informatique de Toulouse (IRIT)—, azterketa sintaktiko-semantikoan jardun dutenez, dagoeneko badute beraien hizkuntzetako preposizio eta aditzen deskribapen eta sailkapen bat. Volem proiektuaren lehenengo egitasmoa, beraz, ikertalde guztien lana bateratzen duen ezagutza-base eleanitza egitea da.

Volem proiektuko ezagutza-basean hurrengo informazioa dago laburbildurik:

- Aditz eta preposizio bakoitzaren izaera sintaktikoaren deskribapena: azpikategorizazioa, hautapen-murriztapenak eta aditzen alternantziak.
- Aditz eta preposizio bakoitzaren informazio semantikoa: *Egitura Lexikal-Kontzeptuala (ELK)*, errol-thematikoak eta aditzen kasuan, WordNet-eko klase semantiko nagusiena.

Hauek guztiak zehatzago aztertuko ditugu hurrengo ataletan.

5.3.2 Sintaxi-semantika elkargunea

Volem proiektuan parte hartzen duen hizkuntza bakoitzaren informazio sintaktikoa, argumentu, azpikategorizazio, hautapen-murriztapen eta alternantziez definitzen dute. Semantika, berriz, zerrendatutako errol-thematikoen zerrenda bat erabilia zehazten dute. Informazio hau guztia adierazteko eta antolatzeke Jackendoff-en ELKak (1990) erabiltzen dituzte: alde batetik aditz eta preposizioen informazio sintaktiko-semantikoa adierazteko egokiak direlako²⁵, eta bestetik, ELKak eleaniztasunarekin bateragarria dela frogatu delako (Dorr, 1993, 1997). Hortaz, frantses, katalan eta gaztelaniako adiera bera duten ale lexikalek ELK bera izango dutenez, ezaugarri sintaktiko-semantiko hauen jokaera eleanitza aztertzeke aukera ematen digu.

Nola adierazten dute, sintaxi-semantika elkargunea? Ikus dezagun Fernández *et al.*-en lanetik (2002) hartutako adibidea:

²⁵Jackendoff-ek aditzak (1990) eta preposizioak aztertzen ditu bereziki, kategoria sintaktiko hauek baitira azpikategorizazio aberatsenak dituztenak.

(52) Common part to the three languages:

thematic grid: [*inic(ag, tc), th*]
 (e.g. agent or causal initiator, theme)

LCS representation (interlingua) to which the verbs point (LCS 37):

[*event CAUSE*([*thing I*],
 [*event BECOME+ char,+ident*([*thing J*],
 [*state STATE*])]

LCS: Literally: I (subject) caused an object J to undergo a change of state in its ontological universe, BE-COMING (achievement) STATE.

Spanish lexical database:

Spanish verb: cerrar(to close)

Sense number: 75

Alternations + examples:

caus-2np:

El viento cerró las ventanas de golpe
 (the wind closed the windows)

caus-aux(hacer)-compl-2np:

El golpe de aire hizo que la puerta se cerrara
 (the wind made the door closed)

anti-pr-np:

La puerta se ha cerrado
 (someone closed the door)

pas-pr-np:

A las 7, se cerraron las puertas
 (At 7, doors are closed)

pas-aux(ser)-part-np:

La caja ha sido cerrada
 (the door was closed)

anti-aux(dejar)-part-np:

Ha dejado cerrada la puerta
 (He left the door closed)

result-aux(estar)-part-np:

La puerta está cerrada
 (the door is closed)

Catalan lexical database: Catalan verb: tancar (to close)

Sense number: 75

The Catalan uses are very similar to the Spanish ones.

Alternations + examples:

caus-2np:

El vent va tancar les finestres de cop
 (the wind closed the windows)

caus-aux(hacer)-compl-2np:

El cop de d'aire va fer que la porta es tanqués
 (the wind made the door closed)

anti-pr-np:

La porta s'ha tancat
 (someone closed the door)

pas-pr-np:

A les 7 can tancar les portes
 (At 7, doors are closed)

pas-aux(ésser)-part-np:

La capsa ha estat tancada
 (the door was closed)

anti-aux(estar)-part-np:

La porta està tancada
 (He left the door closed)

result-aux(deixar)-part-np:
Ha deizat tancada la porta
 (the door is closed)

French lexical database:

French verb: fermer (to close)
 Sense number: 75

Alternations + examples:

caus-2np:

Le vent ferme les fenêtrés d'un coup
 (the wind closed the windows)

caus-aux(hacer)-compl-2np: is not acceptable:

Le coup de vent fait que la porte se ferme
 (the wind made the door closed)

anti-refl-np:

La porte s'est fermée (d'un coup)
 (approx.: someone closed the door)

pas-pr-np:

A 7 heures, les portes ferment
 (At 7, doors are closed)

pas-aux(être)-part-np:

La porte est fermée
 (the door was closed)

anti-aux(avoir)-part-np:

Il a laissé la porte fermée
 (He left the door closed)

anti-pr-np: (middle reflexive)

La porte se ferme facilement
 (The door closes easily)

Lehenik, adiera bereko ale lexikoek (kasu honetan, *cerrar*, *tancar* eta *fermer*) hizkuntza guztietan duten antzekotasun semantikoa (*Common part to the three languages*) errol-thematiko eta ELK baten bidez definitzen dute; eta gero, hizkuntza bakoitzak sarrera lexiko hori (zehaztutako adiera horrekin) izan ditzakeen alternantziak zerrendatzen dituzte. Ezagutza-base eleanitza denez, azalpenak eta argibideak ingelesez ematen dituzte.

Alderdi eleanitza (alternantziak) desberdina izan arren, hizkuntza guztiek beti izango dute komunean adierazpen semantikoa, hots, *cerrar* aditzak beste adiera bat balu, honi legokiokeen ELK beste hizkuntzetan ere bera izan behar luke. Hala, Volem proiektuan ale lexikalaren adiera bakoitza banaka erabiltzen dute, eta adiera aldaketarekin batera gehienetan ezaugarri sintaktikoak ere ondorioztatzen dute.

Bestalde, proiektu honetan adiera ikuspuntu zabal batetik lantzen dute (WordNet-ekin erkatuz adibidez²⁶), era horretara hizkuntzen adieren artean parekotasunak errazago egin daitezkeela uste dutelako.

Nahiz eta adiera mailan WordNet zehatzegia iruditu, aditzen adierak WordNet-eko klase semantiko nagusien arabera antolatzen dituzte (*verbs of possession, verbs of movement, verbs of consumption...*). Klase oso orokorrak direnez, aditz bat klase semantiko bakarrean kokatzea ataza zaila izan daiteke, baina klase hauen azpitalde edo hiponimo multzoak (zehaztapenak direnak, noski) kontuan hartuz gero, jakin dezakegu aditz hori zer klaseri dagokion.

These classes may be quite large, but considering subclasses, with a more specialized meaning, greatly contributes to the homogeneity and to the coherence of the descriptions. Conversely, we can also better identify contrasts between semantically close verbs and make explicit their differences in meaning. (Fernández *et al.*, 2002, 128. orr.)

²⁶Ikus 5.2. atala.

5.3.3 Aditza, izena, preposizioa/postposizioa, adberbioa eta adjektiboa

Honezkero, Volem proiektuak aditzak eta preposizioak lantzen dituela erakutsi dugu, eta adibide bat ere azaldu dugu²⁷, non informazio sintaktiko-semantikoa ageri den, hau da, azpikategozizazioa, hautapen-murritzapenak eta errol-thematikoak.

- Azpikategorizazioa:

Aditzak hartzen dituen argumentu semantikoaren kopurua zehazten dituzten arren (*Common part to the three languages* delakoan), beraien kategoria sintaktiko posibleak alternantzetan definitzen dituzte. Beraz, argumentuaren alderdi semantikoa eta sintaktikoa ez daude zuzenean lotuta, hau da, (52)n adibidez, ez zaigu esaten *egilea* zer kategoriarekin agertzen den.

- Hautapen-murritzapenak:

Jackendoff-en eredu jarraitzen dutenez, hautapen-murritzapenak egitura kontzeptualean adierazten dituzte, ELKko primitibo eta kategoria kontzeptualetako informazioa baliatuz²⁸.

- Errol-thematikoak:

Proiektu honetan ikertaldeek errol-thematikoen zerrenda bateratu bat adosten dute, zeinek argumentu-egitura osatzeko. Proposatzen duten errol-thematikoen zerrenda ondorengoa da:

- (a) *Initiators*: Ekintza bat nahita edo nahigabe “gertarazten” duen *egile* edo *gai kausala*.
- (b) *Themes*: Ekintza jasaten duena eta hauek *holistic* (ez afektatuak) edo *incremental* (*affected*) *beneficiary or victims* delakoak izan daitezke.
- (c) *Localizations*: *spatial*, *temporal* eta *abstract* bezalakoak direnak.
- (d) *Isolated roles*: *quantity*, *instrument*, *accompaniment* eta abar bezalakoak.

Hauetaz gain, Jackendoff-en primitiboak darabiltzate²⁹.

²⁷ (52) adibidean aditzaren errepresentazioa agertu dugun arren, preposizioena ere oso antzekoa da.

²⁸ Argibide gehiagorako ikus 3.1 atala.

²⁹ Argibide gehiagorako ikus 3.1 atala.

5.3.4 Erabilera

Proiektu honetan gaztelaniako, frantseseko eta katalaneko 1700 aditz-forma aztertu dira, eta informazio hori eskuragarri dago hurrengo web orrian: <http://www.irit.fr/recherches/ILPL/Site-Equipe/demonstrations.html>.

Egun, proiektu honen bigarren zatiari ekin zaio (Volem2), zeinetan Volem1-eko aditz eta preposizioei euskara eta okzitaniararen informazioa gehitu nahi zaizkion.

Ezagutza-base hau LNPrako interesgarria izan daitekeen arren, egun oraindik eraikitze bidean dagoenez, honekin ez dira aplikazio ugari ezagutzen. Dena den, hasiberria den proiektu bat aipatu dezakegu: SenSem (*Sentence Semantics*) proiektua MCyT (BFF2003-06456). Proiektu honetan corpus etiketatu bat eraikitzen ari dira erdi-automatikoki eta bere helburu nagusia, Volem-eko informazioa orraztu eta aberastea da. Horretarako, corpusean predikatuekin agertzen diren alternantziak Volem-en zerrendatuak dituzten predikatuekin erkatzen dituzte, ezagutza-basean dituztenak zuzenak diren ala ez egiaztatzeko eta ez dituztenak gehitzeko.

6 ONDORIO NAGUSIAK ETA ETORKIZUNERAKO LANAK

Sarreran aipatu dugun bezala, IXA taldeak LNPn dihardu, eta orain arte batez ere morfologia eta sintaxiaren arloak landu ditu. Arlo hauetan lan handia egiteke dagoen arren, hurrengo aurrerapauso garrantzitsua semantika jorratzea da. Semantikan eta beste ataza konputazionalaetan aurrera egin ahal izateko (egitura sintaktikoen desanbiguazioan, hitzen adieren desanbiguazioan, anaforaren ebazpenean eta itzulpen automatikoan besteak beste), IXA taldeak EBLs baten beharra du. EBLs honetan ale lexikalen semantika adierazita egoteaz gain, hauen informazio sintaktiko-semantikoa ere adieraztea nahitaezkoa da.

Zeregin honetan, lehenago egin diren EBLsen aurreazterketa bat egin behar da, euskarakoa sortzerakoan jarrai dezakegun eredu bat izateko. Horixe izan da lan honen helburua: EBLs-ak landu dituzten zenbait formalismoren azterketa egin, gerora hortik euskararako EBLsa eratzeko. Ezagutza-baseak erabilera konputazionala izango duenez, oihartzun konputazionala izan duten lan eta formalismoetara mugatu gara.

Bestalde, lan hauek fitxa baten bidez deskribatu ditugu (2. atalean zehaztu dugun bezala), eta baita gure ustez EBLs egokia izateko beharrezkoak diren irizpideak zehaztu ere (1.4 atalean). Guztira hamar teoria eta lan aztertu dira: Jackendoff (1990), Levin (1993), Pustejovsky (1995), LFG (Bresnan eta Kaplan, 1982), GPSG (Gazdar *et al.*, 1985), HPSG (Pollard eta Sag, 1994), FrameNet (Fillmore eta Baker, 2001a), WordNet (Miller, 1985; Fellbaum, 1998)), EuroWordNet (Vossen, 1997), MCR (Atserias *et al.*, 2004) eta Volem (Fernández *et al.*, 2002).

Ondorio nagusi gisa, eta aipatu ditugun irizpideen arabera, IXA taldearen beharretara gehiago egokitzen den EBLs formalismoa WordNet eta honen ildotik abiatuta garatu diren EuroWordNet eta MCR direla esango dugu, erabaki horren arrazoiak ondoren azpimarratuko ditugularik.

Honekin batera, WordNet, EuroWordNet eta MCRk informazio gabezia non duten agerian jarriko dugu, nahiz eta informazio hori eransteke erraztasuna baduela ere ikusiko dugun. Honi lotuta azalduko dugu, aztertu ditugun gainontzeko lanetatik euskarako EBLsrako zein informazio izan daitekeen baliagarria eta zergatik.

6.1 Proposamenaren arrazoiak

- a) **Hizkuntzaren ikuspuntu orokorra:** WordNet (EuroWordNet eta MCR) lexiko zabala eta garatua duten EBLSak dira. Hauek adieran oinarritutako ontologiak dira, hizkuntzaren lexikoa ezagutza-base batean jaso nahi dutenak, ale lexikalak, ale lexikalen adierak, klase semantikoak, kategoriak, eta hauen guztien arteko erlazio semantikoak kontuan izanda (5.2, 5.2.5 eta 5.2.6 ataletan azaldu dugun bezala). Noski, hizkuntzaren lexikoak ez du mugarik eta etengabe garatzen dauden ezagutza-baseak dira. Hala ere, hizkuntzaren ikuspuntu orokorra eman dezaketen ezagutza-baseak ditugu. Esate baterako, WordNet-ek 115.424 *synset* ditu (79.689 izen, 13.508 aditz, 18.563 adjektibo eta 3.664 adberbio)¹. MCRk WordNet ezagutza-basearen tamaina bera du, baina erlazio semantiko gehiagorekin (milioi bat inguru).
- b) **Eredu deskribatzailea:** WordNet ez dago teoria bakar bati lotua, hots, teoria ezberdinek erabil dezaketen EBLSak dira. Bestalde, EuroWordNet eta MCR berak, WordNet-en garapenak dira, WordNet beste oinarri eta ikuspuntu teoriko eta konputazionalatik informazio gehiagorekin aberastu dutelarik.
- c) **Inplementazioa:** WordNet, EuroWordNet eta MCR inplementatutako EBLSak dira, hots, praktikotasunera jotzen dute. EBLSa sortu eta lantzeaz gain, hainbat erabilera izan ditzakeen ezagutza-base publikoak dira, kontsultagarriak, alegia. Hauek aztertzerakoan aipatu ditugu beraien erabilera nagusienak (hiztegi eta thesaurus gisa adibidez).

EuroWordNet eta MCRrekin inplementatzeko aukerak areago doaz, EBLS hauek eleanitzak direlako, ingeleseko WordNet-ari beste hainbat hizkuntza gehitu dizkitelako (daniera, italiera, gaztelania, alemana, frantsesa, txekiera, estoniera...) eta horien artean euskara txertatzen hasi gara (Agirre *et al.*, 2002).

- d) **LNPn erabilgarria:** Hiru EBLS hauek oso erabiliak izan dira LNPrean arlo oso ezberdinetan: galdera-erantzunean, informazio erauzketan, itzulpen automatikoan eta adiera desanbiguazioan besteak beste (hauei buruzko argibide gehiago 5.2.4, 5.2.5 eta 5.2.6 ataletan).
- e) **Beste lan batzuekiko gertutasuna:** Txosten honetan aipatutako formalismo eta lan teoriko askok gerora WordNet eta EuroWordNet-en adiera edo/eta klase semantikoekin aberastu dituzte². Esate baterako, Dorr-ek (1997) Jackendoff-en lanarekin egin duen bezala. Dorr-ek Jackendoff-en *Egitura Lexikal-Kontzeptualetan* (ELKetan) oinarritutako EBLSa eraiki du eta ELK hauek WordNet-eko adieretara lotuak daude. Lan horretan bertan, Dorr-ek Levin-en aditz-klaseetako aditzak ere WordNet-eko aditzekin lotzen ditu. Ildo honetatik jarraitu duen formalismoa Volem izan da. Bestalde, Pustejovsky-ren lexikoaren ezaugarri batzuk WordNet-ekoekin lotzeko saiakera ere egin da (Buitelaar, 1998). Formalismo ezberdin hauen arteko uztardura oso baliagarria eta aberatsa da, beti ere, WordNet-en ildotik euskararako egingo den EBLSak hauez guztiez balia baitaiteke euskarako EBLSan informazio gehiago erans-teko. Beraz, garbi dago WordNet eta EuroWordNet-ek LNPrean arloan baliabide oso erabiliak izan direla, eta egun oraindik hainbat esperimendu eta ikerlanetarako iturburu direla.

¹WordNet-en azkeneko bertsioaz ari gara, 2.0 bertsioaz, alegia.

²MCR orain dela gutxi EBLSa izanda, oraindik ez da horrela erabili.

6.2 Gabeziak

WordNet eta EuroWordNet ezagutza-base hauetako informazio sintaktiko-semantikoa mugatua da. Adibidez, ez dituzte azpikategorizazioa, hautapen-murritzapenak eta errol-thematikoak zehazten. Hala ere, hautapen-murritzapenen kasuan, WordNet eta EuroWordNet-en hauek inplizituki badaude, baina EBLSan oraindik ez daude erlazio baten bidez esplizituki adierazita, hau da, WordNet eta EuroWordNet-eko EBLSek ez dute eskaintzen *edan* eta *edari* sarrera lexikalak hautapen-murritzapen gisa lotzeko biderik. Hau oztopo bat da euskarako EBLS bat hauetan oinarrituta egiteko, lanaren hasieratik esan dugun bezala (1.4 atalean), euskarako EBLSan ale lexikalen adierez gain, hauen informazio sintaktiko-semantikoa adierazita etortzea nahiko genukeelako.

Den den, MCRn horrelako informazioa esplizitua egiten saiatzen dira, esate baterako, hautapen-murritzapenak MCRn adierazita daude *Role* erlazio semantikoaren bitartez (*Role agent*, *Role patient* besteak beste³). Hortaz, erlazio semantiko honen bitartez ale lexikal batekin ager daitezkeen ale lexikal motak, hots, hautapen-murritzapenak, eta azken hauek har ditzaketen errol-thematikoak bereizteko gai dira. MCRn ale lexikalak ontologian kategoriaka antolatuta daudenez (WordNet eta EuroWordNet-en bezala) *Role* erlazioak inplizituki ere azpikategorizazioaren berri ere eman dezake. Hortaz, *Role* erlazioak *edan* aditzaren *Role patient* erlazioaren bidez aditz horri buruz esango liguke, bere hautapen-murritzapena *edari* dela⁴, honek izena kategoria duela, eta bere errol-thematikoa *gaia* (*patient*) dela. Hurrengo urratsetan, sintaxi-semantikari buruzko informazio gehiago txertatu nahi da, hala nola, funtzio gramatikalak.

Euskarako EBLSa MCRren ereduan eraikiz gero, honek WordNet eta EuroWordNet-en hezurdura izango luke, hots, adieraka antolatutako EBLS semantiko eleanitz baten abantailak izango genituzke, eta gainera, bi ezagutza-base hauetan dagoen informazioarekin batera, MCRn gehitu (eta gerora gehituko) den informazio sintaktiko-semantikoa eskuragarri izango genuke.

6.3 Beste lanetako informazioa

Hala, euskarako EBLSa MCRren eredua jarraituz egin behar dugu eta MCRk beste iturrietako informazioa jasotzeko oinarri sendoa duenez, ikerlan honetan landu ditugun beste lan eta formalismoetatik baliagarria izan dakigukeenari probetxu atera diezaiokegu hau MCRn sartuz gero. Alde batetik, EBLSan ale lexikalak sailkatzeko erabiltzen dituzten ezaugarri batzuk, eta MCRn ez daudenak, aprobeitza genitzake. Eta bestetik, MCRn sarrera lexikalak jasotzen ez duen informazioa beraietatik har genezakeen. Kontuan izan beharrekoa da lan hauek ingeleserako pentsatuak daudela, eta beraietan dagoen informazioa euskararako EBLSari gehitu baino lehen, informazio hori hizkuntzatik independentea den (unibertsala den) edo behintzat euskararako baliagarria den frogatu beharko genukeela⁵.

Jarraian, lan honetan landutako ikerlan eta formalismoetatik atera ditugun ondorio batzuk aipatuko ditugu, eta hauetatik MCRn sartzeko baliagarria izan daitezkeen informa-

³Ikus 5.2.6 atala.

⁴*Edan* aditzaren hautapen-murritzapena *edari* eta honen hiponimo guztiak dira.

⁵Beste ikerkuntza lanean horrelako esperimendu baten berri ematen dugu. Ingeleserako corpusetatik automatikoki lortutako hautapen-murritzapenak, euskaratu eta euskararako baliagarriak diren aztertu dugu (Agirre *et al.*, 2003; Pociello, 2004).

zioa dagoen ala ez ere adieraziko dugu.

Hizkuntzalaritza teorikotik hiru lan aztertu ditugu: Jackendoff (1990), Levin (1993) eta Pustejovsky (1995). Hasteko, hiru lan hauek ordura arte ez zegoen formalismo berri baten adierazle dira, beraz, ez daude beste formalismoetatik gertu, bakarrak dira. Ikusiko dugun bezala, ondorengo lanetan uztartuko dituzte formalismo hauek beste formalismo ezberdinekin. Honetaz gain, bakoitzetik beste ondorio batzuk atera ditugu.

- a) **Jackendoff:** Jackendoff-en *Egitura Lexikal-Kontzeptualak* (ELKak) ale lexikalari buruzko informazio ugari dakar, batez ere sintaxi-semantika elkarguneari buruzko informazioa. Aditzak eta preposizioak azpikategorizazio aberatsena duten kategoriak direnez, eta hortik informazio sintaktiko-semantiko ugari lor daitekeenez, Jackendoff kategorია hauen adierazpenetara mugatu da, izenak, adjektibo eta adberbioak alde batera utziaz.

Bestalde, Jackendoff sintaxi-semantika elkargunearen errepresentazioaz arduratzen denez, lan honetan adiera bigarren mailan dago, hots, hitzen anbiguotasuna ez zuen esplizituki kontuan hartu. Adierarekin bezala, klase semantikoak ere ez ditu esplizituki lantzen, nahiz eta inplizituki batzuen berri ematen duen (adibidez, *ukipen aditzak* (*contact verbs*) aipatzen ditu baina ez du klase hau osatzen duten zerrenda ematen).

Inplementazioari begira, Jackendoff-en ELKak konputazionalki implementatuak (eta aberastuak) izan dira, esaterako, Dorr-ek (1997) Jackendoff-en ELKetan oinarritutako aditzen EBLSa egin zuen, hauek Levin (1993) eta WordNet-eko adiera eta klase semantikoekin aberastuz. Hauetan oinarrituta tutore sistemak eta itzulpengintza automatikoa landu ditu⁶.

Ildo beretik, Volem (Fernández *et al.*, 2002) proiektuak (ikus 5.3 atala) garatutako EBLSa dago. EBLS hau gaztelaniako, frantseseko eta katalaneko aditz eta preposizioetara mugatzen da, aditz eta preposizio bakoitzaren izaera sintaktikoaren deskribapena (azpikategorizazioa, hautapen-murritzapenak eta alternantziak) eta informazio semantikoa (ELKa, errol-thematikoak eta aditzen WordNet-eko klase semantiko nagusia) ematen duelarik.

Jackendoff-en lanetik abiatutako bi EBLS hauek Jackendoff-en lanari alderdi semantiko eta beste ikuspuntu teorikoak gehitu arren, aditz eta preposizioetara murritzten dira eta ondorioz, hauek ere ez dute hizkuntza bere osotasunean adierazten. Hala ere, Jackendoff-en lanetik abiatutako bi EBLS hauetan aditzen klaseak WordNet-eko adieretara lotuak daudenez, euskarako aditzen sailkapenerako erabilgarria izan dakiguke, noski, lehendabizi bertan dagoen informazioa euskararen izaera sintaktiko-semantikora egokitzen dela egiaztatu beharko genukeen⁷.

- b) **Levin:** Levin-en ikerlana (1993) ikusiko ditugun lanetatik pixka bat aldentzen da. Levin-ek ez baitu EBLSa bat egiten, horren ordez, Levin-ek aditzen EBLS bat osatzeko behar diren baliabide linguistikoak eskaintzen ditu. Bere liburuan ingeleseko aditzen sintaxia eta semantika sakonki aztertzen du eta liburuan bertan landutako aditzen zerrenda ematen du, bakoitzaren adiera, klase semantiko, diatesi-alternantzia eta informazio sintaktiko-semantikoaren idatzizko deskribapenarekin batera.

⁶Lan honen berri 5.2 atalean eman dugu.

⁷IXA taldean Volem proiektuaren jarraipenean parte hartu duenez, horrelako esperimenduak egiteko aukera emango digu, euskara EBLS honekin bateragarria den aztertuko dugulako.

Euskararako EBLsaren proposamenerako baztertu egin dugu aditzen deskribapena bakarrik egiten duelako. Bestalde, gure EBLs-ak aditzak lantzeko baliabide horien azterketatik informazioa deduzitu daitekeen arren, aditz bakoitzerako informazio hori ateratzea, lan handia eskatzen duen ataza da. Gorago aipatu dugun bezala, Dorr-ek Levin-en klase semantikoak WordNet-era lotuak ditu. Horrela bada, MCRren ildotik egindako euskarako EBLs-erako, Levin-en lanetik zuzenean informazioa atera ordez, Dorr-en lanetik abiatzea errazagoa litzaiguke. Horretarako, bete beharreko lehenengo pausua, Levin-en aditz-klaseak eta MCRkoak zer puntutaraino pareka daitezkeen aztertzea litzateke.

Horrekin batera, Aldezabal-en (2004) tesi-lanean Levin-en lana erabili da euskal aditzaren azpikategorizazioa jorratzeko, eta bertan adierazten da Levin-en alternantzia asko euskararako ez direla baliagarriak (ingeleseko 80 alternantzietatik 24 bakarrik datoz euskararekin bat). Hortaz, hizkuntzen arteko desegokitasun hau dela eta, alde aurretik esan daiteke alternantziatan oinarritutako sailkapena ez dela bat etorriko; edota modu orokorrago batean planteatu behar direla.

c) **Pustejovsky:**

Hizkuntzalaritza teorikoari dagokion atalean, aztertutako azken lana Pustejovsky-rena (1995) izan da. Autore honen ustez lexikoak sistema kognitiboaren oinarritzko egitura kontzeptuala adierazteko gai izan behar du, eta gainera, ahalmen sortzaile ere izan behar du. Horretarako, kategoria guztien deskonposaketan funtsatutako teoria proposatzen du. Alde horretatik beraz, euskararako nahi dugun EBLs-ekin bat dator, hots, hizkuntza bere osotasunean aztertu nahi duen lana da.

Inplementazioari dagokionez, Pustejovsky-ren lanean oinarritutako ikerlan konputazional gutxi ezagutzen dugu, eta bere teoriatik abiatuta lexiko errealik ez dago. Hortaz, euskarako EBLs-arentzat nahiko genukeen beste ezaugarrietako bat ez du. Ezagutzen den implementazioetako bat Buitelaar-ena da (1998). Buitelaar-ek Pustejovsky-ren alderdi semantiko batzuk (*alderdi konstitutiboa* adibidez) WordNet-en dauden antzeko harreman semantikoekin erkatzen ditu. Berriro ere, euskarako EBLs-ari begira, WordNet-erako lotura hau ondo etor dakiguke Buitelaar-en lanetik lortutako emaitzak gure EBLs-ak eransteko.

d) **LFG, GPSG eta HPSG:** Hizkuntzalaritza teoriko eta konputazionalaren erdibidean dauden lan hauen inguruan, hasteko, hizkuntzalaritza teorikoko lanei buruz esandako gauza bera errepikatuko dugu: hiru lan hauek ordura arte ez zegoen formalismo berri baten adierazle dira, beraz, ez daude beste formalismoetatik gertu, bakarrak dira.

LFG, GPSG eta HPSG euskararekin erabiltzeko saiakera bat egin da (Gojenola, 1998), eta horretan agerian geratu zen hauen hasierako ingeleserako formulazioko ezaugarri eta erregela ugari ez zetozeela bat euskararen izaera linguistikoarekin. Hortaz, LFG, GPSG eta HPSG euskarari aplikatu ahal izateko lan linguistiko handia egin beharko litzateke.

e) **FrameNet:** Azkenik, hizkuntzalaritza konputazionalerako lanak izan ditugu aztergai: FrameNet (Fillmore eta Baker, 2001a), WordNet eta honen ildotik etorritakoak (Miller, 1985; Fellbaum, 1998; Vossen, 1997; Atserias *et al.*, 2004), eta Volem proiektua (Fernández *et al.*, 2002). WordNet, EuroWordNet, MCR eta Volem-i buruzko ondorioak gorago aipatu ditugunez, zuzenean FrameNet-i buruz jardungo gara.

FrameNet proiektua (Fillmore eta Baker, 2001a) ingeleserako baliabide lexikografikoa eraikitzen ari da, *Frame Semantics* (Fillmore, 1985) teoriaraino oinarritua eta corpus errealeko datuekin lagunduta.

Formalismo hau nahiz eta teoria bati lotua egon, corpus errealeko datuetan oinarritzen da, beraz, WordNet-ekin bezala inplementa daitekeen EBLSa da, hots, praktikotasunera jotzen dute. EBLSa sortu eta lantzearekin batera, corpus etiketatu bat eratzen ari dira eta horrek hainbat erabilerari bidea zabaltzen die (baita konputazionalak ere). Horretaz gain, FrameNet EBLS publikoa da⁸.

Oso EBLS interesgarria da, batez ere ikuspegi konputazionaletik, LNPre arlo ezberdinen ikasketarako oso baliagarria delako (ikus 5.1.5 atala) baina epe luzerako EBLS da, hau da, eremu batzuetara (komunikazioa, legedia, hezkuntza...) mugatutako lexikoa da, denborarekin hizkuntza bere osotasunean adierazteko helburua duena. Gure euskarako EBLSa, ordea, ezin da eremu zehatz horietara mugatu, hizkuntzaren bere osotasunean adierazteko gai izan behar du.

Bestalde, beraien sarrera lexikalak gobernatzaileak dira, (*target*), hots, aditzak, objektuak edo adjektiboak. Beste hitz batzuetan, EBLS hau predikatuetan oinarritzen da eta hauetatik kanpo geratzen dena ez dute aztertzen, hala nola, preposizioak eta adberbioak ez daude. Hala, ez dituzte kategoria guztientzako sarrera lexikalak, eta berriro ere, ez du gure euskarako EBLSarentzako nahi dugun osotasuna.

Azkenik, EBLS hau egiteko corpus bat erabiltzen ari dira, eta gerora corpus horretatik lortutako datuekin proposatzen dituzte predikatu baten azpikategorizazioa, *frame*-ak eta abar. Oso baliabide egokia da proposatutako *frame*-ak hizkuntza errealean erabiliak/egokiak diren ala ez egiaztatzeke. Noski, FrameNet EBLSarekin batera corpus bat etiketatzeak lan handia dakar, eta horrek EBLSaren garapena moteldu egiten du. Euskararako ez sintaktikoki etiketatutako corpusa (Aranzabe *et al.*, 2003) dugu, baina ez semantikoa. FrameNet-en metodologia jarraituz gero, EBLSa sortzen ari garen aldi berean eratu beharko genukeen corpus sintaktiko-semantikoa, eta gure EBLSaren eraikuntza atzeratuko lukeen arren, orain dugun corpus etiketatuaren etiketatze-ereduarekin ez litzateke bat etorriko.

Atal honetan arrazoitzen saiatu gara, euskarako EBLSa egiteko WordNet-en eredu (zehazkiago MCRrena) jarraitzea. Erabaki hori hartu dugunez euskarako EBLSrako nahiko genituzkeen ezaugarriak ondo definitu ondoren. Hauen arabera mugatu dugu gure proposamena. Honekin esan nahi duguna da ikerkuntzaren beste ikuspegi, helburu eta erabileraren arabera, aukeraketa hau ezberdina izango litzatekeela.

Bestalde, behin MCRren alde aukeraketa eginda, eredu hau beste lan eta formalismoe-tako informazioarekin osa dezakegula adierazi dugu. Hala ere, formalismo desberdinak direnez eta batzuetan beraien artean kontraesanean daudenez, hauen arteko hautaketa bat ere egin beharko genukeen, hau da, MCRren ildotik egingo den euskarako EBLSari hauetako zeinekin osatzea komeniko litzatekeen erabaki beharko genuke.

Dena den, lan hori ez dugu ikerlan honetan jorratuko, etorkizunerako lan gisa proposatuko dugu.

⁸FrameNet: www.icsi.berkeley.edu/framenet web.

6.4 Etorkizunerako lana

Etorkizunean euskarako EBLSa MCRren ereduari lantzen hasiko gara: izenen, aditzen, adjektiboen eta adberbioen adierak. Hau epe luzerako lana izango denez, bitartean adierak lantzearekin batera, gerora landu beharko dugun sintaxi-semantika elkargunea zehazten joango gara.

MCRren ildotik sortu behar dugun euskarako EBLSari informazio sintaktiko-semantiko gehiago gehitu behar zaio. Hori egin baino lehen, lehendabizi osatu beharreko informazio sintaktiko-semantikoaren ereduaren finkatzea da, eta hori egin ahal izateko lan honetan aztertutako formalismoen gain beste hainbat landu nahi ditugu, hala nola, Propbank (Palmer eta Kingsbury, 2003), VerbNet (Kipper *et al.*, 2000), Pirapides (Vázquez *et al.*, 2000), EAGLES (EAGLES, 1999, 1996), Fernando Gómez (Gómez, 1998), besteak beste.

Bestalde, kontuan izan beharrekoa da lan hauek ingeleserako pentsatuak daudela, eta hauek dagoen informazioa euskararako EBLSari gehitu baino lehen, informazio hori hizkuntzatik independentea den (unibertsala den) edo behintzat euskararako baliagarria den frogatu beharko genukeela. Horretarako, zenbait esperimentu egin beharko genituzke.

Irudien zerrenda

1	<i>run</i> aditzaren ELK.	14
2	<i>into</i> preposizioaren ELK.	15
3	<i>drink</i> aditzaren ELK.	17
4	<i>open</i> aditzaren sarrera lexikoa Pustejovsky-ren teorian.	24
5	<i>door</i> izenaren sarrera lexikoa Pustejovsky-ren teorian.	27
6	<i>David yawned</i> perpausaren adierazpena LFGn.	30
7	<i>David yawned</i> perpausaren osagai-egitura LFGn.	31
8	Osagai-egitura eta egitura funtzionalaren arteko uztardura LFGn.	31
9	<i>yawned</i> ale lexikalaren adierazpena LFGn.	31
10	Sintagmen adierazpena LFGn.	32
11	Sintaxi-semantika elkargunea LFGn (Bresnan eta Kaplan, 1982).	32
12	Funtzio gramatikalen tasunak LFGn (Bresnan eta Kanerva, 1989).	33
13	<i>give</i> aditzaren adierazpena LFGn proto-errolak erabilita.	35
14	<i>she</i> izenordainaren adierazpena HPSGn.	37
15	Sintagmen adierazpena HPSGn.	40
16	<i>gives</i> aditzaren adierazpena HPSGn.	40
17	<i>to</i> preposizioaren adierazpena HPSGn.	41
18	FrameNet-eko interfazearen adibide bat.	45
19	FrameNet-eko <i>understanding</i> izenaren etiketatzea FrameNet-en.	45
20	<i>avenge</i> aditzaren azpikategorizazioa corpuseko agerpenetan oinarrituta.	46
21	<i>revenge</i> izenaren azpikategorizazioa FrameNet-en.	50
22	<i>revenge</i> aditzaren azpikategorizazioa FrameNet-en.	51
23	EuroWordNet-eko Goi-ontologia.	60
24	<i>Run</i> aditzaren <i>synset</i> bat eta bere hiperonimoak EuroWordNet-eko interfazean.	61

Bibliografía

- Aduriz I., Agirre E., Aldezabal I., Alegria I., Ansa O., Arregi X., Arriola J., Artola X., de Ilarraza A.D., Ezeiza N., Gojenola K., Maritxalar A., Maritxalar M., Oronoz M., Sarasola K., Soroa A., Urizar R., eta Urkia M. A framework for the automatic processing of Basque. In *Proceedings of Workshop on Lexical Resources for Minority Languages*, Granada, 1998.
- Aduriz I., Alegria I., Arriola J., Artola X., de Ilarraza A.D., Ezeiza N., eta Urkia M. Euslem: un lematizador/etiquetador de textos en euskera. In *Actas del X congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*, Córdoba, 1994.
- Agirre E., Aldezabal I., eta Pociello E. A pilot study of English selectional preferences and their cross-lingual compatibility with Basque. In *Proceedings on International Conference on Text Speech and Dialogue (TSD)*, Czech Republic, 2003.
- Agirre E., Alegria I., Arregi X., Artola X., de Ilarraza A.D., Maritxalar M., Sarasola K., eta Urkia M. Xuxen: A spelling checker/corrector for Basque based in two-level morphology. In *Proceedings of ANLP'92*, Povo Trento, 1992.
- Agirre E., Ansa O., Arregi X., Arriola J., de Ilarraza A.D., Pociello E., eta Uria L. Methodological issues in the building of the Basque WordNet: quantitative and qualitative analysis. In *Proceedings of First International WordNet Conference*, Mysore (India), 2002.
- Agirre E., Atutxa A., Gojenola K., eta Sarasola K. Exploring portability of syntactic information from English to Basque. In *Proceedings of the 4rd International Conference on Language Resources and Evaluations (LREC)*, Lisbon, Portugal, 2004.
- Agirre E., García E., Lersundi M., Martínez D., eta Pociello E. The Basque task: did systems perform in the upperbound? In *Proceedings of the SENSEVAL-2 Workshop*, Toulouse, France, 2001.
- Agirre E. eta Lersundi M. Extracción de relaciones léxico-semánticas a partir de palabras derivadas usando patrones de definición. In *Proceedings of the Annual SEPLN meeting*, 2001.
- Agirre E. eta Lersundi M. Semantic interpretations of postpositions and prepositions: a multilingual inventory for Basque, English and Spanish. In *Workshop on The linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications*, Tolouse, France, 2003.

- Agirre E. eta Martínez D. Learning class-to-class selectional preferences. In *Proceedings of the Workshop "Computational Natural Language Learning"*, Toulouse, France, 2001.
- Agirre E. eta Martínez D. Integrating selectional preferences in WordNet. In *Proceedings of First International WordNet Conference*, Mysore, India, 2002.
- Agirre E. eta Martínez D. Exploring automatic word sense disambiguation with decision lists and the Web. In *Proceedings of the Semantic Annotation And Intelligent Annotation workshop organized by COLING*, Luxembourg, 2000.
- Aldezabal I. *Aditz-azpikategorizazioaren azterketa sintaxi partzialetik sintaxi osorako bidean. 100 aditzen azterketa. Levin-en (1993) lana oinarri hartuta eta metodo informatikoak baliatuz*. PhD thesis, UPV-EHU, 2004.
- Aldezabal I., Aranzabe M., Atutxa A., Gojenola K., Sarasola K., eta Goenaga P. Extracción masiva de información sobre subcategorización verbal vasca a partir de corpus. In *Actas del XVII Congreso de la SEPLN Universidad de Jaén*, Jaén, Spain, 2001.
- Alsina A. *Predicate Composition: A Theory of Syntactic Function Alternations*. PhD thesis, Stanford University, 1993.
- Aranzabe M., Arriola J., Atutxa A., Balza I., eta Uria L. *Guía para la anotación sintáctica manual de Eus3LB (corpus del euskera anotado a nivel sintáctico, semántico y pragmático)*, 2003.
- Arriola J. *Euskal Hiztegia-ren azterketa eta egituratzea ezagutza lexikalaren eskuratze automatikoari begira. Aditz-adibideen analisisa Murriztapen-gramatika baliatuz, azpikategorizazioaren bidean*. PhD thesis, UPV-EHU, 2000.
- Arriola J., Artola X., Maritxalar A., eta Soroa A. A methodology for the analysis of verb usage examples in a context of lexical knowledge acquisition from dictionary entries. In *Proceedings of EACL'99, Linguistically Interpreted Corpora*, Bergen, Norway, 1999.
- Artiagoitia X. *Hatsarreak eta parametroak lantzen*. Diputación Foral de Alava, 2000.
- Atserias J., Villarejo L., Rigau G., Agirre E., Carroll J., Magnini B., eta Vossen P. The MEANING Multilingual Central Repository. In *Proceedings of the 2nd Global WordNet Conference*, Brno, Czech Republic, 2004.
- Banerjee S. eta Pedersen T. An adapted lesk algorithm for word sense disambiguation using WordNet. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING-02)*, Mexico City, 2002.
- Barwise J. eta Perry J. *Situations and Attitudes*. Bradford Books. MIT Press, 1983.
- Boas H.C. Bilingual FrameNet dictionaries for machine translation. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, Las Palmas, Spain, 2002.
- Boguraev B. eta Briscoe T. *Computational Lexicography for Natural Language Processing*. Longman, John Wiley and Sons, London, New York, 1989.

- Bresnan J. eta Kanerva J.M. Monotonicity and the theory of relation-changes in LFG. In *Linguistic Inquiry*, volume 20. 1989.
- Bresnan J. eta Kaplan R.M. *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Massachussetts, 1982.
- Broadwell G.A. Directionals as complex predicates in Choctaw. In *On-line Proceedings of the LFG98 Conference*, 1998.
- Buitelaar P. *Systematic Polysemy and Underspecification*. PhD thesis, Brandeis University, 1998.
- Butt M. *The Structure of Complex Predicates in Urdu*. PhD thesis, Standford University, 1996.
- Cahill A., McCarthy M., Genabith J., eta Way A. Parsing with PCFGs and automatic f-structure annotation. 2002.
- Carroll J., Rigau G., Magnini B., Agirre E., Rodríguez H., eta Atserias J. Meaning: cycle 1: Acquisition. Technical report, 2003.
- Chomsky N. *Aspects of the Theory of Syntax*. MIT Press, Cambridge; Massachussetts, 1965.
- Chomsky N. *Lectures on Governmet and Binding. The Pisa Lectures*. Mounton de Gruyter, Berlin; New York, 1987.
- Copestake A. eta Flickinger D. An open source grammar development environment and broad-coverage English grammar using HPSG. In *International Conference on Language Resources and Evaluation (LREC)*, Atenas, 2000.
- Cuyppers I., Sánchez A., Schippers L., Adriaens G., Louw M., eta Forest P. Test specifications for EuroWordNet: internal data quiality and application in multilingual information retrieval. Technical report, University of Amsterdam, 1997.
- Dalrymple M. *Lexical Functional Grammar*, volume 34 of *Syntax and Semantics*. Academic Press, London, England, 2001.
- Demonte V. *Detrás de la palabra. Estudios de gramática del español*. Alianza Editorial, Madrid, Spain, 1991.
- Demonte V. *Teoría sintáctica: de las estructuras a la rección*. Colección Lingüística. Síntesis, 1995.
- Dorr B. Machine translation. A view from the lexicon. *Computational Linguistics*, 20(4), 1993.
- Dorr B. Large-scale acquisition of LCS-based lexicons for foreign language tutoring. Washington, DC, 1997.
- Dowty D. *Word Meaning and Montague Grammar*. Reidel, Dordrecht, 1979.
- Dowty D. Thematic proto-roles and argument selection. *Language*, 67(3):547–619, 1991.

- EAGLES. *Recommendations on Subcategorization*. <http://www.ilc.pi.cnr.it>, 1996.
- EAGLES. *Preliminary Recommendations on Lexical Semantic Encoding*. <http://www.ilc.pi.cnr.it>, 1999.
- Falk Y.N. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax*. CSLI Publications, Stanford University, 2001.
- Fellbaum C. *WordNet. An Electronic Lexical Database*. MIT Press, Cambridge, Massachusetts, 1998.
- Fellbaum C., Palmer M., Dang H.T., Delfs L., eta Wolf S. Manual and automatic semantic annotation with WordNet. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources*, Pittsburgh, 2001.
- Fernández A., Saint-Dizier P., Vázquez G., Kamel M., eta Benamara F. The Volem Project: a framework for the construction of advanced multilingual lexicons. In *Proceedings of Language Engineering Conference (LEC'02)*, Hyderabad, India, 2002.
- Fillmore C.J. Frames and the semantics of understanding. In *Quaderni di Semantica*, volume 6.2. 1985.
- Fillmore C.J. eta Baker C.F. FrameNet: Frame Semantics meets the corpus. In *Proceedings of WordNet and Other Lexical Resources Workshop*, Pittsburgh, 2001a.
- Fillmore C.J. eta Baker C.F. Frame Semantics for text understanding. In *Proceedings on WordNet and other Lexical Resources Workshop*, Pittsburgh, 2001b.
- Fillmore C.J., Baker C.F., eta Sato H. Seeing arguments through transparent structures. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, Las Palmas, Spain, 2002.
- Gazdar G., Klein E., Pullum G., eta Sag I. *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge, Massachusetts, 1985.
- Gilarranz J., Gonzalo J., eta Verdejo F. An approach to conceptual text retrieval using the EuroWordNet multilingual semantic database. In *Proceedings of AAAI-96 Spring Symposium Cross-Language Text and Speech retrieval*, 1996.
- Gildea D. eta Jurafsky D. Automatic labeling of semantic roles. *Computational Linguistics*, 2002.
- Gojenola K. *Guneak zuzendutako egitura sintagmatikoen gramatika (HPSG) eta Euskarako aplikazioa*, 1998.
- Gómez F. Linking WordNet verb classes to semantic interpretation. In Harabagiu S., editor, *Use of WordNet in Natural Language Processing Systems: Proceedings of the Conference*, pages 58–64. Association for Computational Linguistics, Somerset, New Jersey, 1998.
- Hale K.L. eta Keyser S.J. A view from the middle. Technical report, Center of Cognitive Science, Cambridge, Massachusetts, 1987. Lexicon Project Working Papers 10.

- Hale K.L. et al. Keyser S.J. On argument structure and the lexical expression of syntactic relations. In Hale K.L. et al. Keyser S.J., editors, *The View From Building*, volume 20, pages 53–110. MIT Press, Cambridge, Massachusetts, 1993.
- Harabagiu S.M. et al. Moldovan D.I. An intelligent system for question answering. In *Proceedings of the 5th Conference on Intelligent Systems*, Reno, 1996.
- Jackendoff R.S. *Semantic Structure*. MIT Press, Cambridge, Massachusetts, 1990.
- Jackendoff R.S. *Brain, Meaning, Grammar, Evolution*. Oxford University Press, 2000.
- Johnson C.R. et al. Fillmore C.J. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000)*, Seattle WA, 2000.
- King T.H., Crouch R., Riezler S., Dalrymple M., et al. Kaplan R.M. The PARC 700 Dependency Bank. In *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora, held at the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)*, Budapest, 2003.
- Kipper K., Dang H.T., et al. Palmer M. Class-based construction of a verb lexicon. In *AAAI/IAAI*, pages 691–696, 2000.
- Knight K. Building a large ontology for machine translation. In *Proceedings of the ARPA Human Language Technology Workshop*, Princeton, 1993.
- Levin B. *English Verb Classes and Alternations. A Preliminary Investigation*. The University of Chicago Press, Chicago & London, 1993.
- Lyons J. *Semantics*. Cambridge University Press, 1977.
- Magnini B. et al. Strapparava C. Using WordNet to improve user modelling in a web document recommender system. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources*, Pittsburgh, 2001.
- Mandala R., Takenobu T., et al. Hozumi T. The use of WordNet in information retrieval. In *Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*, Montreal, 1998.
- Mann G. Building proper noun ontologies for question answering. In *Proceedings of the Coling 2002 Workshop "SemaNet'02: Building and Using Semantic Networks"*, 2002.
- Matwin S., Szpakowicz S., et al. Li X. A WordNet-based algorithm for word sense disambiguation. 1995.
- Milhacea R. et al. Moldovan D.I. Word semantics for information retrieval: moving one step closer to the semantic web. In *International Conference on Tools in Artificial Intelligence*, 2001.
- Miller G.A. WordNet: a dictionary browser. In *Proceedings of the First International Conference on Information in Data*, Waterloo, 1985.

- Miller G.A., Chodorow M., Landes S., Leacock C., eta Thomas R.G. Using a semantic concordance for sense identification. In *Proceedings of the ARPA Human Language Technology Workshop*, San Francisco, 1994.
- Minnen G. Selective magic HPSG parsing, 1999.
- Montague R. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven, 1974.
- Moon Y.J. eta Kim Y.T. Concept-based verb translation in the Korean-English machine translation system. *Journal of the Korea Information Science Society*, 1995.
- Nishida K., Torisawa K., eta Tsujii J. Efficient HPSG parsing algorithm with array unification, 1999.
- Oepen S., Flickinger D., Toutanova K., eta Manning C.D. A rich and dynamic Treebank for HPSG. Sozopol, Bulgaria, 2002.
- Ohara K.H., Fujii S., Saito H., Ishizaki S., Ohori T., eta Suzuki R. The Japanese FrameNet project: A preliminary report. In *Proceedings of Pacific Association for Computational Linguistics (PACLING03)*,.
- Osenova P. eta Simov K. The Bulgarian HPSG Treebank: Specialization of the annotation scheme. In *Proc. of The Second Workshop on Treebanks and Linguistic Theories (TLT2003)*, Växjö, Sweden, 2003.
- Palmer M. eta Kingsbury P. From TreeBank to PropBank. 2003.
- Pasca M. eta Harabagiu S.M. The informative role of WordNet in open-domain question answering. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources*, Pittsburgh, 2001.
- Pociello E. Aditzen hautapen-murriztapenak: kirol domeinura mugatutako ingeleseko hautapen-murriztapenak eta euren baliagarritasuna euskararako. Hastapeneko lana. Master's thesis, UPV-EHU, 2004.
- Pollard C. eta Sag I. *Information-based Syntax and Semantics*, volume 1. CSLI, 1987.
- Pollard C. eta Sag I. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, 1994.
- Popowich F. eta Vogel C. Chart parsing Head-Driven Phrase Structure Grammar. Technical Report 90-1, 1990.
- Poznanski V. eta Sanfilippo A. Detecting dependencies between semantic verb subclasses and subcategorization frame in text corpora. In *Proceedings of the ACL-SIGLEX WSHP on Extracting Lexical Knowledge from Text*, 1993.
- Pustejovsky J. The generative lexicon. *Computational Linguistics*, 17(4), 1991.
- Pustejovsky J. *The Generative Lexicon*. MIT Press, Cambridge, Massachusetts, 1995.
- Pustejovsky J., editor. *Semantics and the Lexicon*. Kluwer Academic Publishers, 1993.

- Rigau G., Agirre E., eta Atserias J. The MEANING project. In *Proceedings of the "XIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*, Alcalá de Henares (Madrid), 2003.
- Rigau G., Rodríguez H., eta Turmo J. Automatically extracting translation links using a wide coverage semantic taxonomy. In *Proceedings of the 15th International Conference in Language Engineering, IA-95*, Montpellier, France, 1995.
- Ruiz J.C., Zubizarreta J., eta Abaitua J. Un compilador de LFG y su aplicación al euskara. In *SEPLN*, Donostia, 1991.
- Saint-Dizier P. Constructing verb semantic classes for French: Methods and evaluation. In *Proceedings of the Coling*, 1996.
- Sells P. *Teorías sintácticas actuales (GB, GPSG, LFG)*. Teide, Barcelona, 1989.
- Siddharthan A. Resolving relative clause attachment ambiguities using machine learning techniques and WordNet hierarchies. In *Proceedings of 5th National Colloquium for Computational Linguistics in the UK (CLUK 2002)*, 2002.
- Subirats-Rüggeberg C. eta Petruck M.R.L. Surprise: Spanish FrameNet! In *Workshop on Frame Semantics, International Congress of Linguists*, Prague, Czech Republic, 2003.
- Talmy L. Lexicalization patterns: Semantic structure in lexical forms. In *Language Typology and Syntactic Description*, volume 3. Cambridge University Press, 1985.
- Vendler Z. *Linguistics in Philosophy*. Cornell University Press, Ithaca, New York, 1967.
- Verkuyl H. *On the Compositional Nature of the Aspects*. Reidel, Dordrecht, 1972.
- Vossen P., editor. *EuroWordNet: A multilingual database with lexical semantic networks*. Kluwer Academic Publishers, 1998.
- Vossen P. EuroWordNet: a multilingual database for information retrieval. In *Proceedings of the DELOS workshop on Cross-language Information Retrieval*, Zurich, 1997.
- Vázquez G., Fernández A., eta Martí M.A. *Clasificación Verbal. Alternancias de diátesis*. Number 3 in *Quaderns de Sintagma*. Edicions de la Universitat de Lleida, 2000.
- Way A. Translating with examples: The LFG-DOT models of translation. In *Recent Advances in Example-Based Machine Translation*,. Kluwer Academic Publishers, 2003.
- Zaenen A. *Semantics and the Lexicon*. Kluwer Academic Publisher, Dordrecht, 1993.