# GRAMÁTICA DE ESTRUCTURAS SINTAGMÁTICAS DIRIGIDAS POR EL NÚCLEO (HPSG) Y SU APLICACIÓN AL EUSKARA

## Resumen

En este trabajo se examina el formalismo Head-Driven Phrase Structure Grammar (HPSG), diseñado para describir la sintaxis y semántica del lenguaje natural, de cara a comprobar su adecuación al tratamiento del euskara. El formalismo HPSG se encuentra en auge, destacando entre los formalismos basados en unificación, siendo los principales motivos su declaratividad y coherencia (descripción de todo el formalismo en un solo nivel).

El trabajo se estructura en tres partes. En la primera se hace una descripción general del formalismo, utilizando para ello [PS, 87] como principal referencia. En la segunda parte se presentan una implementación de HPSG y la descripción de un subconjunto del euskara. Finalmente se presentan los principales problemas encontrados, sus posibles soluciones, y aspectos a tratar en una continuación del trabajo.

# GUNEAK ZUZENDUTAKO EGITURA SINTAGMATIKOEN GRAMATIKA (HPSG) ETA EUSKARARAKO APLIKAZIOA

## Laburpena

Lan honetan HPSG lengoaia naturalen sintaxia eta semantikarako formalismoaren azterketa egin nahi izan da, euskararen sintaxiaren deskribapenerako egokitasuna aztertzeko. HPSG formalismoa "modan" dauden unifikazioan oinarritutatko formalismoen artean nagusi da momentu honetan, bere deklaratibotasuna eta koherentzia (gramatika osoa maila bakar batean deskribatu ahal izatea) arrakasta horren zergati nagusiak izan daitezkeelarik, eta hau izan da euskararekin aproba bat egiteko saio honen arrazoia.

Lana hiru zati nagusitan banatu da. Lehenengoan formalismoaren aurkezpen orokorra egingo da, eta honetarako [PS 87] formalismoaren erreferentzi-liburua erabili da. Bigarren zatian HPSGren inplementazio bat eta hau erabilita euskararen deskribapenerako azpimultzo baten garapena deskribatuko dira. Azkenik, aurkitutako zailtasunak, soluzio posibleak eta jarraitzeko bideak proposatuko dira.

# HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR (HPSG) AND ITS APPLICATION TO BASQUE


## Abstract

This work examines the HPSG formalism, designed for the description of the syntax and semantics of natural languages, in order to study its adequacy for the treatment of Basque. The HPSG formalism is now in fashion, mainly among the unification-based formalisms, the main reasons being its declarativeness and coherence (all the formalism described at just one level).

This work is structured in three parts. Firstly a general introduction to the formalism, based mainly in [PS, 87]. Next, an implementation of HPSG is presented and used for the specification of a subset of Basque. Finally, the main problems encountered are discussed and some solutions proposed, together with future directions to work on.

# GUNEAK ZUZENDUTAKO EGITURA SINTAGMATIKOEN GRAMATIKA (HPSG) ETA EUSKARARAKO APLIKAZIOA

# 1. Sarrera.

Lan honetan HPSG lengoaia naturalen sintaxia eta semantikarako formalismoaren azterketa egin nahi izan da, euskararen sintaxiaren deskribapenerako egokitasuna aztertzeko. HPSG formalismoa "modan" dauden unifikazioan oinarritutako formalismoen artean nagusi da momentu honetan, bere deklaratibotasuna eta koherentzia (gramatika osoa maila bakar batean deskribatu ahal izatea) arrakasta horren zergati nagusiak izan daitezkeelarik, eta hau izan da euskararekin aproba bat egiteko saio honen arrazoia.

Lana hiru zati nagusitan banatu da. Lehenengoan formalismoaren aurkezpen orokorra egingo da, eta honetarako [PS, 87] formalismoaren erreferentzi-liburua erabili da. Bigarren zatian HPSGren inplementazio bat eta hau erabilita euskararen deskribapenerako azpimultzo baten garapena deskribatuko dira. Azkenik, aurkitutako zailtasunak, soluzio posibleak eta jarraitzeko bideak proposatuko dira.

# 2. HPSG formalismoaren deskribapena.

## 2.1. Sarrera.

Guneak zuzendutako egitura sintagmatikoen gramatika (HPSG) lengoaien <u>sintaxia eta semantika deskribatzeko formalismoa</u> da, zeinek linguistika, informatika eta filosofiako arlo desberdinetako ikuspuntuak hartu dituen: teoria sintaktikoak (Categorial Grammar, Lexical-Functional Grammar, Government-Binding, Generalized Phrase Structure Grammar), teoria semantikoak (Situation Semantics), berbaldiaren teoria (Discourse Representation Theory), ezagutzaren adierazpena eta unifikazioan oinarritutako formalismoak.

Hainbeste lekutatik hartutako ideien bilduma honen ezaugarririk aipagarrienak hauek izan daitezke:

- <u>Unifikazioan oinarrituta</u>. HPSG gramatikan erabiltzen den elementurik garrantzitsuena zeinu linguistikoa da. Zeinua informazio partziala duen ezaugarri-egitura da. Teoria honetan elementu lexikalak, erregela gramatikalak, gramatikaren printzipio orokorrak eta semantika, denak egitura partzial horien bidez deskribatzen dira, honela uniformitate handia mantenduz teoriaren aspektu desberdinen deskribapenean, unifikazioaren abantailekin batera: monotonikoa, deklaratiboa, alderantzgarritasuna (produkzioa eta interpretazioa deskribatzeko), oinarri matematiko sendoa edukitzearen laguntza, eta, azkenik, formalismo desberdinak konparatzeko tresna bezala erabili ahal izatea, azalezko desberdintasunak eta oinarri formal gutxiko deskribapenak agerian uzteko.

- Semantika gramatikaren oinarrizko elementua da, teoria honetan fonologia, sintaxia, semantika eta berbaldia integratu nahi bait dira, lengoaia osoa deskribatzeko oinarria emateko. Teorian lengoaia esaldien eta esaldiek deskribatzen dituzten elementuen arteko murrizpen-multzoa da. Egoeren semantika erabiliko da semantika analizatzeko.

- Beste teoria sintaktikoetan egiten den bezala, teoria honetan printzipio orokorrak edo unibertsalak definitzen dira, eta hauekin batera lengoaia konkretu baten printzipio lokalak. Printzipio orokor hauen bidez maila altuko generalizazioak egiten dira, lengoaiaren zeinu guztiek bete behar dituztenak.

- Teoria sintaktiko modernoek hartu duten bidea jarraituz, lexikalizazio-maila oso altua du teoria honek, gramatikaren informazio gehiena hiztegian gordetzen delarik. Honen ondorioz, erregela sintaktikoak oso eskematikoak izango dira eta hiztegia, berriz, oso konplexua. Honek duen informazioa modu erabilgarrian izateko, ezagutzaren adierazpenaren eta datu-moten teoriaren munduko ideiak hartu dira, osagai lexikalen hierarkia definituz, bertan heredentzia eta erregela lexikalak aplikatu egin daitezkeelarik, erredundantzia gutxitzeko.

## 2.2. Ezaugarri-egiturak. Kategoria eta ezaugarri sintaktikoak.

### 2.2.1. Ezaugarri-egiturak.

Teoria honetan zeinu linguistikoa da gramatika osoaren deskribapenean erabiliko den oinarrizko elementua. Honen definizioa hau da: informazio partziala duen egitura. Informazio partzialeko egiturak ezaugarri-egituren bidez deskribatzen dira. Ezaugarri-egitura bat ezaugarri-balio bikoteez osatutako egitura da.

Ezaugarri baten balioa atomikoa edo beste ezaugarri-egitura izan daiteke. Adibidez:

| | | | |
|---|---|---|---|
| IZENA | JOSE | | |
| ADINA | 11 | | |
| HELBIDEA | | KALEA | XXX |
| | | ZENB | 21 |

Egitura hauen bidez lengoaiaren aspektu guztiak deskribatuko dira: fonologia, sintaxia, semantika, erregela gramatikalak eta printzipio orokorrak.

Bi ezaugarri-egitura A eta B emanda, Ak Bk baino informazio gutxiago badu, orduan Ak B subsumitzen duela esango dugu. Honela ordena partziala definitzen da ezaugarri-egituren artean, punta batean egitura hutsa ([]) edo informazio gutxien duena dagoelarik, eta bestean informazio gehiegi duena, beste inork subsumitzen ez duen egitura. Honela algebra matematikoa definitzen da, lengoaia osoa deskribatzeko oinarri matematiko sendoa emanez.

Ezaugarri-egiturekin egingo den eragiketarik garrantzitsuena <u>unifikazioa</u> da: A eta B ezaugarri-egiturak emanda, euren unifikazioa bi egituren informazio guztia eta ez gehiago duen egitura da. Hau da, bi egiturak subsumitzen dituen egiturarik txikiena (informazio gutxien daukana) da. Unifikazioak bi elementuen informazioa biltzen du, bateragarria den kasuetan.

HPSG teoriak lengoaiaren erabileran (sorkuntza, analisia, itzulpena, ...) zeinuen esanahia modu monotonikoan biltzen dela suposatzen du; sarrera lexikalen, gramatika-erregelen eta printzipio orokorren informazioa bilduz. Unifikazioa izango da honetarako erabiliko den eragiketarik erabiliena.

Adibidez (zenbakiak erabiltzen dira, indize moduan, zenbait ezaugarrik balio bera dutela adierazteko):

| | | | | |
|---|---|---|---|---|
| a) | SUBJ (1) | | | |
| | PRED (1) | AGR | GEN | FEM |
| b) | PRED | AGR | PER | 3 |
| | | | NUM | SNG |

a) eta b) egituren unifikazioa hau da:

c) | SUBJ (1) | PRED (1) AGR | PER 3 | NUM SNG | GEN FEM |

| | | | |
|---|---|---|---|
| c) | SUBJ (1) | | |
| | PRED (1) | AGR | PER 3 |
| | | | NUM SNG |
| | | | GEN FEM |

Ezaugarri-egiturek **mota** edukiko dute, mota honek egitura baten balioak nolakoak diren emango duelarik. Honela ezaugarrien balio posibleak murriztuko dira. Adibidez, zeinu lexikalek sintaxia, fonologia eta semantika deskribatzeko ezaugarriak izango dituzte, eta ez alabei buruzko informazioa, hau zeinu egituratuek (lexikalak ez direnak) izango dutelako:

- zeinu lexikalak:

| | | |
|---|---|---|
| | PHON | ... |
| | SYN | ... |
| sign | SEM | ... |

- zeinu egituratuak:

| | | |
|---|---|---|
| | PHON | ... |
| | SYN | ... |
| | SEM | ... |
| phrasal-sign | DAUGHTERS | ... |

Badaude, unifikazioaz aparte, beste eragiketa erabilgarriak ezaugarri-egiturekin:

- Disjunzioa.

| | | |
|---|---|---|
| CASE | NOM | ACC |
| AGR | GEN FEM<br>NUM SING | NUM PLU |

- Inplikazioa. Printzipio orokorrak deskribatzeko.

- Ukapena. Ezaugarri-egitura batek balio bat ez edukitzea deskribatzeko. Honen esanahia ez dago argi egitura konplexuetarako, eta balio atomikoa duten ezaugarriekin bakarrik erabiliko da:

$$\left[\ \mathrm{CASE} \quad \neg\mathrm{NOM}\ \right]$$

Ezaugarri baten balioa batzuetan ez da ezaugarri-egitura normala izango, eta komenigarriak izango dira, egingo diren analisietarako, aukera hauek:

- Listak: Ezaugarri baten balioa ezaugarri-egituren lista bat da. Honek azpikategorizazioa, egitura sintaktikoa eta urruneko menpekotasunak tratatzeko balioko du.

- Multzoak. Hauek semantikan erabiliko dira, objektu erreferentzialak (izen-sintagmak deskribatzen dituztenak) gordetzeko.

- Dependentzia funtzionalak: hauen bidez ezaugarri baten balioa beste balio batzuekin erlazionatuko da. Hau kasu desberdinetan izango da erabilgarria: esaldien osagaien ordena, azpikategorizazioa eta semantika tratatzeko, ...

### 2.2.2. Ezaugarri eta kategoria sintaktikoak.

HPSG teorian sintaxia hitzen konbinazioak adierazteko informazioa da. Hau sarrera lexikaletan, erregela gramatikaletan eta printzipio orokorretan adieraziko da. Puntu honetan zeinu sintaktikoen ezaugarriak deskribatuko dira. Hauetatik gehienak GPSG teorietatik hartu dira, nahiz eta aldaketak egon, GPSG teorian sintaxiaren pisurik handiena erregelek hartzen zutelako, eta HPSG teorian gehiena printzipio orokorren eta lexikoiaren bidez adierazten delako.

Bi mota bereizten dira ezaugarri sintaktikoen artean:

- <u>Ezaugarri lokalak</u>, zeinu baten berezko propietate sintaktikoak deskribatzeko. Lexikoian oinarrituta daude:

    * <u>Gune-ezaugarriak</u> (HEAD): sintagma bakoitzak badu hitz berezi bat, bere propietate sintaktikoak

ematen dizkiona (gunea). Ezaugarri hauek, gune horretatik dagokion sintagmara teoriaren printzipio baten arabera (HFP) "mugituko" dira. Printzipioak egitura sintagmatiko batean amaren eta gune-alabaren gune-ezaugarriak berdinak direla dio:

phrasal_sign| | ⇒

| SYN|LOC|HEAD (1) |
| DTRS|HEAD-DTR|SYN|LOC|HEAD (1) |

Printzipio honek inplikazioa erabiltzen du, eta bere esanahia honela eman dezakegu hitzez: zeinu egituratu batean (*phrasal-sign* motakoa) amaren eta gune-alabaren (DTRS|HEAD-DTR) gune-ezaugarriak berdinak izan behar dute. Horrela ezaugarriak gune lexikaletik proiekzio guztietara "propagatuko" dira.

Ingleserako proposatutako gune-ezaugarriak hauek dira (gehienak GPSGtik hartuta), eta denek balio atomikoak hartzen dituzte:

- MAJ: Hitz- edo sintagma-mota. Sarrera lexikalek eta beren proiekzioek izango dute. Balioak: N, V, A, P, ADV, D.

- CASE: kasua. Balioak: NOM, ACC, GEN.

- VFORM: aditzaren forma. Balioak: FIN, BSE, PRP, PAS, PSP, INF, GER.

- NFORM: izenak eta beraien proiekzioak dute ezaugarri hau, eta honek izen-sintagma arruntak "it" eta "there" formetatik bereizteko balio du. Balioak: NORM, IT, THERE.

- PFORM: sintagma preposizionalak (aditzen azpikategorizazioan gehienbat) bereizteko. Balioak: ON, TO, BY,...

- AUX: aditz erroa eta laguntzailea bereizteko ezaugarri boolearra.

- INV: galderazko esaldietako aditz laguntzailea bereizteko ezaugarri boolearra.

- PRD: sintagma predikatiboak tratatzeko. Ezaugarri boolearra.

Adibidez, ikus dezagun gune-ezaugarrien portaera esaldi batean (adibidean erregela gramatikalak aplikatzen dira, hauek geroago tratatuko dira). Esaldi baten gunea aditza izango da:

* Azpikategorizazioa (SUBCAT): zeinu batek beste zeinuekin konbinatzeko gaitasuna deskribatzeko. Bere balioa ezaugarri-egituren lista bat izango da, eta honek zenbat eta zein motatako elementuekin konbinatuko den adieraziko du: aditza-osagarriak, izena-determinantea, ...

* Ezaugarri-egitura bat lexikala den ala ez adierazteko ezaugarri boolearra (LEX). Honek ondorioak izango ditu azpikategorizazioa eta osagaien ordena tratatzeko momentuan. Gutxi

gorabehera X-bar teoriako mailen papera egiten du.

- Beste motetako dependentziak tratatzeko ezaugarriak (loturazko ezaugarriak); ez dira lokalak (urruneko menpekotasunak sortzen dituzte) eta ez dira gune-alabetatik sortzen. Ezaugarri hauen portaera BIP (Binding Inheritance Principle, GPSGko FFP printzipioaren antzekoa) printzipioaren bidez tratatzen da. Lan honetan ez dira ezaugarri hauek aipatuko, eta analisiak menpekotasunik sortzen ez duten esaldiekin egingo dira. Hiru dira honetarako erabiliko diren ezaugarriak:

    * SLASH: hutsuneak tratatzeko (topikalizazioa). Bere balioa ezaugarri-egituren multzoa da, zeinek lotu gabe dauden elementuak gordeko dituen. Hutsunea eta hutsunea betetzen duen elementuaren arteko lotura ziurtatzeko erabiltzen da.

    * REL: izenordain erlatiboak eta euren erreferentzia. Balioa indize semantikoa izango da.

    * QUE: galderak tratatzeko. Balioa kuantifikadore logikoen multzoa da.

Guzti honen ondorioz, zeinu baten egitura orokorra honakoa izango da:

$$\begin{bmatrix} \text{PHON} \ldots & \\ \text{SYN} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MAJ} \ldots \\ \text{VFORM} \ldots \\ . \\ . \end{bmatrix} \\ \text{SUBCAT} \langle \ldots \rangle & \\ \text{LEX} \ldots & \end{bmatrix} \\ \text{BIND} & \begin{bmatrix} \text{SLASH} \{ \ldots \} \\ \text{REL} \{ \ldots \} \\ \text{QUE} \{ \ldots \} \end{bmatrix} \end{bmatrix} \\ \text{SEM} \ldots & \end{bmatrix}$$

Zeinu sintaktiko egituratuetan, halere, badaude ama eta alaben arteko dependentziak adierazteko ezaugarriak. Adibidez, gunea-ezaugarriak egituretan hiru ezaugarri daude:

- HEAD-DAUGHTER (gune-alaba). Honek sintagma baten gunea zein den gordetzen du. Bere balioa zeinua da.
- COMP-DAUGHTERS (osagarrien zerrenda). Gune baten osagarriak gordeko ditu. Bere balioa zeinuen lista izango da.
- ADJ-DAUGHTERS (adjuntuen zerrenda). Osagarriak ez diren elementuak gordetzeko zeinuen lista. Adjuntuak azpikategorizatzen ez diren elementuak dira, ez dutelako lotura zuzena gune batekin eta nahi beste aldiz errepika daitezkeelako. Adibidez: adjetiboak, sintagma preposizionalak inglesez, aditzlagun eta izenlagunak euskaraz, ...

## 2.3. Semantika.

Esanahia, egoeren semantikaren (*situation semantics*) arabera, egoera-moten arteko erlazioa da.

Esanahi linguistikoa, beraz, esaldi-mota baten eta esaldi horrek deskribatzen duen egoera-motaren arteko erlazioa izango da. Adibidez:

Jon lotan dago -> **hitz-katea**

Jon izeneko norbait lo egiten dago -> **egoera**

Beraz, lengoaia erabiltzeko ezagumendua honela banatzen da:

- Lexikoia
- Erregela gramatikalak
- Esaldiek deskribatzen duten **mundua**

Esaldiek aipatzen duten mundua deskribatzeko gizakiak munduko gauzak nola klasifikatzen dituen jakin behar da. Egoeren semantikaren arabera, gizakiak honelako elementuak bereizten ditu munduan:

- Banakoak: erregea, etxea, astoa, ... Munduko entitateak; somatu egiten ditugu eta elkarreragina dugu haiekin.

- Erlazioak: gertaera, non banakoak parte hartu dezaketen. Argumentu-kopurua eta argumentuetarako rolak dituzte. Izen arruntak, adjektiboak, preposizioak eta aditzek sortzen dituzte. Mota desberdinetakoak izan daitezke:

    * Propietateak: gorria izatea, katua izatea, ...

    * Gertaerak: maitatzea, haserre egotea, ...

    * Prozesuak: ibiltzea, irakurtzea, ...

- Egoerak: banakoak propietateekin eta erlazioetan denbora-espazioko baldintzetan.

- Zirkunstantziak: munduaren deskribapen partziala izan daitezkeen proposizioak (gertaera potentzialak). Munduaren deskribapen partziala emanda egiazkoa ala faltsua izan daiteke proposizio bat.

  Oinarrizko zirkunstanzia bat deskribatzeko elementu hauek erabiliko dira:

    * Erlazio bat.

    * Erlazio horretako rol guztiei objetuen asignazioa.

    * Polartasuna (positiboa edo negatiboa).

    Adibideak:

    << sleep, sleeper: Jon, 0 >>

    << believe, believer: Jon, belief:
    << sleep, sleeper: Jon, 0 >>, 1 >>

Zeinu linguistiko baten balio semantikoa ezaugarri-egituren bidez emango da, eta hauek egoerak eta beraien banakoak, propietate eta erlazioak deskribatuko dituzte, egoera horretan betetzen diren zirkunstantzien bidez.

Horrela lehen mailako logikaren antzeko sistema deskribatuko da. Bien arteko diferentziarik argiena egoeren logikak erlazio errealistak deskribatzen dituela da, eta ez matematikakoak.

Oinarrizko Zirkunstantzia baten adibidea hau izan daiteke:

| | |
|---|---|
| RELN | LIKE |
| LIKER | JON |
| LIKED | MARIA |
| POL | ONE |

Zirkunstantzia hauek zeinu linguistikoen ezaugarri-egituretan agertuko dira, zeinua eta deskribatzen duen egoera erlazionatzeko:

| | |
|---|---|
| PHON | likes |
| SYN\|LOC\|SUBCAT< | (3) NP [ CONT (1) ],<br>(4) NP [ CONT (2)] > |
| SEM\|CONT | RELN LIKE<br>LIKER (2)<br>LIKED (1) |

Adibide horretan aditz bat erlazio batekin lotzen da, eta horrela aditz horren semantika ematen da. Hala ere, azpikategorizazioaren bidez, erlazio semantikoaren (SEMANTIC|CONTENT ezaugarriaren balioa) rolak ere asignatzen dira (azpikategorizazioaren puntuan ikusiko denez, aditz horren subjektu eta objektuarekin), indizeen bidez.

Semantikaren printzipio unibertsalak, gutxi gorabehera, ama eta gune-alaben balio semantikoak berdinak direla dio:

headed-structure | DTRS | =>

| |
|---|
| SEM\|CONT (1) |
| DTRS\|HEAD-DTR\|SEM\|CONT(1) |

Printzipio bakar horren bidez zeinuen balio semantikoa emateko mekanismoa definitu da, eta ez da behar, GPSG teorian bezala, erregela bakoitzeko ekuazio semantikoen definizioa. Sinplifikazio honen ondorioa zeinu lexikalen balio semantikoaren konplexutasuna izango da.

Adibidez (SUBCAT ezaugarria sintaktikoa denez, honen aurretik ez da jarri ezaugarrien bidea, modu inplizituan SYN|LOC|SUBCAT ulertzen baita; berdin egiten da semantikarekin: CONT izena SEMANTIC|CONTENT bidearen laburpena da):

(4)

## 2.4. Azpikategorizazioa.

Hau SUBCAT ezaugarriaren bidez egiten da, eta gune lexikal bat eta bere osagarrien arteko dependentziak tratatzeko erabiltzen da. Ezaugarri honen balioa ezaugarri-egituren lista da, hau da, ordena kontutan hartzen da. Dependentziak sintaktikoak eta semantikoak dira:

- Zenbat eta nolako elementuak behar duen guneak konbinatzeko.

- Komunztadura.

- Rol semantikoen asignazioa: honek sintagma baten osagarrien eta sintagmak deskribatzen duen egoeraren elementuen arteko konexioa ematen du.

- Kasuaren asignazioa.

SUBCAT ezaugarriaren balioak printzipio unibertsal baten bidez tratatzen dira. Printzipioak ama-alabak motako egitura batean amaren azpikategorizazioko balioak gune-alabenak ken osagarri-alabenak izango direla dio.

Azpikategorizazioaren Printzipioa:

headed-structure | DTRS | =>

| SYN|LOC|SUBCAT (2)
| DTRS | HEAD-DTR|SYN|LOC|SUBCAT append((1), (2))
|      | COMP-DTRS (1)

Adibidez, "saw" zeinu lexikala emanda, ikus dezagun nolako elementuekin konbinatzen den:

| PHON saw
| SYN|LOC|SUBCAT< NP [ ACC (1)], NP [NOM (2)] >
| SEM|CONT | RELN see
|          | SEER (2)
|          | SEEN (1)

Zeinu honek bi izen-sintagma azpikategorizatzen ditu: bata (objektuarena) kasu akusatiboan, eta bestea (subjektua) nominatiboan. Rol semantikoen asignazioa ere azpiindizeen bidez ematen da.

Mota askotako elementuak azpikategorizatu daitezke:

| | | | |
|---|---|---|---|
| PHON | forced | | |
| SYN\|LOC | HEAD | MAJ | V |
| | | VFROM | FIN |
| | SUBCAT< VP [ INF], NP, NP > | | |

Teoria honetan erlazio gramatikalak SUBCAT zerrendako elementuen ordenaren arabera definitzen dira. Honek generalizazio desberdinak izango ditu osagaien ordena, kontrol-teoria, izenordainen eta erreflexiboen lotura, eta erregela lexikalak tratatzeko. Adibidez, zerrendako azken elementua subjektua izango da, azkenaurrekoa objektua, ...

SUBCAT zerrendan informazio hau ematen da:

- Informazio funtzionala, zerrendako elementuen ordenaren arabera. Adibidez, azken elementua subjektua izango da.
- Informazio sintaktikoaren espezifikazioa, hau da, zehaztu daitezke osagarrien SYNTAX ezaugarriaren murrizpenak.
- Informazio semantikoa (SEMANTICS ezaugarriaren murrizpenak, rol semantikoen asignazioa, komunztadura, ...).

Ezaugarri honek betetzen du unibertsala kontsideratzen den beste propietate bat (Locality Principle): SUBCAT zerrendetan zeinu lexikoek informazio sintaktikoa eta semantikoa bakarrik erabil dezakete, eta ez alaben informazioa.

## 2.5. Erregela sintaktikoak.

Teoria honen arabera, lengoaia baten egitura honakoa da:

Euskara = $P_1$ & $P_2$ ... & $P_n$ & ... $P_{n+m}$ & ( $L_1$ |... |$L_p$ | $R_1$ | ... | $R_q$ ) non:

- $P_{1...}P_n$: Printzipio unibertsalak (HFP, Azpikategorizazioaren printzipioa, ...).
- $P_{n+1...n+m}$: Lengoaiaren murrizpen espezifikoak.
- $L_i$: Zeinu lexikalak.
- $R_i$: Gramatikaren erregelak.

Gramatika-erregelek osagaiak dituzten zeinu sintagmatikoen egitura deskribatuko dute. Puntu honetan gunedun egiturak aztertuko ditugu, eta hauen barruan gunea-osagarriak motako egiturak.

<u>Gramatika-erregelak eskematikoak dira</u>, zeinu lexikaletatik eta printzipio orokorretatik atera daitekeen informazio guztia erregeletatik kanpo uzten delako.

Zeinu sintagmatiko egituratuek DAUGHTERS izeneko ezaugarria dute, zeinek zeinu baten alaben informazioa duen. Gunea-osagaiak motako egituretan HEAD-DAUGHTER (gune-alaba) eta COMP-DAUGHTERS (osagarrien zerrenda) ezaugarriak egongo dira.

Erregela 1[1] : $\left|\text{SUBCAT}\langle\rangle\right| \rightarrow \text{H} \left|\text{LEX -}\right|, \text{C}$

erregela hori erregela sintaktikoen formatu klasikoan emanda dago baina HPSG formalismoan, lengoaiaren aspektu guztiekin egiten den bezala, formulazioa ezaugarri-egitura baten bidez ematen da:

$$\begin{vmatrix} \text{SYN|LOC|SUBCAT}\langle\,\rangle & \\ & \\ \text{DTRS} & \begin{vmatrix} \text{HEAD-DTR|SYN|LOC|LEX -} \\ \text{COMP-DTRS}\langle[\,]\rangle \end{vmatrix} \end{vmatrix}$$

Honek "S -> NP VP" eta "NP -> DET N" moduko erregela asko hartzen ditu (gunearen konbinazioa azkeneko elementuarekin).

Erregela hau oso eskematikoa da, eta bere esanahia hobeto ulertzen da azpikategorizazio eta gune-ezaugarrien printzipioekin unifikatzen badugu (baita semantikaren printzipioa ere, baina adibidean ez da agertzen sinplifikatzearren):

$$\begin{vmatrix} \text{SYN|LOC} & \begin{vmatrix} \text{HEAD}(1) \\ \text{SUBCAT}\langle\,\rangle \end{vmatrix} \\ & \\ \text{DTRS} & \begin{vmatrix} \text{HEAD-DTR|SYN|LOC} & \begin{vmatrix} \text{HEAD}(1) \\ \text{SUBCAT}\ \langle(2)\rangle \\ \text{LEX -} \end{vmatrix} \\ & \\ \text{COMP-DTRS}\ \langle(2)\rangle & \end{vmatrix} \end{vmatrix}$$

---

[1] <> zerrenda hutsa adierazteko erabiltzen da.

<[]> elementuko bateko zerrenda adierazteko erabiltzen da.

Erregela 2: | SUBCAT <[]> | -> H |INV -, LEX + |, C*

SYN|LOC|SUBCAT<[]>

DTRS|HEAD-DTR|SYN|LOC | LEX +
INV - |

Bigarren erregela honek "VP -> V NP", "VP -> V NP PP", eta antzeko erregela tradizionalen mordoa subsumitzen du (gunea osagarri guztiekin, azkena ezik, konbinatzen denean).

Hurrengo adibidean erregela gramatikalen eta printzipioen interakzioa ikus dezakegu. Hiru printzipio orokor eta bi erregelen bidez, inglesaren zati handia deskribatzen da:

Irudi honetan erregela gramatikalen aplikazioak R1 eta R2 izenekin markatzen dira, lehen eta bigarren erregela adierazteko. Halere, H hitzak gunea eta C hitzak osagarriak adierazten dute. Erregelan gune-ezaugarrien eta azpikategorizazioaren printzipioak ere aplikatzen dira.

Erregela 3: | SUBCAT <> |-> H | INV +, LEX + |, C*

SYN|LOC|SUBCAT<[]>

DTRS|HEAD-DTR|SYN|LOC | LEX +
INV + |

Hirugarren erregelak galderazko perpausak tratatuko ditu. Bertan INV ezaugarria (galderazko aditz laguntzaileak bereizteko erabiltzen dena) "+" balioarekin duen guneak osagarri guztiekin konbinatzea deskribatzen du.

Hiru erregela hauek eta printzipio orokorrak hartuta, inglesezko esaldi enuntziatiboak eta galderazkoak tratatu daitezke.

## 2.6. Osagaien ordena.

Teoria honetan osagaien ordena eta erregela gramatikalak independenteak dira, GPSG teorian egiten den bezala, eta printzipio orokor baten eta aurrekaritza linealeko murrizpenen bidez tratatuko da:

"Constituent Order Principle":

$$\textit{phrasal\_sign}\left[\;\right] \Rightarrow \left[\begin{array}{ll} \text{PHON} & \text{order-constituents (1)} \\ \text{DTRS (1)} & \end{array}\right]$$

"order-constituents" dependentzia funtzionalak ondoko erregelak (LP edo aurrekaritza linealeko murriztapenak) betetzen direla konprobatuko du:

$$\text{LP1:} \quad \text{HEAD}\left[\text{LEX} +\right] < \left[\;\right]$$

"<" erlazioak bi alaben arteko lehentasuna deskribatzen du, hau da, gune lexikala edozein elementuren aurretik doa (adibidez: aditza bere osagarriak, subjektua ezik, baino lehenago; preposizioa sintagma preposizionaletan).

$$\text{LP2:} \quad \text{COMPLEMENT}\left[\text{MAJ} \;\; \neg\text{V}\right] \ll \left[\text{LEX} -\right]$$

"<<": murrizpen hau SUBCAT listan ezkerreko elementua eskubikoa baino lehenago agertzen denean betetzen da: bi alaba SUBCAT zerrendan agertzen badira, orduan aditzen proiekzioak ez direnak lexikalak ez direnak baino lehenago joango dira. Ordena hau hartzeko badaude beste puntu batzuk: erregela lexikalen aplikazioa, kontrola, lotura eta komunztadura.

Adibidez: *I want you to go.*

*I gave the man the apple.*

Hauetan, osagarriak azpikategorizazioko listen arabera ordenatuko dira, eta beste esaldi hauek baztertzen dira:

*I want to go you .*

*I gave the apple the man .*

Bi erregela hauekin gunea eta osagarrien arteko murrizpen linealak tratatzen dira: gunea osagarri ahizpen aurretik doa, eta osagarriak SUBCAT listako ordenaren arabera ordenatzen dira (gutxi gorabehera, murrizpenak orokorragoak baitira).

LP3: $\left[ \text{MAJ} \quad \neg \text{N} \right] < \left[ \text{FOCUS} + \right]$

(berbaldiaren araberako murrizpena: FOCUS edo markatutako elementua izenen proiekzioak baino atzerago doa)

Ez dugu hemen esplikatuko murrizpen honen esanahia, baina honekin mota askotako murrizpenak adieraz daitezkeela ikusten dugu.

Hiru murrizpen hauekin mota desberdinetako fenomenoak tratatzen dira: pasibak, adjuntu adberbialak, osagarri sententzialak. Aurrekaritza linealeko murrizpenek informazio sintaktikoa, gramatikala eta berbaldiarena erabiltzen dituzte.

## 2.7. Lexikoiaren antolaketa.

HPSG teorian oso lexikalizazio-maila handia dago, lehen agertutako arrazoiengatik:

- Azpikategorizazioa printzipio orokor baten eta erregela eskematikoen bidez egiten da.
- Zeinuen edukin semantikoa modu esplizituan kodetzen da, eta Semantikaren Printzipioaren bidez tratatzen da.
- Erregela sintaktiko gutxi daude.

Honen ondorioa elementu lexikalak oso konplexuak izatea da, informazio asko gordetzen dutelako. Hauek egituratzeko mota lexikalen sistema taxonomikoa erabiliko da. Honela informazio lexikala hitz-mota batzuen hierarkien bidez emango da, beraietan hitz guztiak konpartitutako propietate sintaktiko, semantiko eta morfologikoak sailkatuta egongo direlarik. Adibidez:

PHON tried
SYN|LOC
HEAD
MAJ V
VFORM FIN
AUX -
INV -
PRD -
SUBCAT< SYN|LOC HEAD MAJ V
VFORM FIN
SUBCAT < (1) >
SEM|CONT (2)
,NP[NOM,NORM(1)]>
SEM CONT
RELN try
TRYER (2)
CIRCUMSTANCE (1)
LOCATION (3)
INDS{ VAR (3)
RESTRICTION RELN TEMPORALLY-PRECEDES
ARG1 (3)
ARG 1d }

Hitz bakoitzeko informazio hori adieraztea lan ikaragarria izan daiteke, baina kontutan hartzen badira jeneralizazio batzuk erredundantzia gutxitu daiteke:

- Hitz guztiek konpartitzen duten informazioa:

$$\begin{bmatrix} \text{SYN}|\text{LOC}|\text{LEX} & + \end{bmatrix}$$

- Aditz guztiek konpartitzen duten informazioa:

$$\begin{bmatrix} \text{SYN}|\text{LOC}|\text{HEAD}|\text{MAJ} & \text{V} \\ \text{SEM}|\text{CONT} & \textit{basic-circumstance} \end{bmatrix}$$

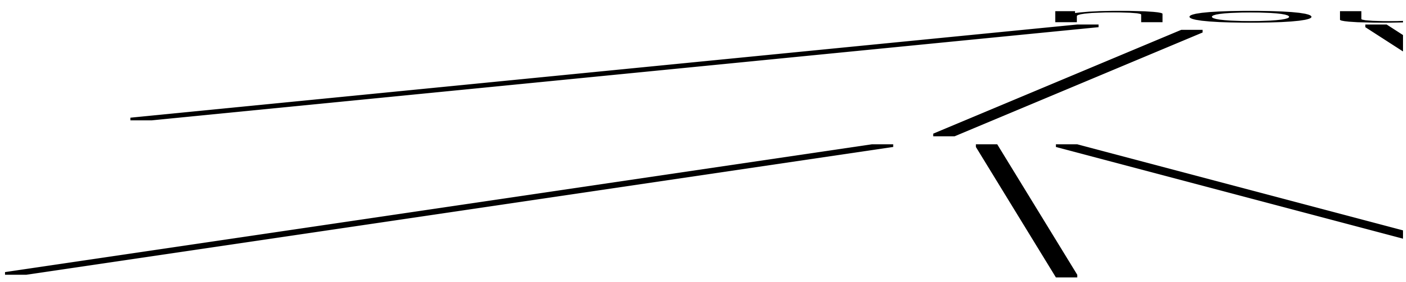
- Aditz jokatu guztiek konpartitzen duten informazioa:

$$\begin{bmatrix} \text{SYN}|\text{LOC} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{VFORM} & \text{FIN} \\ \text{PRD} & - \end{bmatrix} \\ \text{SUBCAT} \langle \ldots, \text{NP[NOM]} \rangle & \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{CONT}|\text{LOCATION} & (3) \\ \text{INDS}\ \{\ [\ \text{VAR}\ (3)\ ]\ \} & \end{bmatrix} \end{bmatrix}$$

- "tried" hitzaren informazio espezifikoa:

$$\begin{bmatrix} \text{PHON} & \text{tried} \\ \text{SYN}|\text{LOC}|\text{SUBCAT} & \langle\ [\ \text{SEM}|\text{CONT}\ (2)]\ ,\ [\ \ ]\ (1)\ \rangle \\ \text{SEM}|\text{CONT} & \begin{bmatrix} \text{RELN} & \text{try} \\ \text{TRYER} & (1) \\ \text{CIRCUMSTANCE} & (2) \end{bmatrix} \end{bmatrix}$$

Ezagumendu hau erabiliz zeinu lexikalen taxonomia deskribatuko da. Bertan elementu bakoitzak mota bat du, eta mota batzuek beste elementuen azpimotak izango dira, bere ezaugarri guztiak heredentziaren bidez hartuz. Azkenean, zeinu bakoitzeko bere ezugarri espezifikoak besterik ez dira zehaztu behar (irudian maiuskulaz agertzen diren izenak (HEAD, SUBCAT) ez dira motak, moten sailkapenerako irizpideak baizik):



Taxonomia honekin batera erregela lexikalak definitu daitezke: inflekzioa, deribazioa eta ereduak (pasiba, singularreko hirugarren pertsonaren erregela, ...) tratatzeko. Erregela hauek hierarkiaren motei aplikatzen zaizkie, eta forma berriak deskribatzeko balio dute.

Erregela lexikalak hitz-motei aplikatzen zaizkien funtzioak dira, emaitzatzat beste hitz-mota bateko elementua ematen dutenak.

Adibidez, pasibaren tratamendurako erregela:

$$ {}_{\text{base \& trans}}\left| \begin{array}{l} \text{PHON (1)} \\ \text{PAST-PART (2)} \\ \text{SYN|LOC|SUBCAT} < \ldots,\ (3),\ (4) > \\ \text{SEM|CONT (5)} \end{array} \right| \text{->} $$

$$ {}_{\text{passive}}\left| \begin{array}{l} \text{PHON } f_{PSP}\ \ ((1), (2)) \\ \text{SEM|LOC|SUBCAT} < (\text{PP[BY]}\ (4)), \ldots,\ (3) > \\ \text{SEM|CONT (5)} \end{array} \right| $$

Erregela horretan ondoko aspektuak deskribatu dira:

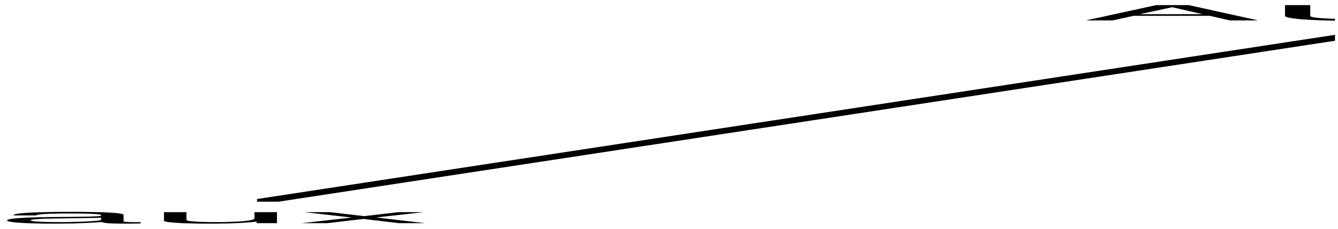
- Erregela "base" (aditz erroa) eta trantsitiboak diren aditzei aplikatzen zaizkie. Emaitza pasiba motakoa da. Honekin, lortutako egiturak, erregela lexikalaren emaitzean agertzen direnetaz aparte, pasiba hitz-motako ezaugarri guztiak (heredentziaren bidez) hartuko dituela esaten da.

- Zeinu berriaren fonologia funtzio batek ematen du. Funtzio honek bi parametro hartzen ditu: argumentuaren fonologia, forma normalak egiteko ("ed" atzizkia gehituz) eta PAST-PART ezaugarriaren balioa formak irregularrak direnerako.

- Azpikategorizazioan aldaketa ematen da: forma berrian argumentuan subjektuaren papera egiten zuen elementua preposizio-sintagma izango da.

- Rol semantikoen asignazioa berdina izango da (kontutan hartu behar da forma berrian preposizio-sintagmaren elementuak hasierako forman subjektuak betetzen zuen papera hartzen duela).

Erregela hauek aplikatuz erredundantzia asko kenduko da zeinu lexikalen taxonomian:



Adibidean bi erregela lexikal agertzen dira: bata pasiba eta bestea gerundioa lortzeko.

## 3. Euskararako gramatikaren adibidea.

### 3.1. HPSG formalismoaren inplementazioa.

HPSG formalismoa probatzeko HPSG-PL inplementazioa [KO 92] [PV 91] erabili da. Sistema honen ezaugarririk nagusienak hauek dira:

- Prolog lengoaiaz idatzita, guztira 2.000 bat lerro. Ez da sistema sendoa, baizik eta gramatikak garatzeko zenbait tresnen bilduma (lexikoiaren konpiladorea, parserra, eta sarrera-irteerako modulua, gutxi gora behera).

- HPSG formalismoan idatzitako gramatikentzako inplementazioa. Ez da, beraz, PATR-IIren antzeko sintaxiaren tratamendurako inplementazio orokorren parekoa, baizik eta HPSGren barruan esperimentatzeko tresna.

- Sistema ez da erabat deklaratiboa, kasu batzuetan gramatika eta inplementazioa lotuta daudelako. Kasu hauetan, gramatikaren zati bat aldatzeko programa aldatu behar da, honek eduki ditzakeen zailtasun eta arrisku guztiekin. Programarekin lotura duten HPSGko ezaugarriak hauexek dira: aurrekaritza linealeko murrizpenen tratamendua, urruneko menpekotasunak, semantikarekin lotura duten ezaugarriak, eta azpikategorizazioa.

- Sistema ez da oso lagungarria gramatiken garapenean, eta problema bat dagoenean, azkenik Prologezko kodea aztertu behar izan da batzuetan. Honek argi erakusten du gramatika bat garatzeko sistema batean beharrezkoa dela "debugger" gramatikala erabili ahal izatea.

Lehengo puntuetan aipatu denez, HPSGren ezaugarri aipagarrienak unifikazioaren erabilera eta lexikalizazioa dira. Honen ondorioa elementu lexikal oso konplexuak eta mota lexikalen sistema taxonomikoa erabili behar izatea da. Beraz, gramatika bat garatzeko, lexikoia da landu behar den elementu nagusia.

printzipio orokorrak
ezaugarri gramatikalak "finkoak"
Gramatika -> erregela gramatikalak
osagai lexikalen hierarkia

Gramatikarien lana

Dena den, HPSGko gramatika ingleserako planteatu denez, badaude egin beharreko aldaketak erregela gramatikal, printzipio orokorretan eta ezaugarrietan euskararen gramatika egin nahi denean.

Lexikoiaren definiziorako ezaugarri-egituren taxonomia definitu behar da, eta hau mota-definizioen bidez egingo da:

*identifikadorea* **type** *egitura* .

Adibidez:

```
phon        type [phon, _ ].

sem         type [sem, _ ].

agr         type      [agr, [!num, !per, !gen]].

maj         type      [maj, [!n,!v]].

head        type   [head, [@maj, !min, !form, !inv, !aux, !case, @agr, !heads]].
```

Notazioa:

- !ident -> [ident, _] edo egitura definitua ez duen ezaugarria (aldagaia).
- @ident -> ident izeneko egitura-motaren ezaugarri guztiak (heredentziaren bidez) dituen egitura.

Honela hierarkia osoa definitu daiteke:

```
lex         type      [lex, _ ].

subcat      type      [subcat, _ ].

loc         type      [loc, [@head, @subcat, @lex]].

syn         type      [syn, [@loc, @non_loc]].

dtrs        type      [dtrs, _ ].

sign        type      [@dtrs, @phon, @sem, @syn].
noun        type           @sign with
                  syn:loc:head:form = norm with
                  syn:loc:head:min = undef with
                  syn:loc:head:maj:n = plus with
                  syn:loc:head:maj:v = minus.
np(Case)    type      @noun & @sat with
                  syn:loc:head:case = Case.

transv(Phon,Form)     type @word(Phon) & @trans & @mainv(Form).
```

Hori emanda, hiztegia honela definituko da:

```
entry    @adject(good).
entry    @cn(cat).
entry    @det(the).
entry    @pn(mary).
entry    @preposition(to).
entry    @transv(eats,@fin).
```

Gramatikaren printzipio orokorrak eta erregelak:

```
hfp      type @psign with        % gune-ezaugarrien printzipioa
    syn:loc:head = Sign with
     dtrs:head_dtr:syn:loc:head = Sign.


principles type @hfp & @subp & @cop.   %    printzipio    orokorren
bilduma


1      rule    @principles & @csign with   % lehenengo erregela
                dtrs:head_dtr:syn:loc:lex = minus with
              syn:loc:subcat = [] with
              dtrs:comp_dtrs = [_].
```

Sistema hori erabilita, hurrengo puntuan euskararen gramatika txiki baten definizioa emango da.

## 3.2. Euskararako gramatika.

Gramatika egiteko esaldi enuntziatibo sinpleak (aditz bakarrekoak) hartu dira. Azken puntuan gramatika horren luzapenak proposatuko dira.

Aukeratutako esaldiak:

| |
|---|
| ni |
| guk |
| guri |
| etxe polita |
| zenbait neska politi |
| bi gizon ak |
| gizon biek |
| mutil bat dator |
| nik katu txikia daukat |

Tratatutako egiturak

- izen sintagmak zenbatzaile, adjektibo eta kasuarekin
- aditz-sintagma sinpleak -> intrantsitibo eta trantsitiboak

Gramatikaren oinarria hartzeko euskararako proposatu diren bi gramatika hartu ziren [GO 80], [AB 88]. Lehenengoa gramatika deskriptiboa da eta bigarrena, berriz, formalismo konputazional baterako (LFG) pentsatuta dago. Dena den, bien arteko diferentziak ez dira handiak tratatzeko aukeratu den euskararen azpimultzoaren kasuan. Euskararako gramatika bat egiteko orduan modu errekurrentean agertzen den problema morfemen banaketa da: nola tratatu maila sintaktikoan hitz-forma baten osagaiak? Hitz-forma da unitate minimoa sintaxian edo bera osatzen duten morfemak? Lan honetan bi egile horien bidea jarraitu da eta suposatu da analisi sintaktikorako unitateak morfemak direla, analizatzaile morfologiko baten emaitza izan daitezkeenak (XUXENen [AG 92a] bezalakoak adibidez). Orduan, honakoa izango da analizatzailearen sarrera:

*gizon ak txori polit a dakar*

Gramatika egiteko mota desberdinetako zailtasunak agertu dira:

a) Ezaugarri sintaktikoen definizioa.

b) Gune-osagarriak erlazioak bereiztea.

c) Erregela gramatikalen baliagarritasuna.

d) Adjuntuen tratamendua.

e) Aurrekaritza linealeko murrizpenen inplementazioa.

a) Ezaugarri sintaktikoen definizioa.

Ingleserako proposatutako ezaugarrien artean badaude batzuk euskararako (edo behintzat euskararen azpimultzo honetarako) interesgarriak kontsideratu ez direnak:

* INV: honek ingleseko galderen ordenarekin du zerikusia. Euskararen kasuan badaude aldaketak galderen elementuetan ordena normalarekiko, baina hauek ez daudenez hartutako azpimultzoan, ez da erabili ezaugarri hau.

* NFORM: honen ordez, izenordainak eta izen arruntak bereizteko antzeko ezaugarria (izen-mota) erabili da, azken hauek adjuntuak hartzeko gaitasuna dutelako.

b) Gunea-osagarriak-adjuntuak erlazioak bereiztea.

HPSG formalismoaren pisurik handiena sintagma bakoitzaren gunean datza, kontzeptu honetatik abiatzen direlako formalismoaren mekanismo guztiak: erregela gramatikalak, azpikategorizazioaren printzipioa eta semantikaren printzipioa. Honen ondorioz, beharrezkoa da sintagma-mota guztietan elementuen arteko nolako erlazioa dagoen aztertzea gramatika egiteko. Hartutako erabakiak hauek dira:

* Izen-multzoa + Kasua

Beti dago kasua eta gunea da, honekin XBAR teorian proposatu diren ideiak hartuz. Nominatiboaren kasuan hartu den soluzioa kasua 0 bezala jartzea da. Honek gunea kasua izateko posibilitatea ematen du, ez lukeelako zentzurik edukiko gunea aukerazkoa dela esateak. Dena dela, soluzio hau bat dator XUXEN analizatzaile morfologikoak lortzen duen emaitzarekin.

Adibideak:

gizon 0
gizon polit a
gizon bat k

Kasua eta determinantea lotuta agertzen direla suposatu da, XUXENek ere horrelako emaitzak ematen dituelako. Honen ondorioa analisi ez simetrikoa ematea da zenbatzaileen kasuan, baina beste justifikazio bat normalean hauek determinantearen antzeko papera batetzen dutela da:

gizon ak
gizon horrek
gizon bat k
gizon bi ak

* (Izena + Adjetiboak*) + Zenbatzailea

Analisi honetan izen-multzoaren gunea izena dela suposatu da, honek ematen diolako informazio nagusia sintagmari. Zenbatzaileen kasuak dudak sortu zituen: beharrezkoak ez direnez ezin dira guneak izan; izena gunea bada, ordea, zenbatzailea azpikategorizatuko du, eta hau aukerazkoa denez, aukerazko osagarria definitu beharko da (inplementazioak disjunziorik onartzen ez duenez, sarrerak biderkatu egin behar izan dira).

Adjetiboen kasuan, hauek adjuntuen adibide argia dira, errepikatzeko posibilitatearengatik eta edozein izenek hartu dezakeelako.

Adibidez:

neska bat
zenbait neska
bi gizon polit
gizon polit bi

Izen-multzoaren barruko komunztadura tratatzeko aukera guzti hauek tratatu behar ziren:

gizon 0*
gizon a
gizon ak

(zenbatzailerik gabeko izenak determinantea behar du)

emakume bat 0
emakume bat a*
emakume bat ak*

("bi" ez diren zenbatzaileek bakarrik har dezakete mugagabea)

emakume bi 0
emakume bi a*
emakume bi ak

("bi" zenbatzaileak mugagabea edo mugatu plurala har ditzake)

Hau ebazteko bi ezaugarri berri behar dira, euskararen kasuan GPSG-ALVEY [AG 92b] inplementazioan proposatutakoak:

mugatasuna (+, -) eta numeroa (sg, pl)

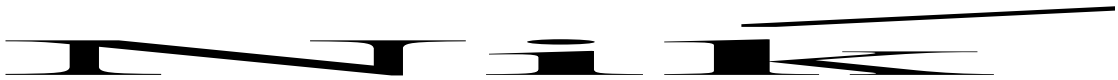
**Kasuak**:

| | |
|---|---|
| gizon (mug +) | 0 (mug -) |
| bat (mug -, num sg) | a (mug +, num sg) |
| bi (mug _, num pl) | ak (mug +, num pl) |

* Aditz erroa, laguntzailea eta osagarriak.

[AB 88] lanean proposatzen denez, bi aukera daude aditz erroa eta aditz laguntzailearen arteko erlazioa aztertzeko: gunea erroa edo gunea laguntzailea izatea.

Lehen kasuaren aldeko arrazoi nagusia aditz nagusiak bere osagarriak zehaztea da, zeren laguntzailea izanez gero, zaila izan daiteke adieraztea komunztadura ez duten osagarriak: gunea laguntzailea bada, honek ez du informaziorik kasu nuklearrak ez direnei buruz. Adibidez:



Adibide honetan ikusten da laguntzailea beste osagarrien maila berean geratzen dela. [AB 88] lanean bi osagai hauek sintagma osatzen dutela proposatzen da, baina hau egiteko bereiztu egin beharko litzateke aditz laguntzailea beste osagarriengandik (erregela gramatikal desberdinak planteiatuz).

Bigarren kasuan problema agertzen da aditz laguntzailea gunea bada, kasu honetan ezinezkoa delako, orain dauden printzipio orokor eta erregela gramatikal hartuta behintzat, kasu nuklearrak ez diren osagarriak azpikategorizatu.

[AB 88] lanean proposatzen den soluzioa, bi kasuetan, aditz laguntzailearen eta erroaren informazioa biltzea da, batetik bakarrik hartu barik. Hau errazagoa da LFG bezalako formalismo batean, ekuazioen bidez egin daitekeelako. HPSG formalismoan, esan denez, printzipioak edo erregelak aldatu beharko lirateke, eta honek konplikazioak izan ditzake, aldaketa hauek gramatika osoan dutelako eragina.

c) Erregela gramatikalen baliagarritasuna.

Inglesaren kasuan, bi erregela gramatikal proposatzen dira gunea-osagarriak motako egiturak tratatzeko: gunea azken osagarriekin (S -> NP VP edo NP -> DET N bezalako egiturak) eta beste batean gunea beste osagarri guztiekin. Euskararen kasuan, ez dugu arrazoirik ikusten bi erregela hauek bereizteko eta, beraz, erregela bakarra definitu dugu: gunea osagarri guztiekin (HPSG teoriaren azken bertsioan ere [PS ateratzeko] azken osagarri horien tratamendua aldatu egin da).

| SUBCAT<> | -> H | LEX+ | , C*

Ingleserako hirugarren erregela ere, galderen tratamendurako dena, aldatu beharra ikusten dugu euskararen kasuan. Dena dela, galderazko esaldirik tratatu ez dugunez, ez dugu proposamenik egingo.

d) Adjuntuen tratamendua.

[PS 87] lanean bi aukera proposatzen dira adjuntuen tratamendurako: guneak bere adjuntuak aukeratzea edo adjuntuak bere guneak. Azterketa honetan [PV 92] laneko proposamena hartu da eta adjuntuak bere gune posibleak aukeratuko ditu. Honela adjuntuen erregela berria definitzen da:

```
% Unsaturated head combines with one adjunct at a time
4 rule @principles & @asign & @unsat with
        dtrs:adj_dtrs = [@adjunct(A)] with
        dtrs:head_dtr:syn:loc:lex = minus with
        dtrs:head_dtr = select1(A).
```

Kasuak:

Iz -> Iz Adj*
IM -> IM Izlg*
Aditz-S -> Aditz-S Adlg*

e) Aurrekaritza linealeko murrizpenen inplementazioa.

Euskararako gramatikan ezin izan dugu, ingleserako bezala proposatutako hiru murrizpenekin, murrizpen orokorrik diseinatu eta, ondorioz, kasurik kasu tratatu dira. Inplementazioaren desegokitasun handia murrizpen hauek Prologez kodetu behar izatea izan da. Tratatu direnak hauek dira:

* Adjektiboa izenaren atzetik
* Zenbatzailea "Ordena" ezaugarriaren arabera
* Izen-multzoa kasuaren aurretik

Lan honetan hauek dira tratatu gabe geratu diren puntuak:

- Aditz erroa eta laguntzailearen arteko erlazioa definitu behar da. Lehen begirada batean zaila ematen du konponketa dotorea lortzea teoriaren erregela gramatikalak edo printzipio orokorrak aldatu gabe, honek eman dezakeen lan linguistikoa kontutan hartuz.

- Esaldi bateko osagarrien tratamendua. Inglesaren kasurako elementu baten osagarrien listak ordenari buruzko informazioa ematen du. Hau aprobetxatu izan da inplementazio honetan inglesaren ordena ia finkoa tratatzeko. Euskararen kasuan, nahiz eta teoriarekin problemarik ez egon (maila teorikoan ere, [GU 87] lanean printzipio orokorren aldaketak planteatzen dira), inplementazioa aldatu beharko litzateke, zeren elementu bat osagarrien listako lehen elementuarekin bakarrik konbina daiteke, eta ezin delako Nor-Nori-Nork esaldi bateko azpikategorizazioa deskribatu inplementazioa aldatu gabe.

- Esaldi subordinatuak (erlatibozkoak, konpletibak, zehargalderak, ...) tratatzea ez du ematen inglesarekin diferentzia handirik egongo denik, eta antzeko tratamendua aplikatzea posiblea dela suposatu daiteke.

- Aztertu ez den eta oso garratzitsua den aspektua urruneko menpekotasunena da.

# 4. Ondorioak.

HPSG formalismoaren azterketan eta euskararako gramatika txiki baten prestakuntzan egindako lan honen ondorioz, aspektu hauek azpimarra daitezke:

- HPSG teoria, nahiz eta edozein lengoaia tratazeko formalismo orokorra izan, ingleserako planteatu da eta, beraz, euskara bezalako hizkuntzekin aplikatzeko lan linguistiko handia egin behar da. Hau LFG teoriaren kasuarekin konpara daiteke, LFG hasieratik lengoaia desberdinetarako pentsatuta zegoelako eta, beraz, euskararako aplikazioa sinpleagoa zelako.

- HPSG teoriaren aldeko punturik handienak bere deklaratibotasuna (GPSG bezalako teoriak baino askoz modu sinpleagoan definituta eta semantika ondo definituarekin) eta bere homogenotasuna (teoria osoa maila bakarraren bidez definituta: ezaugarri-egituren bidez) dira.

- Nahiz eta teoria oso modularra izan, informazio guztia ahalik eta modu sinpleenean definitu nahi izateak problemak ekartzen ditu batzuetan. Adibidez, gertaera linguistiko bat tratatzeko erregela gramatikala edo printzipio orokorra aldatzen bada, orduan honek gramatika osoan du eragina, eta batzutan interakzioak nahiko konplexuak izan daitezke.

- Unifikazioan oinarritutako teoria guztiak bezala, oraindik frogatu gabe dago sistema hauen bideragarritasuna benetako aplikazioetan erabiltzeko (adibidez, testuen tratamenduan). [US 91] lanak aipatzen duenez, eraginkortasuna hobetzeko mekanismoak gehitu behar zaizkie aplikagarritasuna hitza erabili nahi bada.

## Bibliografia

[AB 88] Abaitua J. 1988. "*Complex predicates in Basque: from lexical forms to functional structures*" PhD. Thesis, University of Manchester, Centre for Computational Linguistics, Department of languages and Linguistics.

[AG 92a] Agirre E., Alegría I., Arregi X., Artola X., Díaz de Ilarraza A., Maritxalar M., Sarasola K., Urkia M., 1992.*"XUXEN: A spelling checker/corrector for Basque based on two-level morphology"*. Proceedings of the 3rd Conf. on Applied Natural Language Processing (ANLP'92) Trento 119-125.

[AG 92b] Agirre E., Sarasola K. 1992. "*ALVEY-ANLT euskararako gramatika*" Donostiako Informatika Fakultateko Lengoaia Naturaleko Taldearen1990/91 ikasturteko mintegia.

[GO 80] Goenaga P., 1980. "*Gramatika bideetan*", EREIN, Donostia.

[GU 87] Gunji T. 1987. "*Subcategorization and word order*" in Japanese Phrase Structure Grammar.

[KO 92] Kodric S., Popowich F., Vogel C. 1992. "*The HPSG-PL system . Version 1.2*" Centre for Systems Science, Technical Report CSS-IS TR 92-05, Simon Fraser University, Canada.

[PS 87] Pollard C.; Sag I., 1987. "*An Information-Based Approach to Syntax and Semantics: Volume 1 Fundamentals*", CSLI Lecture Notes No 13, Chicago University Press: Chicago.

[PS ateratzeko] Pollard C.; Sag I., 1987. "*An Information-Based Approach to Syntax and Semantics: Volume 2*"

[PV 91] Popowich F., Vogel C. 1991. "*A logic-based implementation of HPSG*" Natural language understanding and logic programming III, C. Brown and G. Koch (eds.), Elsevier Science Publishers.

[PV 92] Popowich F., McFetridge, Fass D., Haal G. 1992. "*Processing complex noun phrases in a natural language interface to a statistical database*", Proceedings of COLING-92.

[US 91] Uszkoreit H. 1991. "*Strategies for adding control information to declarative grammars*" Proceedings of the 29th meeting of the Association for Computational Linguistics, Berkeley, USA.

## E.1. Eranskina: Euskararako HPSGko gramatika.

## E.2. Eranskina: Gramatikaren emaitzak.

<u>Adibidea:</u> *gu 0 gabiltza*

```
dtrs _6777
phon < gu, zero, gabiltza >
sem
 indices <
  var _6809
  restriction
   reln gu
   arg1 _6809,
  var _6809
  restriction
   reln zero
   arg1 _6809 >
 cont
  quant _6865
  scope _6871
  index
   var _6883
   restriction _6889
  rlt
   reln gabiltza
   arg1 _6809
syn
 loc
  head
   maj
    n minus
    v plus
   min undef
   form _6955
   izen_mota _6961
   aux minus
   case _6973
   agr
    mug _6985
    num _6991
    per _6997
   heads <  >
   ordena _7009
  subcat <  >
  lex minus
```

Adibidea: *gizon polit bat 0*

```
dtrs _6689
phon < gizon, polit, bat, zero >
sem
 indices <
  var _6723
  restriction
   reln bat
   arg1 _6723,
  var _6723
  restriction
   reln polit
   arg1 _6723,
  var _6723
  restriction
   reln gizon
   arg1 _6723,
  var _6723
  restriction
   reln zero
   arg1 _6723 >
 cont
  quant _6831
  scope _6837
  index
   var _6723
   restriction
    reln zero
    arg1 _6723
  rlt _6873
syn
 loc
  head
   maj
    n minus
    v minus
   min undef
   form _6921
   izen_mota _6927
   aux _6933
   case { nom }
   agr
    mug minus
    num sing
    per 3
   heads <  >
   ordena _6988
  subcat <  >
  lex minus
```

Adibidea: *mutil ak dabiltza.*

```
dtrs _6825
phon < mutil, ak, dabiltza >
sem
 indices <
  var _6857
  restriction
   reln mutil
   arg1 _6857,
  var _6857
  restriction
   reln ak
   arg1 _6857 >
 cont
  quant _6913
  scope _6919
  index
   var _6931
   restriction _6937
  rlt
   reln dabiltza
   arg1 _6857
syn
 loc
  head
   maj
    n minus
    v plus
   min undef
   form _7003
   izen_mota _7009
   aux minus
   case _7021
   agr
    mug _7033
    num _7039
    per _7045
   heads <  >
   ordena _7057
  subcat <  >
  lex minus
```

<u>Adibidea:</u> *gizon bi ak dabiltza.*

```
dtrs _6689
phon < gizon, bi, ak, dabiltza >
sem
 indices <
  var _6723
  restriction
   reln bi
   arg1 _6723,
  var _6723
  restriction
   reln gizon
   arg1 _6723,
  var _6723
  restriction
   reln ak
   arg1 _6723 >
 cont
  quant _6805
  scope _6811
  index
   var _6823
   restriction _6829
  rlt
   reln dabiltza
   arg1 _6723
syn
 loc
  head
   maj
    n minus
    v plus
   min undef
   form _6895
   izen_mota _6901
   aux minus
   case _6913
   agr
    mug _6925
    num _6931
    per _6937
   heads <  >
   ordena _6949
  subcat <  >
  lex minus
```

Adibidea: *zenbait emakume k.*

```
dtrs _6817
phon < zenbait, emakume, k >
sem
 indices <
  var _6849
  restriction
   reln zenbait
   arg1 _6849,
  var _6849
  restriction
   reln emakume
   arg1 _6849,
  var _6849
  restriction
   reln k
   arg1 _6849 >
 cont
  quant _6931
  scope _6937
  index
   var _6849
   restriction
    reln k
    arg1 _6849
  rlt _6973
syn
 loc
  head
   maj
    n minus
    v minus
   min undef
   form _7021
   izen_mota _7027
   aux _7033
   case { erg }
   agr
    mug minus
    num _7070
    per 3
   heads <  >
   ordena _7088
  subcat <  >
  lex minus
```