

Sintaxi partziala

*Aduriz I., **Aranzabe M. J., **Arriola J.M., ***Diaz de Ilarraza A.

*Universitat de Barcelona

**Euskal Filologia Saila, EHU/UPV

***Lengoaia eta Sistema Informatikoak, EHU/UPV

Laburpena

Lan honetan hizkuntzaren prozesamenduko hainbat aplikaziotarako baliagarri den sintaxi partzialaz jardungo dugu. Partzialtasun hori testu libreei¹ analisi sendoa² emateko beharrazatik sortzen da. Sintaxi konputazionalako analisiak izango duen xehetasun maila hizkuntzaren prozesamenduaren helburuaren mende dago. Helburua testu libreak analizatzea denean, sintaxi partzialeko estrategia nagusitzen da. Hurbilpen horretako analizatzaileek informazioaren zati bat, ez guztia, aztertuko dute. Erabiltzen diren teknikek fidagarritasuna eta sendotasuna dute helburu, sakontasuna eta osotasuna neurri batean galduz. Hala ere, hainbat aplikazioen oinarri da eta, gainera, analisi sintaktiko osoagorako ere baliagarri da. Ildo horretatik, euskararako egin dugun ekarpena azaldu aurretik, sintaxi konputazionalaren baitan kokatu, ezaugarri nagusiak azaldu eta, era berean, oinarritzko zenbait kontzeptu argituko ditugu. Euskararako sintaxi partzialak Murriztapen Gramatikaren garapena izan du lehen urrats, ondoren, kateak³ definitzeko gramatikak idatzi eta aplikatu dira. Azkenik, kateen identifikazioan lorturiko emaitzak ebaluatuko ditugu.

1. Sarrera

Sintaxi konputazionalako analizatzaile sintaktikoek edota *parser*-ek⁴ sorturiko analisien xehetasun edota sakontasun maila hizkuntzaren prozesamenduaren helburuaren mende dago. *Parsing* osoa egiten denean esaldien egitura sintaktikoak oso-osorik eman nahiko dira. Horrelakoetan, arazo nagusia izango da hainbat analisiren artetik analisi sintaktiko bakarra aukeratzea. Arazo horren arrazoi nabarmenena hizkuntzaren maila orotan agertzen den anbiguotasuna da.

Gaur egunean, oraindik orain, ez dago testu libreetako edozein esaldi osorik analizatuko duen analizatzailearik. Ingeleserako sortu da ahalmen zabaleko sistemarik, baina sistema horiek ere

¹ *Testu libre* terminoarekin edozein testu hutsi egiten zaio erreferentzia.

² *Analisi sendoa* terminoarekin adierazi nahi da testu libreak analizatuko direla eta, testu horretako hitz guztiei emango zaiela analisisiren bat.

³ *Katea* erabiliko dugu arloko lanetan erabiltzen den *chunk* terminoa adierazteko. Sintagma kategoriako zatia da eta, sintaktikoki erlazionaturiko hitzez osatuta dago.

oztopo itzel baten aurrean aurkitzen dira beren emaitzak aplikatu nahi dituztenean: emaitza posible bakarra ez, baizik eta dozenaka-edo analisi posible lortzen baitituzte aipatu testuetako perpauk arruntak analizatzerakoan; horretan oinarritutako aplikazioak ezinezkoak bihurtuz.

Hala ere, hizkuntzaren prozesamenduko aplikazio guztiek ez dute analisi sintaktiko osorik behar. Gainera, *parsing* osoak askotan behar den baino informazio gehiago ematen du, eta batzuetan gutxiago. Esaterako, informazioaren berreskurapenean nahikoa litzateke izen-sintagmak eta aditz-sintagmak ezagutzearekin. Aldiz, informazio-erazketan, testuen laburpenean eta galde-erantzun sistemetan mota honetako erlazio sintaktiko-semantikoak behar dira: agentea, denbora, kokapena, etab. Aplikazio horietarako ez da hain garrantzitsua egiturazko analisi sintaktikoak egitea.

Gauzak horrela bi lerro nagusi bereizten dira sintaxi konputazionalan. Batetik, konputagailuaren laguntzaz hizkuntzaren deskribapen sakona lortzeko gero eta gramatika osatuagoak eraikitzea; bestetik, perpauk oso-osoen analisia lortzeko asmorik gabe, zatiak (ingelesez *chunk* esaten duten horietakoak: izen-sintagmak, adizlagunak, aditz-kateak, ...) bereizteko tresnak sortzea, tresna horiek hizkuntza-teknologiako aplikazioak garatzeko baliagarriak izango direlako, hala nola, hizketa-sorkuntzan, itzulpen automatikoan, informazio-erazketan, laburtze automatikoan, edo galdera-erantzunetako sistemetan.

Analisi partziala *parsing* osoaren alternatiba bilakatu da. Nahiz eta, *parsing* osoak ematen duen informazio oso zehatza galdu, edozein testu libre analizatzeko erronkari aurre egiten dio eraginkortasunez eta fidagarritasunez, eta, horrezaz gainera, esan bezala, analizatzaile partzialek ematen duten informazio sintaktikoa baliagarri da hizkuntzaren prozesamenduko hainbat aplikaziotarako. Vebrmobil⁵ izeneko proiektuan, *parser* partzialak baliatu ziren itzulpen-sistema sendotzeko, esaterako.

Parser partzialen ezaugarri nagusiak Abney (1997)-n dakarrenari jarraituz hauexek dira:

"Partial parsing techniques aim to recover syntactic information efficiently and reliably from unrestricted text, by sacrificing completeness and depth of analysis".

⁴ *Analisi sendoa* terminoarekin adierazi nahi da testu libreak analizatuko direla eta, testu horretako hitz guztiei emango zaiela analisisiren bat.

⁵ <http://www.dfki.de/pas/f2w.cgi?iuic/vebrmobil-e>.

Sintaxi konputazionalako hurbilpen horretan zenbait ataza egiten dira: testuak kateetan zatitzea, izen-sintagma kateak zehaztea edota klausula-mugak identifikatzea, besteak beste. Zertan datza testua kateetan zatitzea? Zer dira *chunk* edota kate direlako horiek?

Chunk terminoa Abney-ri zor zaio (Abney 1991). Horren orde, esan bezala kate erabiliko dugu. Katea sintagma kategoriako zatia da eta, sintaktikoki erlazionaturiko hitzez osatuta dago. Katearen barneko egitura aztertuz gero, burua eta bere modifikatzaileak bereiziko genituzke. Horrela bada, testua kateetan zatitzea gainjartzen ez diren eta elkarrekin sintaktikoki erlazionaturik dauden hitz multzoak atzematean datza. Hitz multzo horiek, beraz, ez errekurtsiboak izango dira, hau da, ezin dute beren baitan beste hitz multzorik edota katerik izan.

Ildo horretatik, sintaxi partzialak hitzari zuzen-zuzenean loturiko egitura sintaktikoen berri emango digu. Horretarako, esaldia gainjartzen ez diren kateetan banatuko da. Kate horietako hitz bakoitzak kategoria sintaktikoa izango du. Kateen analisi hori esaldi osoaren oinarri izango da. Adibidez, *Udaberrian hegoaldeetik ifarralderantz eta udazkenean ifarralderantz hegazti ugari pasatzen da gure mendien eta herrien gainetik* esaldia honela zatituko genuke kateetan:

- (1) [Udaberrian] [hegoaldeetik] [ifarralderantz] [udazkenean] [ifarralderantz] [hegoalderantz] [hegazti ugari] [pasatzen da] [gure mendien eta herrien gainetik].

Euskararen prozesamendu sintaktikorako jorratutako bideetariko bat sintaxi partzialean kokatzen da. Horretarako, Constraint Grammar (CG) (Karlsson et al 1995) formalismoa jarraitu dugu. Formalismo horretan egiten den analisisa ere partziala eta azalekoa da, hau da, hitz-formei zuzen-zuzenean dagozkien egitura sintaktikoak esleitzen zaizkie. Egitura sintaktiko horiek analisi morfologikoa izango dute oinarri. Horrez gain, sortuko diren analisi aukerak gutxitzeko murriztapenak ezarriko dituzte. Murriztapen horiek, hitzaren azaleko morfologia, hitzen arteko dependentzia sintaktikoak, eta hitz-ordena izango dute kontuan; beraz, ez dira egitura-printzipio abstraktuetan oinarrituko.

Sintaxi partzialari esker euskarazko testu libreak analizatuko dira eta, analisi-prozesua gauzatzeko erabiltzen diren modulu nagusiak hauek dira:

- MORPHEUS : analizatzaile morfosintaktikoa (Alegria et al 1996). Testu-hitzen analisi morfosintaktikoaz arduratzen da, hitz bakunak zein hitz anitzeko unitate lexikalak analizatuko dira.
- EUSTAGGER: euskararako lematizatzailea da eta bere zeregina anbiguotasun morfosintaktikoa murriztean datza. Horrez gain, hitz bakoitzari funtzio-etiketa bakarra esleitzen saiatzen da.
- ZATITZAILEA (*CHUNKER*): kateen analizatzailea. Ezagutuko diren egitura sintaktiko ez-osoak hauek izango dira: entitateak, postposizio konplexuak⁶ eta kateak (sintagmak eta aditz-kateak; ikus 4. atala).

Euskararako analizatzaile sintaktikoa edota *parser*-a edozein testu analizatzeko gai izan behar denez gero, analizatzailea modulu horietan banatu dugu. Hartara, analisi osoari ekin ordez, pausoz pauso egin beharreko oinarritzko analisi partzialak modulu bakoitzari dagozkion zereginak bereiziz bideratu ditugu. Analisi urrats bakoitzean aurreko moduluaren informazioa hartzen da aintzat eta, ondorengo prozesuan informazio hori aberastu egingo da. Ikus ditzagun (1)eko adibidearen bidez analisi partzialaren urrats nagusiak:

- a. MORPHEUS-en emaitza: hitz bakoitzeko analisi guztiak ematen ditu lerroka. Analisi-lerro bakoitzaren hasieran lema eta analisi morfosintaktiko bat. Eta jarraian, funtzio-etiketa @ ikurrarekin.⁷ Zenbaitetan funtzio-etiketa bat baino gehiago ager daitezke.

```
(1a)
/<Udaberrian>/<HAS_MAI>/
  ("udaberri" IZE ARR DEK NUMS MUGM DEK INE @ADLG)
/<hegoaldetik>/
  ("hegoalde" IZE ARR DEK NUMS MUGM DEK ABL @ADLG)
/<ifarralderantz>/
  ("iparralde" /ifarralde/ IZE ARR DEK NUMS MUGM DEK ABZ @ADLG)
/<eta>/
  ("eta" LOT JNT EMEN @PJ)
  ("eta" LOT MEN KAUS @PJ)
/<udazkenean>/
  ("udazken" IZE ARR DEK NUMS MUGM DEK INE @ADLG)
/<ifarraldetik>/
  ("iparralde" /ifarralde/ IZE ARR DEK NUMS MUGM DEK ABL @ADLG)
/<hegoalderantz>/
  ("hegoalde" IZE ARR DEK NUMS MUGM DEK ABZ @ADLG)
/<hegazti>/
  ("hegazti" IZE ARR DEK ABS MG @OBJ @SUBJ @PRED)
  ("hegazti" IZE ARR @KM>)
```

⁶ Postposizio konplexuak: -ren ondoren, 0 gisa ...

⁷ @ ikurrarekin adierazten dira funtzio-etiketak. Analisisetan agertzen diren horiek guztiak eta gainerako sinboloen esanahia artikuluaaren bukaeran jasoko dugu.

```

/<ugari>/
("ugari" ADJ IZO DEK ABS MG @OBJ @SUBJ @PRED)
("ugari" ADJ IZO @<IA)
("ugari" DET DZG MG DEK ABS MG @SUBJ)
("ugari" DET DZG MG @ID>)
("ugaritu" ADI SIN AMM ADOIN @-JADNAG)
/<pasatzen>/
("pasatu" ADI SIN AMM ADIZE DEK INE @OBJ @-JADNAG_MP_OBJ)
("pasatu" ADI SIN AMM ADOIN ASP EZBU @-JADNAG)
/<da>/
("izan" ADL A1 NR_HU @+JADLAG)
("izan" ADT A1 NR_HU @+JADNAG)
/<gure>/
("gu" IOR PERARR NUMP GU DEK GEN DEK ABS MG @IZLG> @<IZLG @OBJ @SUBJ @PRED)
("gu" IOR PERARR NUMP GU DEK GEN @IZLG>)
("guretu" ADI SIN AMM ADOIN @-JADNAG)
/<mendien>/
("mendi" IZE ARR DEK GEN NUMP MUGM DEK ABS MG @IZLG> @<IZLG @OBJ @SUBJ
@PRED)
("mendi" IZE ARR DEK GEN NUMP MUGM @IZLG>)
/<eta>/
("eta" LOT JNT EMEN @PJ)
("eta" LOT MEN KAUS @PJ)
/<herrien>/
("herri" IZE ARR DEK GEN NUMP MUGM DEK ABS MG @IZLG> @<IZLG @OBJ @SUBJ
@PRED)
("herri" IZE ARR DEK GEN NUMP MUGM @IZLG>)
/<gainetik>/
("gain" IZE ARR DEK NUMS MUGM DEK ABL @ADLG)
/<.>/<PUNT_PUNT>/

```

- b. EUSTAGGER-aren emaitza: aurreko analisia desanbiguatuta. Hitz bakoitzeko testuinguru horri dagokion analisi-lerroa izango dugu.

```

(1b)
/<Udaberrian>/<HAS_MAI>/
("udaberri" IZE ARR DEK NUMS MUGM DEK INE @ADLG)
/<hegoalde>/
("hegoalde" IZE ARR DEK NUMS MUGM DEK ABL @ADLG)
/<ifarralderantz>/
("iparralde" /ifarralde/ IZE ARR DEK NUMS MUGM DEK ABZ @ADLG)
/<eta>/
("eta" LOT JNT EMEN @PJ)
/<udazkenean>/
("udazken" IZE ARR DEK NUMS MUGM DEK INE @ADLG)
/<ifarralderantz>/
("iparralde" /ifarralde/ IZE ARR DEK NUMS MUGM DEK ABL @ADLG)
/<hegoalderantz>/
("hegoalde" IZE ARR DEK NUMS MUGM DEK ABZ @ADLG)
/<hegazti>/
("hegazti" IZE ARR @KM>)
/<ugari>/
("ugari" DET DZG MG DEK ABS MG @SUBJ)
/<pasatzen>/

```

```

("pasatu" ADI SIN AMM ADOIN ASP EZBU @-JADNAG)
/<da>/
("izan" ADL A1 NR_HU @+JADLAG)
/<gure>/
("gu" IOR PERARR NUMP GU DEK GEN @IZLG>)
/<mendien>/
("mendi" IZE ARR DEK GEN NUMP MUGM DEK ABS MG @IZLG> @<IZLG @OBJ @SUBJ
@PRED)
("mendi" IZE ARR DEK GEN NUMP MUGM @IZLG>)
/<eta>/
("eta" LOT JNT EMEN @PJ)
/<herrien>/
("herri" IZE ARR DEK GEN NUMP MUGM @IZLG>)
/<gainetik>/
("gain" IZE ARR DEK NUMS MUGM DEK ABL @ADLG)
/<.>/<PUNT_PUNT>/

```

- c. ZATITZAILLE-aren emaitza: EUSTAGGER-aren emaitzari % ikurdun etiketak gehitu zaizkio. Etiketa horiek adierazten dute katea: sintagma edo aditz-kate den eta, kate bakoitza non hasi eta bukatzen den. Etiketa horien berri zehatza 4. atalean emango dugu.

(1c)

```

"<Udaberrian>"<HAS_MAI>"
    "udaberri" IZE ARR DEK NUMS MUGM DEK INE @ADLG HAS_MAI %SINT
"<hegoalde>"
    "hegoalde" IZE ARR DEK NUMS MUGM DEK ABL @ADLG %SINT
"<ifarralderantz>"
    "iparralde" "ifarralde" IZE ARR DEK NUMS MUGM DEK ABZ @ADLG %SINT
"<eta>"
    "eta" LOT JNT EMEN @PJ AORG
"<udazkenean>"
    "udazken" IZE ARR DEK NUMS MUGM DEK INE @ADLG %SINT
"<ifarralde>"
    "iparralde" "ifarralde" IZE ARR DEK NUMS MUGM DEK ABL @ADLG %SINT
"<hegoalderantz>"
    "hegoalde" IZE ARR DEK NUMS MUGM DEK ABZ @ADLG %SINT
"<hegazti>"
    "hegazti" IZE ARR @KM> %SIH
"<ugari>"
    "ugari" DET DZG MG DEK ABS MG @SUBJ %SIB
"<pasatzen>"
    "pasatu" ADI SIN AMM ADOIN ASP EZBU @-JADNAG NOTDEK %ADIKATHAS
"<da>"
    "izan" ADL A1 NOR NR_HU @+JADLAG %ADIKATBU
"<gure>"
    "gu" IOR PERARR NUMP GU DEK GEN @IZLG> %SIH
"<mendien>"
    "mendi" IZE ARR DEK GEN NUMP MUGM @IZLG>
"<eta>"
    "eta" LOT JNT EMEN @PJ AORG
"<herrien>"
    "herri" IZE ARR DEK GEN NUMP MUGM @IZLG>
"<gainetik>"
    "gain" IZE ARR DEK NUMS MUGM DEK ABL @ADLG %SIB

```

"<\$.>"<PUNT_PUNT>"
PUNT_PUNT

Aurreko (1)eko analisisiek erakusten duten bezala, Murriztapen Gramatikaren filosofia jarraitu dugu. Hau da, morfologia da analisi sintaktikoaren abiapuntu eta, ahal den neurrian hitzen analisi aukerak erabat murriztea da jomuga. Hau da, hitz bakoitzeko interpretazio-lerro bakarra. Interpretazio-lerro horretan, analisi morfologikoaren emaitzaz gain, desanbiguatze-lanaren ondoren, funtzio-etiketa sintaktiko bakarra izango dugu xede.

Hala ere, esan beharra dago analizatzaile partzial gehienek sarrera anbiguoen gainean ere lan egin dezaketela, beren emaitzetan ondorioz izan gabe. Euskararako garatu den analizatzaileak ere sarrera anbiguoekin lan egin behar izaten du. Sintaktikoki erabat desanbiguatutako corpusak ez ditugu erdiesten eta. Dena den, ebaluazioaren atalean ikusiko dugun bezala emaitza onak lortzen ditugu.

Ondoren sintaxi partzialean dauden joera nagusien ikuspegi orokorra aurkeztuko dugu, gure hurbilpena hobeki ulertzearen.

2. Corpus libreen prozesamendurako analizatzaileak

Hizkuntzaren prozesamendurako *parser* orokorrek probabilitateetan oinarrituak izan ala ez, ez dute hainbat aplikaziotarako behar den adina doitasun eta estaldura. Testu libreen analisi orokor doiak lortzeko zailtasun horren aurrean bi estrategia jorratuko dira: 1) analisi orokor baina azalekoak egitea, eta 2) parte batzuen analisi oso zehatzak erdiestea. Mota horretako analisiak lortzeko erabiltzen diren teknika multzoak analisi sendoaren izena hartzen du (Martí & Llisterri 2002). Teknika horietatik azaleko analizatzaileen eta analizatzaile partzialen estrategiak hartuko ditugu mintzagaitzat.

Azaleko analizatzaileak, oro har, etiketatzaileen aberastearen ondorio dira. Hitz bakoitzari dagozkion ezaugarri morfologikoak, azaleko funtzio sintaktikoak adierazten dituzten etiketekin aberasten dira. Horien artean kokatzen dira Constraint Grammars (Voutilainen 1994).

Bestalde, analizatzaile partzialek ahalik eta informazio sintaktiko zehatzena lortu nahi dute corpuseko zenbait zatiri buruz. Analizatzaile horiek ez dute analisi osorik behar. Nahikoa dute oinarritzko sintagmak ezagutzearekin. Horretarako, hasiera batean, hizkuntzalariek idatziriko erregeletan oinarritzen ziren *parser*-ak. Baina, kontuan hartuta, erregela horiek gerta daitezkeen

testuinguru guztiak (gehi salbuespenak) jaso beharko dituztela, gramatikak idaztea lan handia da. Hori dela eta, metodo automatikoetara lerratu ziren *parser*-ak sortzen zituztenak. Beraz, ez da harritzekoa *parser* gehienak automatikoki sortutakoak izatea: egoera finituko automaten edota ikasketa automatikoko hurbilpenen teknikak erabiliz nagusiki. Hala ere, badira beste hurbilpen batzuk: heuristikoetan oinarritutako erregelak eta testuingururik gabeko gramatika sinpleak, besteak beste.

Esan bezala, sintaxi partzialak sendotasuna du jomuga, hots, testu libreak bere osotasunean analizatzea. Horrek eragiten ditu: analisisen azalekotasuna eta zenbait parte zehatzago aztertuta azaltzea, besteak beste. Beraz, testuingururik gabeko gramatikak baino egokiagozat jotzen dira egoera finituko automatetan oinarritutakoak (Jurafsky D. eta Martin J.H. 2000). Automata horietan oinarritutako *parser*-ek kaskada metodoak erabiltzen dituztela esaten da (*cascade methods edota finite state cascade methods*). (Ait-Mokhtar eta Chanod 1997)-n ikus daitekeen bezala, egoera finituko teknikak erabil daitezke kateak atzemateko, eta subjektuak eta objektuak ere zehazteko. Automatak baliatuko dituzte Constraint Grammar formalismoan ere. Azaleko sintaxia lantzeko informazio linguistikoa erregelen bidez adieraziko da. Erregela horiek hizkuntzalariak idatzita daude. Eta, zenbait kasutan heuristikoak ere baliatuko dira, hau da, estatistika erabiliko dute.

Azken hamarkadan, ikasketa automatikoko hurbilpenen inguruko interesa handitu da. Funtsean oinarritzko egitura sintaktikoak ikasi nahi dira corpusetatik. Sintaktikoki etiketaturiko corpusen edota *treebank*-en gainean trebatuko dira ikasketa automatikoko algoritmoak. Beraz, oso garrantzitsuak dira *treebank*-ak. Metodo diferenteen arteko konparazioa errazteko CoNLL-2000 (*Conference on Computational Natural Language Learning*) delakoa sortu zen (Sang eta Buchholz 2000) datu multzo estandarra finkatzeko asmoarekin. Hartara, metodo horiek erabiltzen dituzten guztiek beren emaitzak konparatu ahal izateko. Datu horiek *Penn Treebank*-etik⁸ ateratako atalez osatuta daude. Ikasketa automatikoko hurbilpenaren baitan hainbat teknika aurkitzen dira literaturan: *Transformation-based Learning*, *Memory-based Learning*, *Hidden Markov Models*, *Maximum Entropy*, *Support Vector Machines*, etab. Hasierako lanek izen-sintagmak atzemateko helburua zuten nagusiki. (Ramshaw eta Marcus 1995) *Transformation-based learning*; *memory-based* (Daelemans et al 1999).

Bestalde, gaur egun, ikasketa automatikoko hurbilpenaren ildotik *parsing* partzialeko teknika oso sendoak ari dira garatzen, adibidez: RASP (*Robust Accurate Statistical Parsing*).⁹

Gain-gainetik bada ere, sintaxi partziala kokatzeko paratu dugun atal honi bukaera emateko, esan beharra dago gaur egun zabalik dagoen ikerlerroa dela mota honetako analizatzaileak analizatzaile sintaktiko sakonagoekin integratzea.¹⁰ Horrela bada, *parser* partzialak erabiliak dira aurreprozesatzaile gisa zenbait arazori aurre egiteko, eta modu horretara analizatzaile sakonagoen lana errazteko, esate baterako anbiguotasuna jaitsez edota *treebank*-a eraikiz (Hindle-ren analizatzaile partzial ezaguna, Fidditch, alegia, erabili zen *Penn Treebank* eraikitzeke).

3. Euskararako azaleko sintaxia. Murritzapen Gramatika: azaleko funtzio sintaktikoak

Lehen esan bezala, syntaxiko lanei ekiteko CG formalismoa erabili dugu. Formalismo ahaltsua eta sendoa da testu-masa handiak (corpus handiak) eta testu libreak edo errealak analizatzea duelako helburu. Izan ere, forma eta egitura guztiak analizatzea du helburu azaleko analisisa eta analisi partziala aurrera eramanez. Hizkuntzalaritzaren ikuspegitik, bestalde, oso formalismo intuitibo eta erabilerraza da. Batez ere ahaltsua da desanbiguazio-lanetarako, hau da, analisi morfologikoa egin ondoren gelditzen diren anbiguotasunak ebazteko. Ikus horretarako sarrerako adibidea: MORPHEUS-en emaitzatik, EUSTAGGER-en emaitzara dagoen aldea.¹¹ Euskaraz egindako desanbiguazio-gramatika morfosintaktikoa (EUSMG¹² deitua) duela urte batzuk bukatua dago eta luze azaldua dago (Aduriz, 2000) tesi-lanean eta hainbat publikaziotan (Aduriz et al, 2003; Aduriz eta Diaz de Ilarraza, 2004). Halaber, lan honetatik ateratako tresna, lematizatzaile/etiketatzailea (EUSTAGGER) hainbat aplikaziotan erabili izan da: terminoen erauzketan, JALGI bilatzailean, etab.

Desanbiguazio morfosintaktikoz gain, gramatikak desanbiguazio sintaktikoa ere aurrera eramaten du, hau da, funtzio sintaktikoei dagokiena. Izan ere, analisisaren maila

⁸ <http://www.cis.upenn.edu/~treebank/home.html>

⁹ <http://www.cogs.susx.ac.uk/lab/nlp/rasp>

¹⁰ Deep Thought proiektuan adibidez: <http://www.project-deepthought.net>

¹¹ Kontuan izan behar da analisi morfologikoa hitzez hitz egiten dela hitzaren testuinguruari begiratu gabe. Horrela, maila bateko baino gehiagoko anbiguotasuna aurki dezakegu. Adibidez, *omen* forma *omendu* aditzaren aditz-erroa, izen arrunta eta *omen* partikula ere izan daiteke. Edo beste maila bateko anbiguotasuna aipatuz, adibideko *zeunden* hitzeko *-(e)n* mendeko atzizkia, erlatibozko egitura osa dezake, zehar galdera izan daiteke, esate baterako.

¹² EUSkararen Murritzapen Gramatika.

morfosintaktikoan bezalaxe, forma bakoitzak funtzio sintaktiko bat baino gehiago izan dezake, testuinguruari begiratu gabe egiten baita analisisa.¹³ Sarrerako adibidean analisi-lerro bakoitzaren bukaeran aurkitzen ahalko ditugu funtzioak @ markaren ondoren. MORPHEUSen emaitzan (ikus (1a)) analisi-lerro bakoitzak funtzio bat baino gehiago du askotan. EUSTAGGER-en emaitzan, berriz, lerro bakoitzak funtzio bakarra du (ikus (1b)).

Ez dugu sakonduko funtzio sintaktiko hauen desanbiguazio-prozesuan, izan ere, erregela-mota bera da erabiltzen dena bai desanbiguazio morfosintaktikoan bai funtzio sintaktikoen desanbiguazioan, eta goian esan bezala, beste argitalpenetan horren berri zabalagoa aurki daiteke.

Funtzio sintaktiko horiei erreparatuko diegu puntu honetan bereziki, funtzio sintaktikoak baitira azaleko sintaxiaren tratamenduaren oinarria, ondorengo puntuetan ikusiko dugun bezalaxe.

3.1. Funtzio sintaktikoen izaera eta sailkapena.

Gramatika honek analizatuko duen testuak, beraz, analisi morfologikoa eginga izan behar du, hau da, testuaren hitz bakoitza modu isolatuan hartuta, zein informazio morfologikoa duen esana egon behar da (hitz bakoitzaren kategoria(k), azpikategoria(k), eta gainontzeko informazio morfologikoa (ikus sarrerako adibidea). Era berean, testuko hitz guztiek izan behar dute funtzio sintaktiko bat gutxienez. Eskakizun honetatik dator, neurri batean behintzat, puntu honetan azalduko ditugun funtzioen berezitasuna. Izan ere, askotan funtzio sintaktiko klasikoak izango dira, baina beste askotan, goian aipatutako eskakizun horrengatik, funtzio sintaktiko tradizioaletan aurkitzen ez diren etiketatara jo behar izan dugu, etiketa funtzioaletara, alegia.

Bestalde, funtzio sintaktikoek buru/modifikatzaile¹⁴ banaketari jarraitzen diote, eta ondorioz, bi motatakoak dira: funtzio sintaktiko nagusiak, eta funtzio sintaktiko modifikatzaileak (@...> eta @<...¹⁵). Funtzio sintaktiko nagusiak, buruei dagozkien etiketak dira: subjektua, osagarri zuzena,

¹³ -ak bukatu hitzek absolutibo plurala eta ergatibo singularra izan daitezke. Funtzio sintaktiko mailan hauen anbiguotasuna subjektu eta objektuarena izango da.

¹⁴ Ez da nahasi behar hemengo *funtzio sintaktiko modifikatzailea*, gramatika sortzailean erabiltzen den *modifikatzaile* terminoarekin. Gramatika honentzat, etiketa horiek daramaten elementuak, funtzio sintaktiko nagusia duen elementuari nolabait laguntzen doaz. Murriztapen Gramatika ingelesez egin zenean (Constraint Grammar Parser), horrelakoak erabili ziren (Karlsson, 1995:405). Honi buruzko informazio gehiago lortzeko ikus Voutilainen lanak (1994; Voutilainen et al 1992:21). "The functions are largely derived from descriptive grammars like [Quirk, Greenbaum, Leech and Svartvik, 1985]." (Voutilainen 1994:16).

¹⁵ Funtzio sintaktiko modifikatzaileek *handiago-txikiago* zeinuz (> <) lagunduta doaz buru/modifikatzaile erlazioak zein noranzkoan gauzatzen diren adierazteko.

zehir osagarria, adizlaguna, predikatua, aditz nagusia eta laguntzailea. Funtzio modifikatzaileak, berriz, izenlaguna, determinatzailea, etab. Azken hauek, laguntzen duten elementuarekiko noranzkoa markatzen dute etiketan, hau da, buruarekiko direkzioa. Horrela, sintagma barruan dauden dependentzia sintaktikoen berri ematen da. Izan ere, funtzio etiketok klausularen edota perpausaren barnean hitz-forma bakoitzak duen oinarritzko dependentzia-harremana adierazten dute modu inplizitu batean. Adibidez, @IZLG> etiketak zera esan nahiko du: eskuinetara dagoen izenki baten izenlaguna dela (*herrien-@IZLG> gainetik-@ADLG*).

Gauzak horrela, sailkapen bat egin dezakegu orain arte aipatutako informazioa kontuan hartuta.

1. Funtzio sintaktiko “klasikoak” edo ohiko deskribapen gramatikaletan erabiltzen direnen artean hauek genituzke: subjektua (@subj), objektua (@obj), zehir objektua (@zobj), atributua (@atrib), predikatiboa (@pred), adizlaguna (@adlg), izenlaguna (@izlg).
2. Etiketa funtzionalak. Aipatu dugunez, gramatikak eskatzen du testuko hitz bakoitzak analisi morfologikoari dagokion informazioa edukitzea eta funtzio sintaktiko bat edo etiketa funtzional bat gutxienez, ondorengo prozesuen oinarri izango baitira elementu hauek. Hori dela eta, funtzio sintaktiko tradizionalekin zerikusirik ez duten etiketak erabiltzen ditugu hitz guztiak etiketatuak egon daitezzen. Esate baterako hauek genituzke: @ADILOK (aditz lokuzioetako elementuek daramatena *behar izan* eta antzekoak), @HAOS (hitz anitzeko forma baten osagaiak hartzen duena, *hain zuzen ere* lokailuaren osagaiak hartuko luketena, adibidez), @LAB (laburdurek hartzen dutena), @KM> kasua daraman formaren guneak hartzen duena (*hegazti ugari* sintagman, *hegazti* formak hartuko luke), etab.

Multzo nagusi hauetako funtzio bakoitzak bere aldetik, buru/modifikatzaile ezaugarriari erantzuten diola kontuan izanik, funtzio nagusiak ala laguntzaileak izan daitezke:

1. Funtzio sintaktiko nagusiak. Sintagma edo egitura baten buruari egiten diote erreferentzia. Izen-sintagmetan, kasu-marka daraman hitzak izan ohi du funtzio sintaktiko nagusia, postposizio-sintagmetan postposizioa daramanak, aditz-sintagmetan aditz nagusiak eta aditz laguntzaileak, eta oro har, forma beregainek dute funtzio sintaktiko nagusia, hala nola, lokailuak, juntagailuak, etab.

2. Funtzio sintaktiko modifikatzaileak: buruarekiko noranzkoa markatzen dutenak dira. Adibidez: @IZLG> (burua eskuinetara duen izenlaguna da: *etxeko ataria*), ala @<IZLG (burua ezkerretara duen izenlaguna da: *aingeru guardakoa*).

Funtzio sintaktikoen formalki duten itxurari buruz, berriz, sinpleak ala konposatuak izan daitezke.

1. Funtzio sintaktiko sinpleak elementu batez osatutakoak dira, orain arte ikusi ditugunak bezalakoak. Ezaugarri baten berri ematen dute soilik: subjektua den, objektua den, aditz nagusia ala laguntzailea den, etab.
2. Funtzio sintaktiko konposatuek berriz, ezaugarri baten baino gehiagoren berri ematen duten elementu anitzez osatuta daude. Aditzen etiketak, esate baterako, ia denak dira konposatuak hainbat informazio ematen dutelako, alegia, jokatu/ez-jokatu, laguntzaile/nagusi, mendekoa den ala ez, eta mendekoa den kasuan zein funtzio betetzen duen ere adierazten baita. Adibidez, @-JADNAG_MP_SUBJ etiketak honako informazio ematen digu: (@-J) ez-jokatua, (ADNAG) aditz nagusia, (_MP) mendekoa eta (_SUBJ) mendekoaren funtzioa subjektua da.

Hurrengo puntuan kateen gramatikari buruz luze eta zabal jardungo dugu eta orduan ikusiko dugu gramatika horren erregelak aurretik ikusitako funtzio sintaktikoen banaketa funtsezkoan oinarritzen direla, alegia, nagusiak ala modifikatzaileak diren. Izan ere, funtzio nagusia duen elementu batek osa dezake sintagma bat, eta era berean laguntzen duten elementuak izan ditzake bere barruan (funtzio sintaktiko modifikatzaileak). Baina modifikatzaile batek ezingo du bakarrik sintagmarik osatu, modifikadore direnez, beti beste elementuren bati laguntzen baitoaz.

Hori horrela izanik, alegia, ondorengo urratsetan duten zereginaz jabeturik, garbi ikusten da zein garrantzizkoa den funtzio sintaktikoak ondo esleituak eta desanbiguatuak edukitzea, hortik baitator hurrengo urratsen arrakastaren parte handi bat.

4. Kateen gramatika: printzipioak eta kate motak

Urrats honetan, analisi prozesuko EUSTAGGER-aren analisiaren irteera abiapuntutzat harturik, zenbait oinarritzko egitura sintaktiko edo kate (aditz-kateak eta sintagmak) ezagutuko

ditugu. Esaldia osatzen duten eta funtzio sintaktikoen bitartez implizituki adierazita dauden hitzen arteko dependentzia-harremanak azaleratuko dira, egitura sintagmatikoak adieraziz (ikus sarrerako adibideko (1c), hots, ZATITZAILE-aren emaitza). Kate horiek guztiak harrapatzeko osatu diren erregelak (Arriola 2000) tesi-lana dute abiapuntu eta horien aberasketaren ondoren bi gramatika definitu dira:

- Aditzei dagokien gramatika: aditzen funtzio sintaktikoetan eta aditz-kateko partetzat hartu ditugun modalitatea eta egiatasuna adierazten duten partikuletan oinarrituta, aditz-kate jarraituak eta gehienez ere hiru osagai dituzten aditz-kate ez jarraituak ezagutu ditugu. Gramatika honetan erabiltzen diren etiketak, kateak adierazteko etiketak, alegia, hauek dira:
 - %ADIKATHAS: aditz-kate jarraitu bateko lehenengo elementuari esleitzen zaion etiketa.
 - %ADIKATBU: aditz-kate jarraitu bateko azken elementuari esleitzen zaiona.
 - %ADIKAT: elementu batez osaturikoari.
 - %ADIKATETENHAS: aditz-kate ez jarraituetako lehen osagaia.
 - %ADIKATETEN: aditz-kate ez jarraituetako bigarren osagaia.
 - %ADIKATETENBU: aditz-kate ez jarraituetako azken osagaia.
- Sintagmen gramatika: sintagmak ezagutu ahal izateko, aditz-kateekin egin dugun bezalaxe, ondoko etiketak definitu dira: sintagmen hasieran (%SIH) eta bukaeran (%SIB); eta hitz bakarreko sintagmetan (%SINT) .

Sintagma-etiketa hauek aditz-funtzio ez diren gainontzeko funtzio sintaktikodun elementuei esleituko zaizkie. Funtzio sintaktikoetan egin den modifikadore eta funtzio nagusi bereizketari esker ezartzen ditugu sintagma-hasierak eta sintagma-bukaerak, eta baita elementu bakarreko sintagmak ere. Gramatika honetako arau nagusia elementuak lotzen joatea da, harik eta funtzio sintaktiko nagusi bat aurkitu arte. Horrela modifikadore den elementu oro ezin izango da sintagma-bukaera izan (%SIB). Funtzio sintaktiko nagusidunak aldiz, sintagma bukaera (%SIB) etiketa eramango dute, edota elementu bakarreko sintagmaren etiketa (%SINT).

Arau nagusi horrek badu salbuespenik, sintagma koordinatuak osatzerakoan funtzio sintaktiko nagusidun juntagailua azaltzen baitzaigu, baina, jakina, hau ezin da inola ere sintagma baten bukaera izan.

Gramatika horietan definitutako erregelen bidez ezagutzen diren egitura sintagmatikoak edota kateak deskribatuko ditugu jarraian:

4.1. Aditz—kateak

Ezagutzen diren aditz-kateak osagaien kopuruari erreparatuz honela sailka daitezke:

4.1.1. *Osagai bakarreko katea*: ondoko adibide¹⁶ hauetan azpimarratuta dauden osagaiak dira. Osagai hauek %ADIKAT modura markatuko dira analizatzaile sintaktikoan. Zein dira?

- Adizki jokatueta aditz trinkoak; adizkia soilik agertzen denean, hau da, *ahal, ez, omen, ba...* moduko partikularik gabe (*Mirenek badaki; Egiazko tresora lan segituan datzala*)

1) Adizki ez jokatueta, berriz, i) perpaus lokabeetan agertzen diren aditzoinak edota partizipioak: aginduak ematean (*Urrikal jauna!, Etorri berehala!*); ii) galdera erretoriko edo harridurazkoetan (*Nork hari eskua eman?*); iii) errefracu edo itxura bereko esapideetan; iv) kontakizunetan... (*Lehenik zernahi esan, gero zernahi egin eta azkenean zernahi ikus*); v) ("*Sar eskuinetik... ama besarka...*"). Gauza bera perpaus txertatueta ere: vi) errefracueta (*Otsoa non aipa, han gerta*); vii) zehar galderetan (*Ez dakit nondik sar*); viii) hautuzko perpausetan (*Zuk aitor ala hark sala, eni berdin zait*); ix) zenbait postposizioekin (*Bazkal ostein...*) eta zenbait esapidetan aurkituko dugu aditz oina (*Ikus ahala urrun*). Kasu horietan askok partizipioa ere erabiltzen dute. Adjektiboa ageri den testuinguru berberetan ager dakiguken partizipioa ere horrela markatzen da (*Ama hila dugu aspaldion. Sagarrak amari emanak dizkiot*). Partizipioarekin jarraituz, postposizio batzuek eskatzen duten partizipioa ere osagai bakarrekoa izango da (*Ikusi arte*). Osagai bakarreko aditz-kateekin bukatzeko, jokatu gabeko zenbait mendeko perpausen aditza ere horrelakoa izango dugu (*Goizean etorri gizona*); eta azkenik, aditz-izena bera (*Hori irakurtzeak buruko mina ematen dit. Izertitzea ona dela osasunerako*).

4.1.2. *Osagai bat baino gehiagoko kateak*: ondoko adibide hauetan azpimarratuta dauden aditz-kate jarraituak ditugu batetik, eta bestetik beltzago adierazita dauden aditz-kate ez jarraituak. Lehenak, analizatzaile sintaktikoan %ADIKATHAS eta %ADIKATBU etiketen bitartez adierazten diren kateak ditugu; eta bigarrenak %ADIKATETENHAS %ADIKATETENBU modura markatutakoak:

¹⁶ Adibide hauek *Euskal Gramatika Lehen Urratsak-II* (1987) liburutik eta Arriolaren tesi-lanetik (2000) hartuak izan dira.

-Adizki jokatueta aditz laguntzailea behar duten aditz konposatuak, batetik (*Etorri da. Jakin izan banu, etorriko nintzatekeen. Bere inozotasuna aurpegiratu zion*), eta bestetik aditz trinkoak, baldin eta partikularen batekin agertzen badira (*Ez dator*).

- Adizki jokatueta zein jokatugetan aditzaren inguruan agertzen diren bestelako partikulak ditugunean (*Mikel ez dator. Ba al daukazu dirurik hau erosteko?. Ez etorri!. Etortzen ahal naiz. Ezin zuen ogirik erosi. Etorri ohi da*).

- Izen/adjektibo + izan/ukan aditz-esapideak (*behar izan, nahi izan, bizi izan, balio izan, maite izan, komeni izan, merezi izan...*). Aditz-lokuzioak kate bezala hartzeko, horietako osagai izan daitezkeen multzoa zehaztu dugu. Hona hemen bildu ditugun aditz-lokuzioetako osagaietariko batzuk: (*ageri, aiherr, balio, behar, bizi, falta...*). Hala ere, esan beharra dago *lo egin, negar egin...* modukoak ez ditugula aditz-kate gisa ezagutzen. Honelakoetan, aurreko osagaiari absolutibo mugagabe analisia eman diogu; adibidez, *lo* izen absolutibo mugagabe legez analizatuko dugu *lo egin* moduko aditz konposatueta azaltzen zaigunean.

4.2. Sintagmak

Ezagutzen diren sintagmak osagai kopuruaren arabera sailkatu ditugu:

4.2.1. *Osagai bakarrekoak*: ondoko adibideetan azpimarratuta ageri diren izenez (*Pellok egin du. Amak esan dio*), izenordainaz (*Nik zalantza egin dut*) edo aditzondoaz (*Atzo ez zen lanera etorri*) bakarrik osatutako sintagmak. Eta izena elidituta dagoenean, sintagma osa dezaketen gainerako kategoriak bakarrik agertzen direnean (*Honekin ezin da urrutira joan. Gazteek dakite ongien hori*). Hauek, funtzio sintaktiko nagusia daraman osagai bakarrez osatutako sintagmak ditugu. Hauek guztiak, esan bezala, %SINT modura markatuko dira analizatzaile sintaktikoan jarraian:

4.2.2. *Osagai bat baino gehiagokotan*: izenaren eskuinera zein ezkerrean ager daitezkeen izenlagun, adjektibo eta zenbatzailez osatutako sintagmak (*bi aldiz; gezur galanta; bere emaztearekin; emakume zahar bat; aldizkari baten zenbaki bat; silize kristaldu bat; bere ahaleginaren ondorio gaiztoak...*). Zenbait postposizio konplexu (*Borroka harrigarri baten ondoren*). Sintagma koordinatu batzuk (*Haren lehengo aurpegi argi eta ederra. Ogi eta ardoa*).

Esan bezala, analizatzaile sintaktikoan %SIH eta %SIB etiketek adieraziko dituzte sintagma hauen hasierak eta bukaerak.

5. Euskararako sintaxi partzialaren ebaluazioa

Laugarren atalean deskribatu ditugun gramatika horiek ebaluatzeko, 24.000 hitzez osatutako lagina¹⁷ baliatu dugu. Analizatzaile sintaktikoak edo *parser*-ak aztertutako lagin hau eskuz markatu denarekin alderatu da Xiraba izeneko ebaluazio-tresnaren bitartez, ondoko grafikoan ikusten den moduan:

27569	"<ifarraldetik>"	27569	"<ifarraldetik>"
PARSER	<Correct!> "iparralde" "ifarralde" IZE ARR DEK NUMS MUGM DEK ABL @ADLG S:2154 %SINT	ESKUZ	<Correct!> "iparralde" "ifarralde" IZE ARR DEK NUMS MUGM DEK ABL @ADLG S:256 %SINT
27570	"<hegoalderantz>"	27570	"<hegoalderantz>"
PARSER	<Correct!> "hegoalde" IZE ARR DEK NUMS MUGM DEK ABZ @ADLG S:1939 %SINT	ESKUZ	<Correct!> "hegoalde" IZE ARR DEK NUMS MUGM DEK ABZ @ADLG S:256 %SINT

Ebaluazio hau bi ezaugarri erreparatuz egin da:

- 1) Doitasuna: analizatzaile sintaktikoak esleituriko aditz-kate nahiz sintagma-kate etiketak ongi esleituta dauden.
- 2) Estaldura: aditz-kate eta sintagma diren guztietatik zenbat markatu dituen analizatzaileak.

Oro har emaitza onak lortu dira ebaluatu diren aditz-kate eta sintagmetan:

	<i>Doitasuna</i>	<i>Estaldura</i>
%SINT	% 94.7	% 80.3
%SIH %SIB	% 84	% 91.1

Taula 1.-Sintagmen ebaluazioaren emaitzak.

Osagai bat baino gehiagoko sintagmetan, sintagma koordinatuak dira arazo-iturri. Batetik, zenbait koordinazio-egitura ez dira harrapatzen, eta, bestetik, harrapatzen diren batzuk oker daude. Hori dela eta, egitura horiek ezagutzeko erregelak etorkizunean aztertu beharko ditugu.

¹⁷ Adibide hauek *Euskal Gramatika Lehen Urratsak-II* (1987) liburutik eta Arriolaren tesi-lanetik (2000) hartuak izan dira.

6. Ondorioak

Testu libreen analisi sintaktiko partziala egiteko urratsa egin dugu kateen gramatikak landuz. Gramatika horiek aplikatzeko CG formalismoa baliatu dugu, eta emaitzek erakusten duten bezala bide egokia jorratu dugu.

Etorkizuneko lanen artean dugu hiru osagai baino gehiagoz osatutako aditz-kateak ezagutzen ahalegintzea (... zeinari ezin izan baitzitzaion ebazpen materialistarik eman), esaldi nagusiko aditz-katean mendeko perpausari dagokion aditz-katea tartekatzen diren egiturak aztertzea (... eta ezin da ondorio honetara **ailegatzan ez den** doktrina erlijiosorik bururatu), eta sintagma mailako koordinazio-egiturak harrapatzea (Praktikak, esperientzia objektiboak eta natura biziaren behaketak osotzen dute), (Ulermen zein azterketarentzat...).

Azkenik, aztertu ditugun kate horien analisi sintaktikotik areago joan nahi izanez gero, *parsing* osoagoa beharko genuke. Horri erantzuteko, irtenbideetariko bat dependentzia-gramatiken bidea jarraitzea da. Ildo horretatik egiten dute lan Järvinen eta Tapanainen-ek (1997). Modu berean, euskararako garatzen ari garen *parsing* osoago batean, dependentzia-eredua lantzen jarraitzen dugu, batetik, buru/mendeko harremanean dauden elementuak esplizituki adieraziz, eta, bestetik, erlazio sintaktikoak gehiago zehaztuz (Aranzabe et al, 2003). Horrela analisi sintaktiko osoagoa izango dugu, eta esaldiaren egitura sintaktikoa adierazten duten analisisien zuhaitzak lortuko ditugu. Kontuan izan behar da lan honetan ere lagunduko lukeela aditzei buruzko azpikategoriazioa zehazteak; hala sintagma edota aditz-kateen analisisitik esaldi konplexuagoen analisisira jauzia egiteko aukera izango baikenuke.

7. Aipamenak

- Abney S. 1991. "Parsing By Chunks". In Robert Berwick, Steven Abney & Carol Tenny (eds.), *Principle-Based Parsing*. Kluwer Academic Publishers: Dordrecht. 257-278.
- Abney S. 1997. "Part-of-speech tagging and partial parsing". In *Corpus-Based Methods in Language and Speech Processing*. Steve Young and Gerrit Bloothoof (Eds.). Kluwer Academic Publishers, Dordrecht. 118-136.

- Aduriz I. 2000. *EUSMG: Morfoloġiatik sintaxira Murriztapen Gramatika erabiliz*. Doktoretza-tesia, Euskal Filologia Saila, Euskal Herriko Unibertsitatea.
- Aduriz, I., J.M., Arriola & A., Díaz de Ilarraza. 2003. "Desanbiguazio morfoloġikoa, azterketa sintaktikoaren lehen urratsak eta aplikazioak Murriztapen Gramatikaren eredu konputazionala jarraituz". J. M. Makatzaga & B. Oyarçabal eds *Euskal Gramatikari eta Literaturari buruzko ikerketak XXI. mendearen atarian*. Bilbo. Euskaltzaindia. IKER 14 (1). 1-36 or.
- Aduriz, I. & A. Diaz de Ilarraza. 2004. "Morphosyntactic Disambiguation and Shallow Parsing in Computational Processing of Basque". B. Oyarçabal (arg.), *Inquiries into the lexicon-syntax relation in Basque*. Supplements of ASJU XLVI, 1-22 or.
- Ait-Moktar S. & J.P. Chanod 1997. "Incremental Finite-State Parsing". In the *Proceedings of the 5th Conference of Applied Natural Language Processing*. 72-79 or.
- Alegria I., X. Artola & K. Sarasola. 1996. "Automatic morphological analysis of Basque". *Literary & Linguistic Computing* Vol. 11, No. 4. Oxford: Oxford University Press. 193-203.
- Aranzabe M.J., J.M., Arriola, A. Atutxa, I. Balza & L. Uria. 2003. *Guía para la anotación sintáctica manual de Eus3LB (corpus del euskera anotado a nivel sintáctico, semántico y pragmático)*. UPV/EHU/LSI/TR-13.
- Arriola J.M. 2000. *Euskal hiztegia-ren azterketa eta egituratzea ezagutza lexikalaren eskuratze automatikoari begira*. Aditz-adibideen analisisa Murriztapen Gramatika baliatuz, azpikategorizazioaren bidean. Doktoretza-tesia, Euskal Filologia Saila, Euskal Herriko Unibertsitatea.
- Daelemans W., S. Buchholz & Veenstra. 1999. "Memory-based shallow parsing". In the *Proceeding of CoNLL*, Bergen, Norway. 53-60 or.
- Jurafsky D. & H.J. Martin. 2000. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Prentice Hall, Upper Saddle River.

- Karlsson F., A., Voutilainen, J. Heikkilä & A. Anttila. 1995. *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Berlin: Mouton de Gruyter.
- Martí M. A. & J. Llisterri. 2002. *Tratamiento del lenguaje natural*. Barcelona: Edicions de la Universitat de Barcelona.
- Ramshaw L. A. & M.P. Marcus. 1995. "Text chunking using transformation-based learning". In *Workshop on Very Large Corpora*, 82-94.
- Sang E.T.K., Buchholz S. 2000. "Introduction to the CoNLL-2000 shared task Chunking". In *Proceedings of the Fourth Conference on Computational Natural Language Learning (CoNLL) and the Second Learning Language in Logic Workshop (LLL)*, Lisboa, Portugal, 127-132.
- Tapanainen P. & T., Järvinen. 1997. "A non-projective dependency parser". In *Proceedings of ANLP-97, ACL*. 64-71 or.
- Voutilainen A., J. Heikkilä, A., Anttila. 1992. *Constraint grammar of English A Performance-Oriented Introduction*. Publications nr. 21, Dept., of General Linguistics, University of Helsinki.
- Voutilainen A. 1994. *Designing a Parsing Grammar*. Publications of the Department of General Linguistics, 22. University of Helsinki.

8. Glosarioa

(1) Adibidean erabiltzen diren laburtzapenen azalpena alfabetikoki ordenatua

%ADIKATBU: aditz-kate edo sintagma jarraitu bateko azken elementuak darama;

%ADIKATHAS: aditz-kate edo sintagma jarraitu bateko lehenengo elementuak darama;

%SIB: sintagmaren bukaera;

%SIH: sintagmaren hasiera;

%SINT: hitz bakarreko sintagma;

@+JADLAG: aditz laguntzailea;

@+JADLAG_MP: aditz laguntzaile mendekoa;

@+JADNAG: aditz nagusi jokatua;

@+JADNAG_MP: aditz nagusi jokatu mendekoa;
@<IA: eskuineko adjektiboa;
@<IZLG: eskuineko izenlaguna;
@ADLG: adizlaguna;
@ID>: ezkerreko determinatzailea;
@IZLG>: ezkerreko izenlaguna;
@-JADNAG: aditz nagusia ezjokatua;
@-JADNAG_MP_OBJ: objektu funtzioa duen aditz nagusi ezjokatu mendekoa;
@KM>: kasua daraman formaren modifikatzailea;
@OBJ: objektua;
@PJ: perpaus juntadura;
@PRED: predikatiboa;
@SUBJ: subjektua;
A1: DA / DU: izan / *edun-en orainaldia (indikatiobozko orainaldia);
ABL: ablatiboa;
ABS: absolutiboa;
ABZ: hurbiltze-adlatiboa;
ADI: aditza;
ADIZE: aditz izena;
ADJ: adjektiboa;
ADL: aditz laguntzailea;
ADOIN: aditzoina;
ADT: aditz trinkoa;
AMM: aditz-mota morfema;
AORG: -a organikoa daukan hitzak darama;
ARR: (izen) arrunta;
ASP: aspektu-morfema;
DEK: deklinabidea;
DET: determinatzailea;
DZG: determinatzaile zehaztugabea;
EMEN: emendiozko juntagailua;
EZBU: aspektu-marka ezburutua;
GEN: genitibo edutezkoa;
HAS_MAI: letra larriz hasitako hitza;
INE: inesiboa;
IOR: izenordaina;

IZE: izena;

IZO: izenondoa;

JNT: juntagailua;

KAUS: kausala;

LOT: loturazko elementua: lokailu eta juntagailuak hartzen ditu bere baitan;

MEN: mendekoa;

MG: mugagabea;

MUGM = M: mugatua;

NOTDEK: deklinabidetik pasatzen ez den aditza;

NR(_)HU: NOR paradigmako HURA;

NUMP = PL = P: numero plurala;

NUMS = S: numero singularra;

PERARR: izenordain pertsonal arrunta;

PUNT_PUNT: puntua;

SIN: aditz sinplea.