

---

# **Recursos sintácticos para el euskara**

---

---

# Trabajos realizados

- HPSG simple para euskara (1998)
- Gramática de unificación PATR: PATRIXA (2001)
- Analizadores sintácticos de dependencias
  - EDGK, Dependencias, Constraint grammar (2008)
  - MaltIXA, Dependencias, Estadístico, MaltParser (2012)
- Otros recursos

---

# HPSG simple para euskara (1998)

- Herramienta “de juguete”
  - Oración básica (SOV)
  - Sintagma nominal básico: N + ADJ + CASO
- ¿Cuál es la unidad de análisis?
  - Ejemplo: *mendiko umea*
    - “mendi”/monte + “ko”/de + “ume”/niño + “a”/el
  - Palabra? “mendiko umea”
  - Morfema? “mendi ko ume a”

Gojenola K.,

Guneak zuzendutako egitura sintagmatikoen gramatika (HPSG) eta Euskararako aplikazioa  
UPV/EHU/LSI/TR 5-98.

---

# HPSG simple para euskara (1998)

- ¿Cuál es el núcleo del sintagma nominal?  
Ejemplo: etxea/**la casa** (“etxe”/**casa** + “a”/**la**)
  - El núcleo indica los complementos (subcategorización)
- ¿Cuál es el núcleo del sintagma verbal?
  - Ejemplo: ekarri dute /**(lo) han traído**  
“ekarri”/**traer** + “dute”/**SUBJ:3P-OBJ-3S**
  - Los dos aportan información importante:
    - Verbo principal: subcategorización
    - Verbo auxiliar: concordancia
  - Solución Abaitua (1988, LFG): el núcleo es la unión de los 2 verbos

---

# Trabajos realizados

- HPSG simple para euskara (1998)
- **Gramática de unificación PATR: PATRIXA (2001)**
- Analizadores sintácticos de dependencias
  - EDGK, Dependencias, Constraint grammar (2008)
  - MaltIXA, Dependencias, Estadístico, MaltParser (2012)
- Otros recursos

---

# Gramática de unificación PATR: PATRIXA (2004)

- PATR: formalismo básico de unificación
  - Basado en ecuaciones
  - No hay principios generales
- Gramática PATR para el euskara:
  - ~100 reglas de complejidad media/alta
  - Fenómenos tratados:
    - Oraciones simples
    - Oraciones subordinadas
    - Sintagmas nominales y postposicionales

# Gramática de unificación PATR: PATRixa

## ■ Ejemplo de regla (N/ADJ + Caso-Núm):

```
rule(r_lehen_knmdek_arrunta, X0 ---> [X1, X2]@[
    m(1, edo [eta [X1/ezaug/kat      badago [ize, ior, adj, det, eli, adb, snb, bst], % tut+ik
        % adb gaurko+a bezalakoak onartzeko
    edo [X1/ezaug/kas ez [gen],
        eta [X1/ezaug/kas badago [gen],
            X2/'lema-nagusia'/twol ez  ["gandik", "gan", "gana", "gatik", "ganantz", "ganaino"]
        ]
    ] % en+gatik/gana/... ez tratatzeko hemen (r_gen_atzizki-k egingo du)
    ],
    eta [X1/ezaug/kat badago [adl, adt],
        X2/'lema-nagusia'/twol ez  ["gandik", "gan", "gana", "gatik", "ganantz", "ganaino"],
        % dutenen+gatik ez egiteko, hori r_gen_atzizki-k egiten du
    edo [X1/ezaug/kas badago [gen, gel],
        % dakienaren+a. dakite+n+ko + 0, dakite+n+ko + a
        X1/ezaug/erl badago [ert]
        % adizkiak elipsirik ez dagoenean: dakien+a ]
    ],
    m(2, edo [X1/ezaug/azp      ez  [izb, lib],
        % det-dzh guztiak erregela honekin tratatuko dira
        X1/'lema-nagusia'/sarrera/'sarrera-gakoa'/sarrera <=>  "bat",
        % beraz, hau soberan dago det-dzh delako, eta aurrekoa
        % beteko duelako
    eta [X1/ezaug/azp      badago [izb, lib],
        X1/ezaug/kas      badago [gen, gel]  % izan daiteke plu+ ("gabonenek") edo plu- ("peiorenek")
        % *keparen+ek, hemen ez dago izen berezien komuntadura
    ]
]
```

---

# Gramática de unificación PATR: PATRIXA

- Conclusión PATRIXA:
  - Chunker aceptable
    - Buena cobertura y precisión
    - Combina la información de los diversos constituyentes mostrando un resultado elaborado
  - Explosión de ambigüedad al analizar oraciones complejas



---

# Trabajos realizados

- HPSG simple para euskara (1998)
- Gramática de unificación PATR: PATRixa (2001)
- **Analizadores sintácticos de dependencias**
  - EDGK, Dependencias, Constraint grammar (2008)
  - MaltIXA, Dependencias, Estadístico, MaltParser (2012)
- Otros recursos

---

# Analizadores sintácticos de dependencias

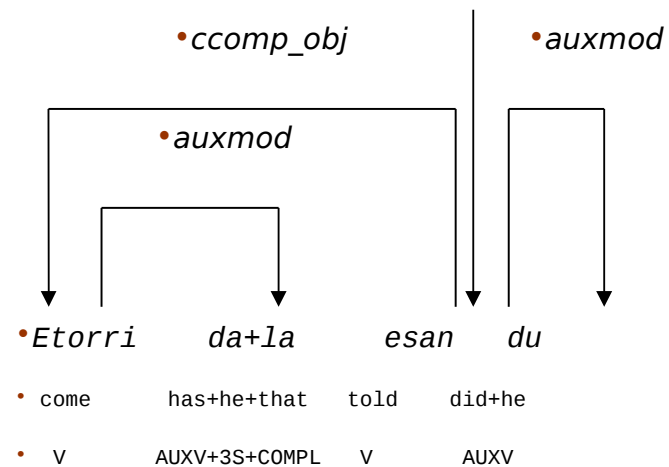
- EDGK: etiquetador de dependencias basado en reglas (Constraint Grammar)
  - Sintaxis parcial
- MaltIXA: analizador sintáctico estadístico basado en dependencias:
  - Treebank de 150.000 palabras para entrenamiento
  - MaltParser, MST, Bohnet

---

# Analizadores sintácticos de dependencias

- Trabajos realizados:
  - Optimización de features
  - Transformación de árboles
  - Combinación de analizadores:
    - Stacking
    - Voting
- Resultado actual: ~ 80% Labelled Attachment Score

# Analizadores sintácticos de dependencias



•Figure 3. Dependency tree for the sentence *Etorri dela esan du* (He told that he would come).

---

# Trabajos realizados

- HPSG simple para euskara (1998)
- Gramática de unificación PATR: PATRixa (2001)
- Analizadores sintácticos de dependencias
  - EDGK, Dependencias, Constraint grammar (2008)
  - MaltIXA, Dependencias, Estadístico, MaltParser (2012)
- **Otros recursos**

---

# Otros recursos

<http://ixa.si.ehu.es/Ixa/Produktuak>

- EPEC Treebank (155.000 palabras marcadas con dependencias, <http://ixa.si.ehu.es/Ixa/resources/Treebank>)
- EDBL Base de datos del léxico del euskara. Información morfológica.
- Morfeus: Analizador morfológico
- Eustagger: Lematizador
- Información sobre subcategorización verbal