

# Itzulpen automatikoa IXA taldean

I. Alegria, A. Díaz de Ilarraza, G. Labaka, M. Lersundi,  
A. Mayor eta **K. Sarasola**



**IXA Taldea. Informatika Fakultatea  
Euskal Herriko Unibertsitatea**



# Ekinbideak

- **Matxin**
  - Erregeletan eta transferentzian oinarritutako prototipoa
  - es->eu eta en->eu
- **OpenTrad**
  - es->eu, es<->ca, es<->ga
- **Itzulpen-memoriak sortu eta erabiltzeko tresnak**
- **Ikerlerroak**



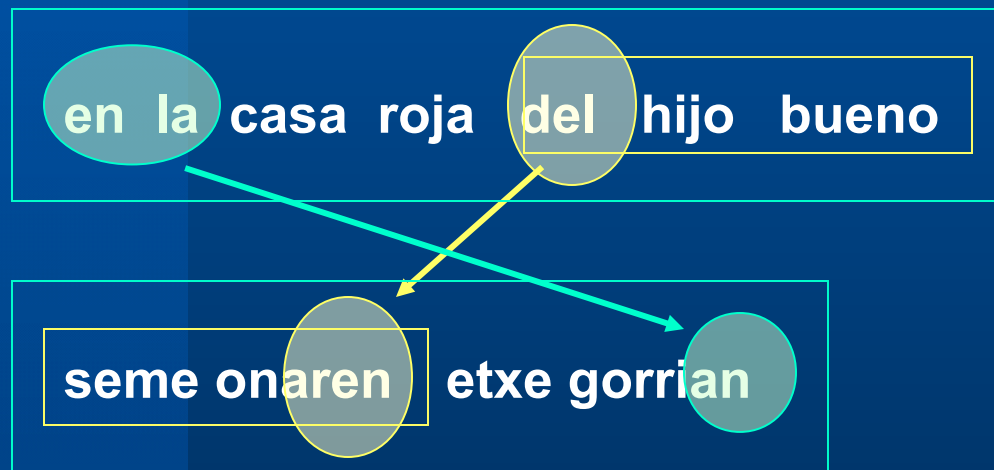
# IArako prototipoa: Matxin

- Eleanitza  
ingelesa-euskara eta espainiera-euskara
- Transferentzian oinarritua
- Estaldura zabaleko baliabide eta tresna  
linguistikoen berrerabilpena  
(EDBL, hiztegi elebidun elektronikoak, Morfeus,  
sintaxi-gramatika, ...)



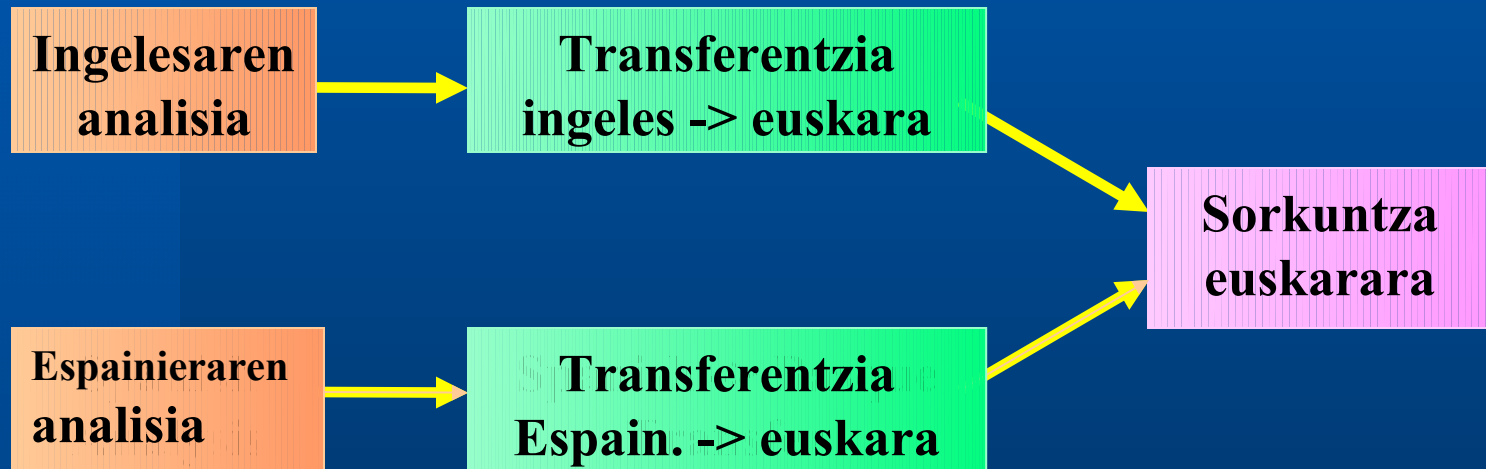
# Euskarara itzultzeko zailtasuna

- Sintagmen egitura eta esaldiaren oso bestelakoak dira ingelesez eta espainieraz



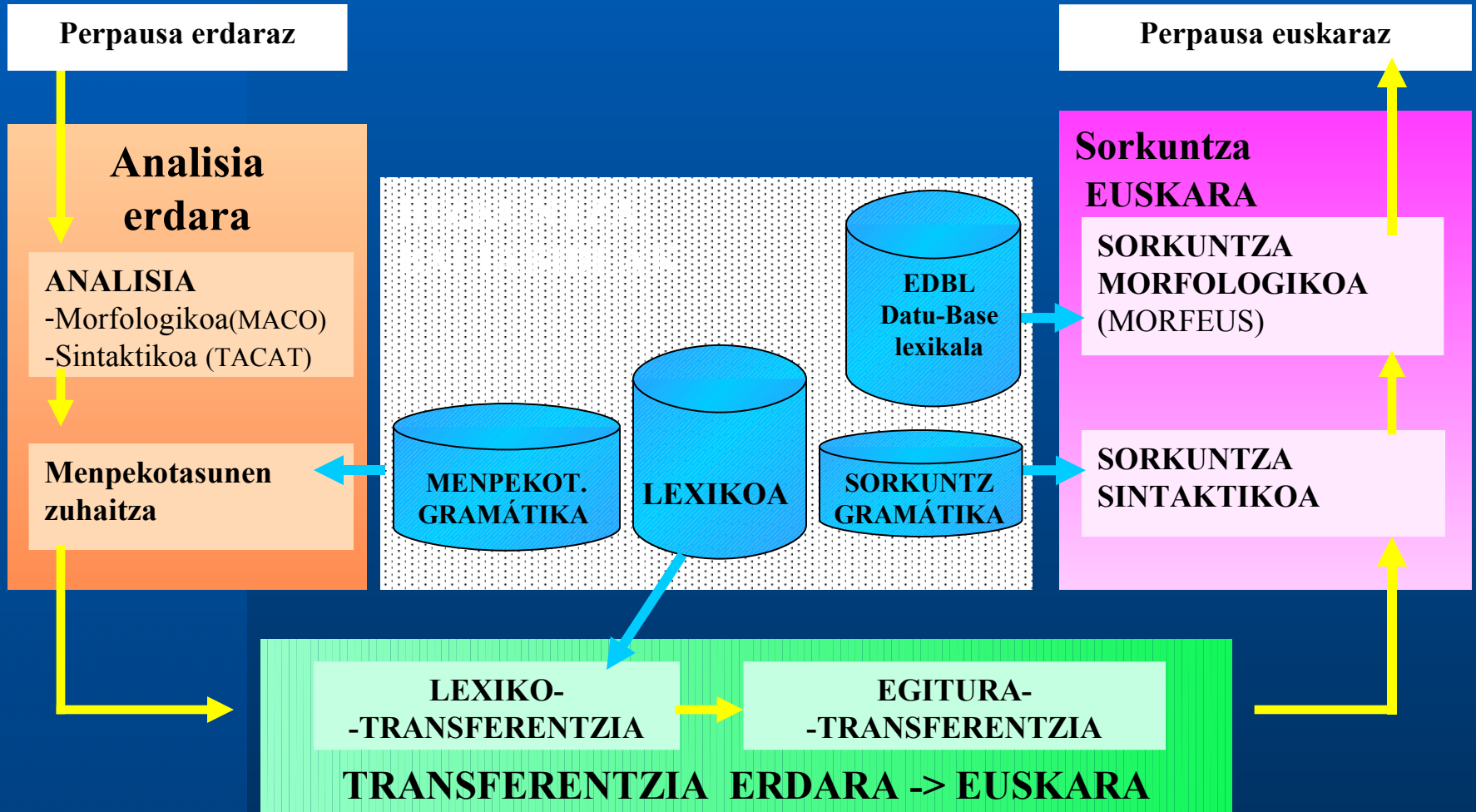


# Sistemaren arkitektura (I)





# Sistemaren arkitektura (II)





# Itzulpenaren adibide bat

“dentro de las casas construidas en aquel bonito pueblo ”

"herri polit hartan eraikitako etxeen barruan"

**Analisis  
ERDARA**

**Sorkuntza  
EUSKARA**

**TRANSFERENTZIA  
ERDARA->EUSKARA**



# Analisisa (I)

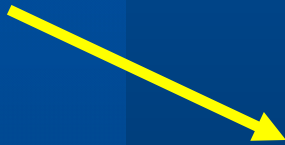
## Analisi morfologiko eta sintaktikoa

“dentro de las casas construidas en aquel bonito pueblo ”



### ANALISIA

-Morfologiko(MACO)  
-Sintaktiko (TACAT)



Forma	Lema	Inf. morfologikoa
dentro_de	dentro_de	[sps00]
las	la	[tdfp0]
casas	casa	[ncfp000]
construidas	construir	[vmpp0pf]
en	en	[sps00]
aquel	aquel	[dd3ms00]
bonito	bonito	[aq0ms00]
pueblo	pueblo	[ncms000]





# Analisisa

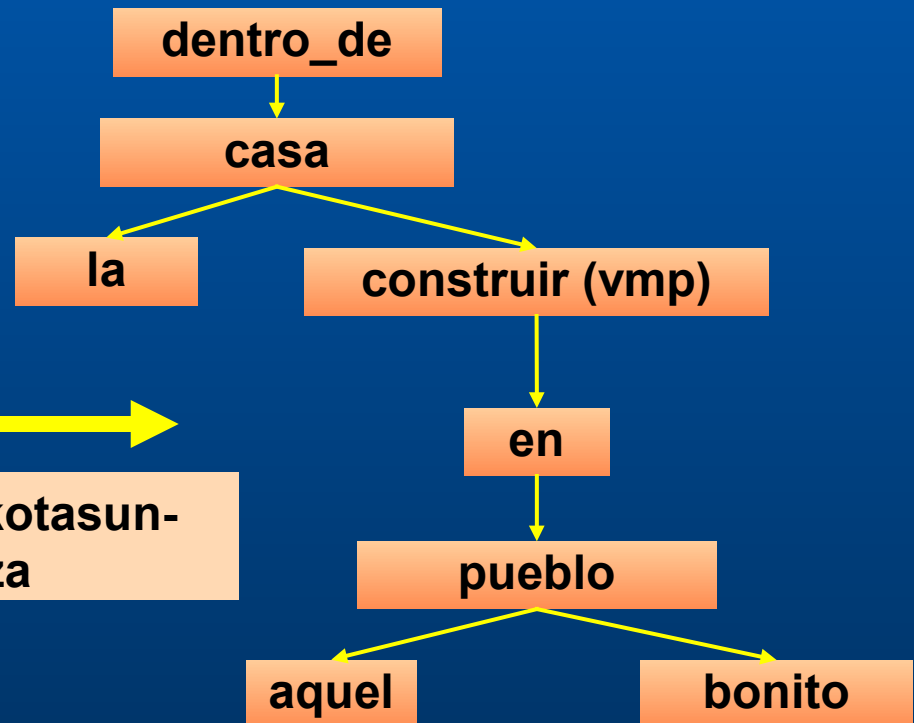
## Menpekotasun-zuhaitza lortzea

“dentro de las casas construidas en aquel bonito pueblo ”

Forma	Lema	Inf. morfológica
dentro_de	dentro_de	[sps00]
las	la	[tdfp0]
casas	casa	[ncfp000]
construidas	construir	[vmpp0pf]
en	en	[sps00]
aquel	aquel	[dd3ms00]
bonito	bonito	[aq0ms00]
pueblo	pueblo	[ncms000]



Menpekotasun-zuhaitza

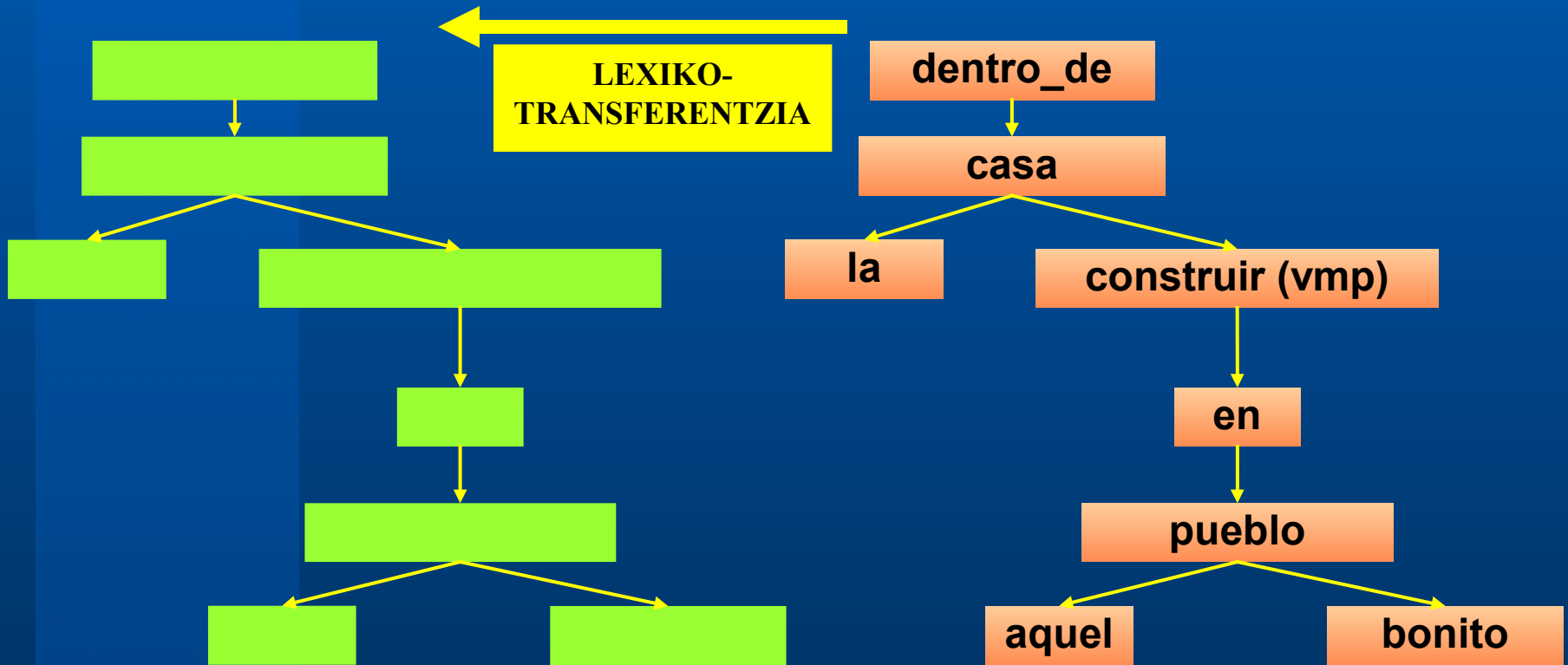




# Transferentzia

## Lexiko-transferentzia (I)

“dentro de las casas construidas en aquel bonito pueblo ”

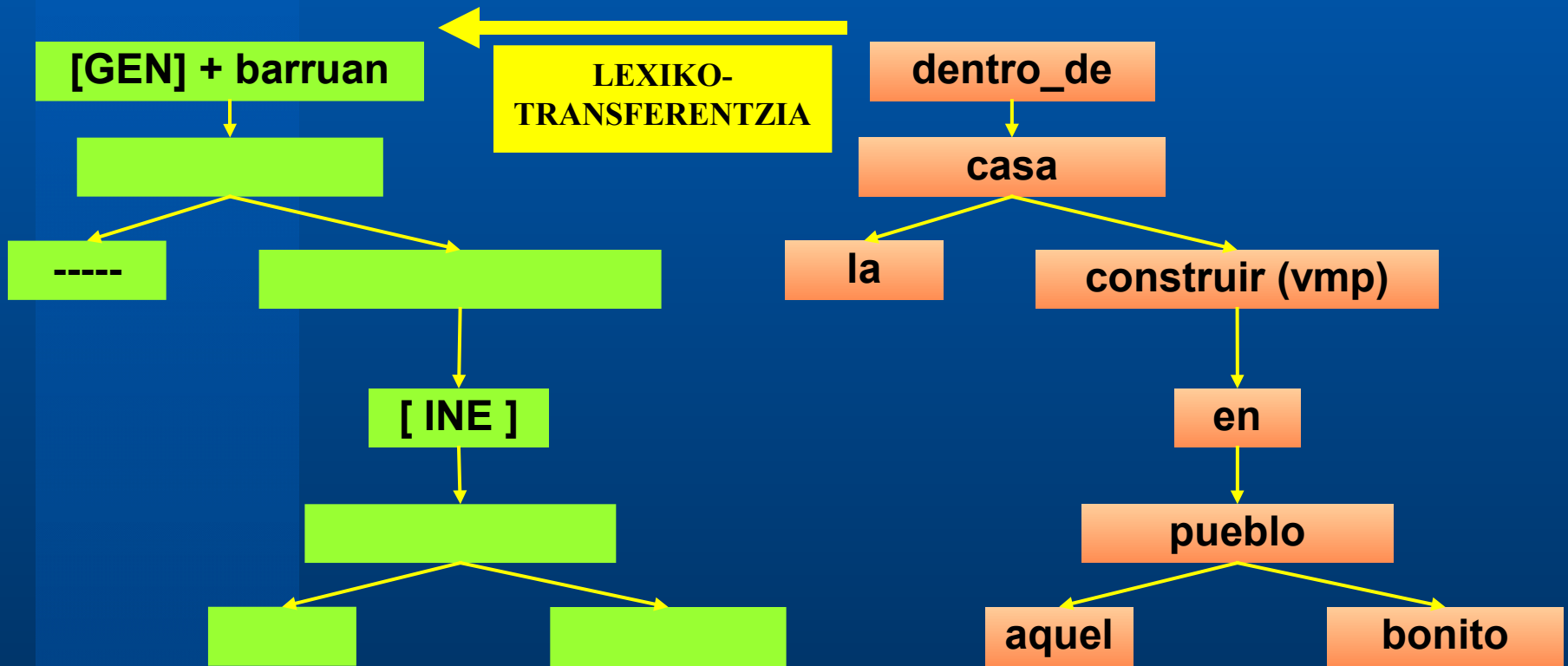




# Transferentzia

## Lexiko-transferentzia (II)

“dentro de las casas construidas en aquel bonito pueblo ”

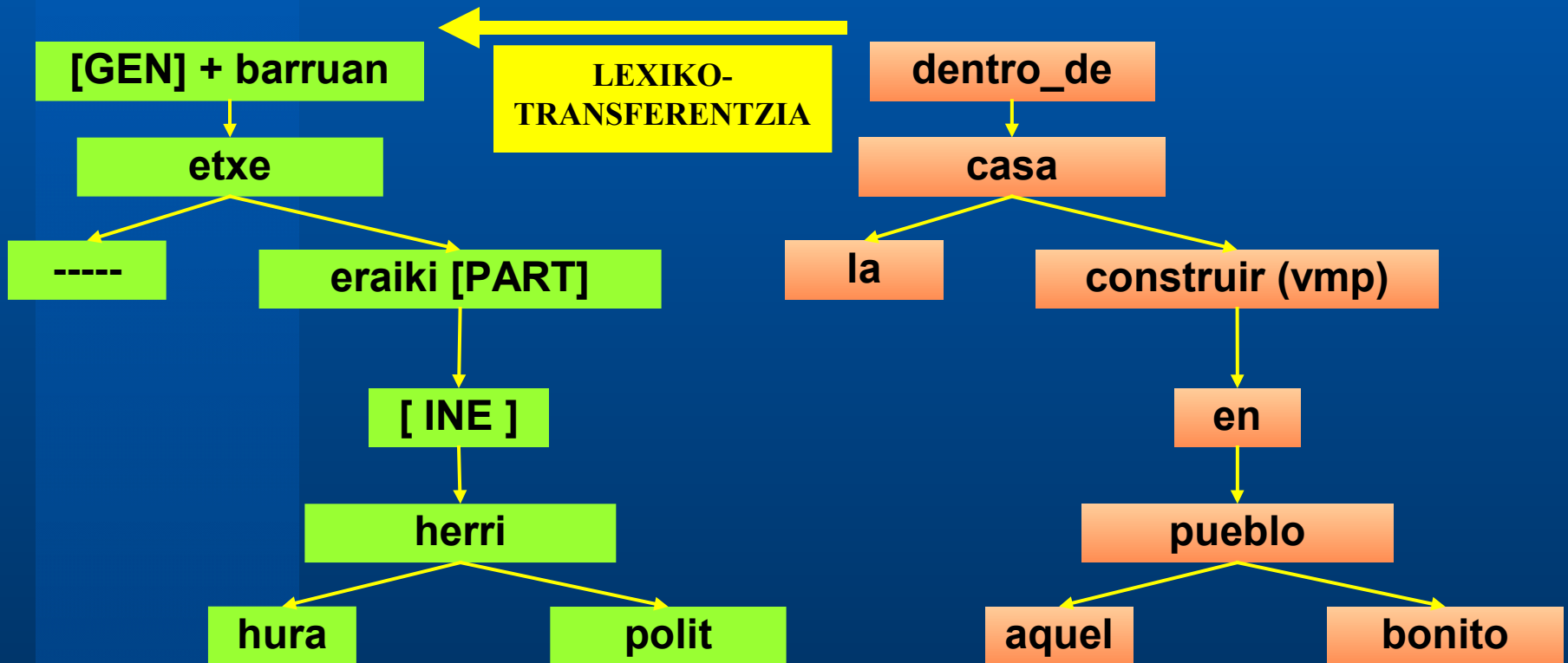




# Transferentzia

## Lexiko-transferentzia (III)

“dentro de las casas construidas en aquel bonito pueblo ”

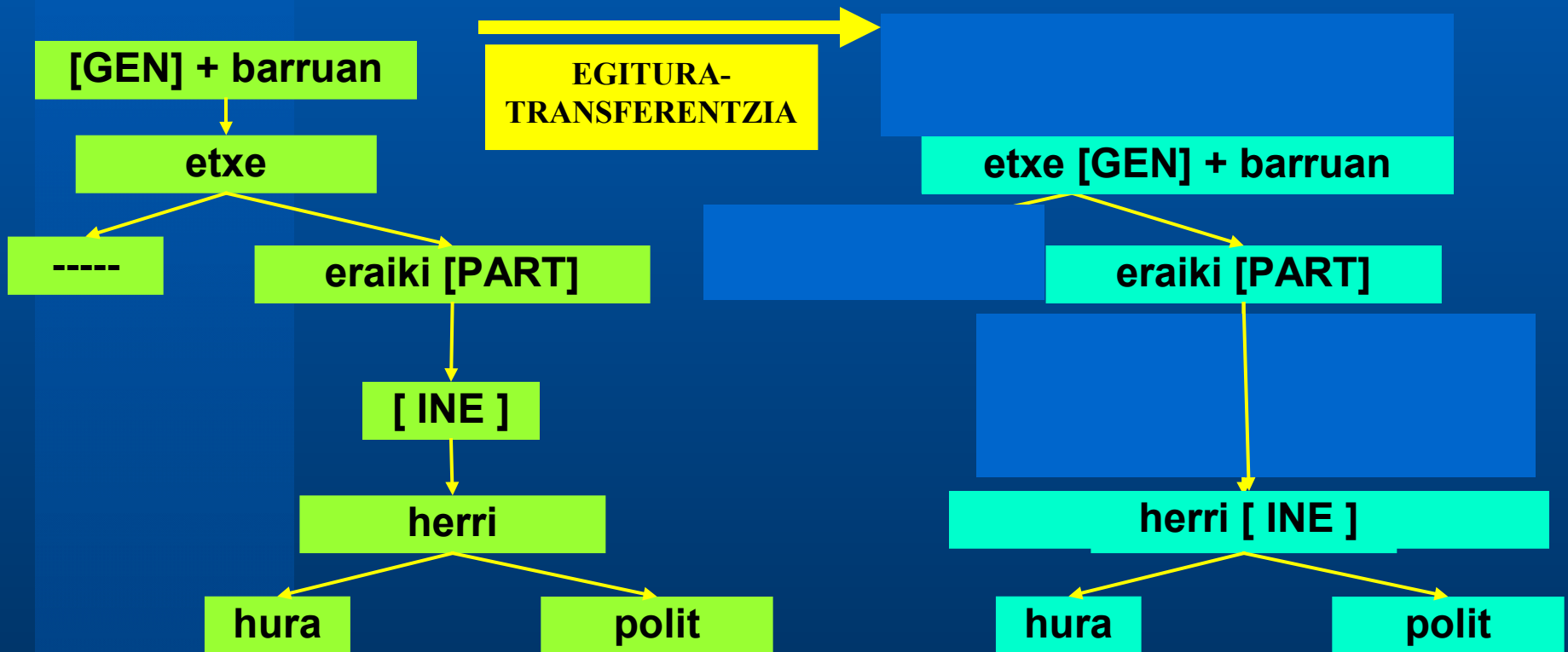




# Transferentzia

## Egitura-transferentzia

“dentro de las casas construidas en aquel bonito pueblo ”





# Sorkuntza

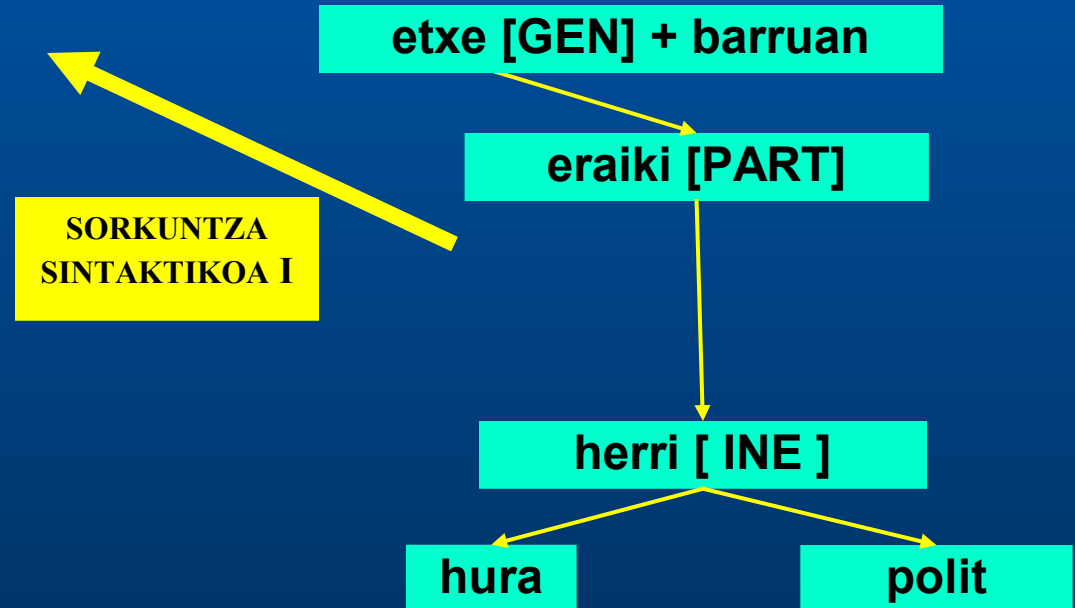
## Sorkuntza sintaktikoa (I)

“dentro de las casas construidas en aquel bonito pueblo ”

herri [ INE ]  
pueblo (en)

Hitzen berrordenazioa  
Gramatika bat erabiliz

- 33 erregela bitar
- PATR gramatikaren egokitzapena





# Sorkuntza

## Sorkuntza sintaktikoa (II)

“dentro de las casas construidas en aquel bonito pueblo ”

herri [ INE ]  
pueblo (en)

polit  
bonito

hura  
aquel

etxe [GEN] + barruan

eraiki [PART]

SORKUNTZA  
SINTAKTIKOA I

hura

polit



# Sorkuntza

## Sorkuntza sintaktikoa (III)

“dentro de las casas construidas en aquel bonito pueblo ”

herri [ INE ]  
pueblo (en)

polit  
bonito

hura  
aquel

eraiki [PART]  
construido

etxe [GEN] + barruan

eraiki [PART]

SORKUNTZA  
SINTAKTIKOA I





# Sorkuntza

## Sorkuntza sintaktikoa (IV)

“dentro de las casas construidas en aquel bonito pueblo ”

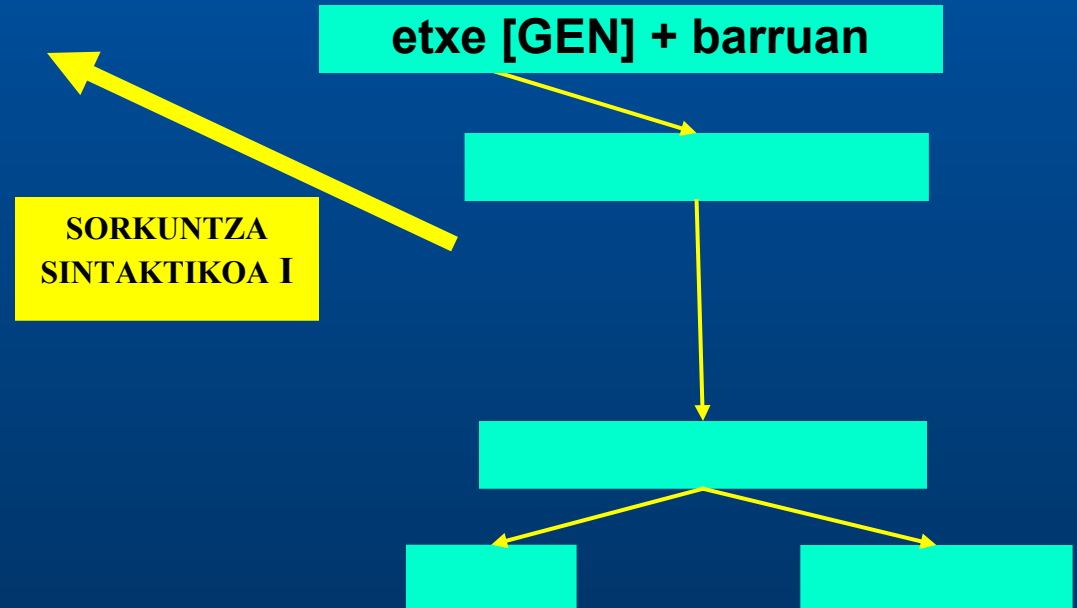
herri [ INE ]  
pueblo (en)

polit  
bonito

hura  
aquel

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro





# Sorkuntza

## Sorkuntza sintaktikoa (V)

“dentro de las casas construidas en aquel bonito pueblo ”

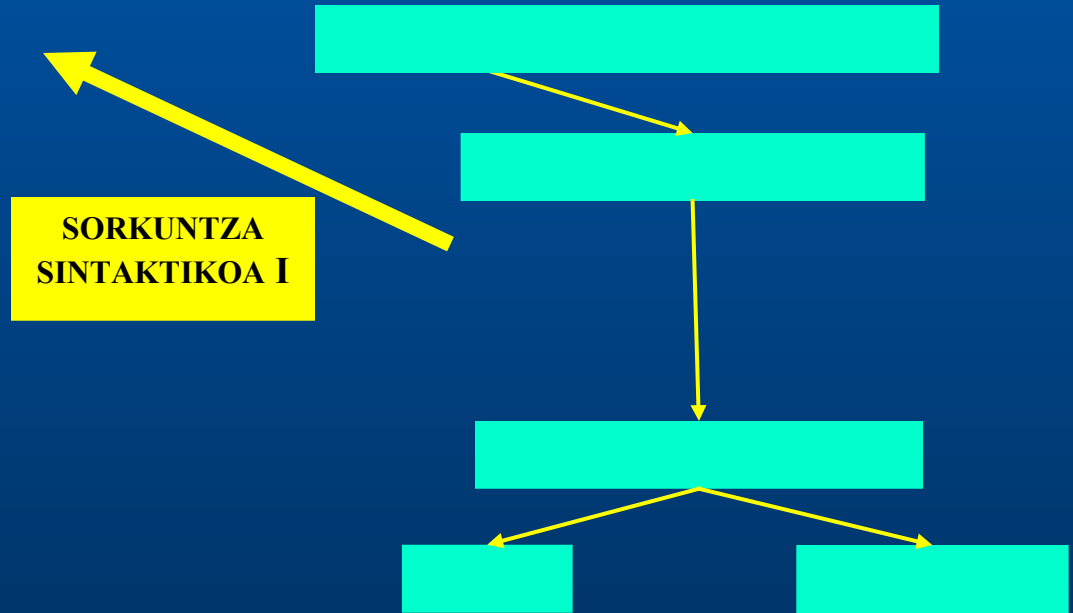
herri [ INE ]  
pueblo (en)

polit  
bonito

hura  
aquel

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro





# Sorkuntza

## Sorkuntza sintaktikoa (V)

“dentro de las casas construidas en aquel bonito pueblo ”

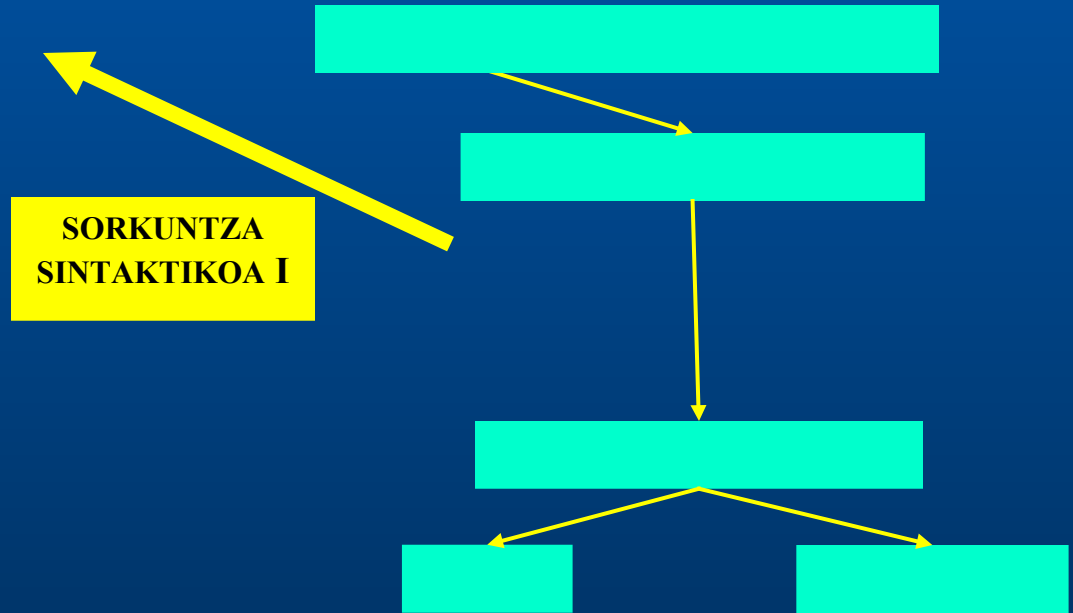
herri [ INE ]  
pueblo (en)

polit  
bonito

hura  
aquel

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro





# Sorkuntza

## Sorkuntza sintaktikoa (VI)

“dentro de las casas construidas en aquel bonito pueblo ”

herri [ INE ]  
pueblo (en)

polit  
bonito

hura  
aquel

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro

SORKUNTZA  
SINTAKTIKOA II

Kasu-markaren  
birkokapena

herri |  
puebl

polit  
bonito

hura [ INE ]  
aque (en)

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro



# Sorkuntza

## Sorkuntza morfologikoa (I)

“dentro de las casas construidas en aquel bonito pueblo ”

hartan  
aquel (en)

SORKUNTZA  
MORFOLOGIKOA

herri  
pueblo

polit  
bonito

hura [INE]  
aquel (en)

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro



# Sorkuntza

## Sorkuntza morfologikoa (II)

“dentro de las casas construidas en aquel bonito pueblo ”

hartan  
aquel (en)

eraikitako  
construido

SORKUNTZA  
MORFOLOGIKOA

herri  
pueblo

polit  
bonito

hura [INE]  
aquel (en)

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro



# Sorkuntza

## Sorkuntza morfologikoa (III)

“dentro de las casas construidas en aquel bonito pueblo ”

hartan  
aquel (en)

eraikitako  
construido

etxeen barruan  
casa (de) +dentro

SORKUNTZA  
MORFOLOGIKOA



herri  
pueblo

polit  
bonito

hura [INE]  
aquel (en)

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro



# Sorkuntza

## Sorkuntza morfologikoa (IV)

“dentro de las casas construidas en aquel bonito pueblo ”

herri  
pueblo

polit  
bonito

hartan  
aquel (en)

eraikitako  
construido

etxeen barruan  
casa (de) +dentro

SORKUNTZA  
MORFOLOGIKOA



herri  
pueblo

polit  
bonito

hura [INE]  
aquel (en)

eraiki [PART]  
construido

etxe [GEN] + barruan  
casa (de) +dentro





# Sorkuntza

## Sorkuntza morfologikoa (V)

“dentro de las casas construidas en aquel bonito pueblo ”

herri  
pueblo

polit  
bonito

hartan  
aquel (en)

eraikitako  
construido

etxeen barruan  
casa (de) +dentro



"herri polit hartan eraikitako etxeen barruan"



# Ekinbideak

- **Matxin**
  - Erregeletan eta transferentzian oinarritutako prototipoa
  - es->eu eta en->eu
- **OpenTrad**
  - es->eu, es<->ca, es<->ga
- **Itzulpen-memoriak sortu eta erabiltzeko tresnak**
- **Ikerlerroak**



# OpenTrad

## Matxin 2004

- Transferentzian oinarritutakoa
- es->eu,

## eta gainera ...

- es<->ca, es<->ga
- Kode irekia eta libre.
- Formatuen estandarizazioa XML lengoaiaren bitartez
- Teknologia: egoera finituak, transduktoreak
  - Ondo dokumentatua eta oso eraginkorra



# OpenTrad - Proiektuko kideak

- Transducens (Alacanteko Unibertsitatea)
- IXA (Euskal Herriko Unibertsitatea)
- TALP (Kataluniako Unibertsitate Politeknikoa)
- Linguistika Informatikoko Mintegia (Vigoko Unibertsitatea)
- Eleka Ingeniaritza Linguistikoa
- Elhuyar Fundazioa
- Galiziako Imaxin Software



# OpenTrad - Emaidzak

- Webgune bat itzulpenak egiteko
- Programaren kode irekia  
(GPL, LGPL edo CC lizentziapean)
- Hizkuntza-baliabideek (lexiko elebiduna...)
  - oinarrizko bat libre izango da
  - bertsio osatuagoak lizentziapean



# OpenTrad - Oinarriak

- **Matxin.** Transferentzia bidezko teknologia
- ***InterNostrum.***
  - ca-es Transducens taldea (Alacanteko Unibertsitatea)
  - dohain erabil daiteke web bidez.
  - *Egoera finituak, transduktoreak*
- ***FreeLing***
  - UPCK garatua
  - Espainieraren eta ingelesaren azterketarako software librea da
  - CVS bidez zabaltzen da. Edonork erabil edo alda dezake
  - MACO, TACAT eta RELAX (Matxin-ekoak) programen bertsio librea
  - Txala: FreeLing + sintaxi-menpekotasunak



# OpenTrad – Osagaiak



# Analisis

“Los amigos han comido manzanas rojas ”



**Analisis**

comido

|

han

amigos

|

los

manzanas

|

rojas

*SUJ*

*OBJ*

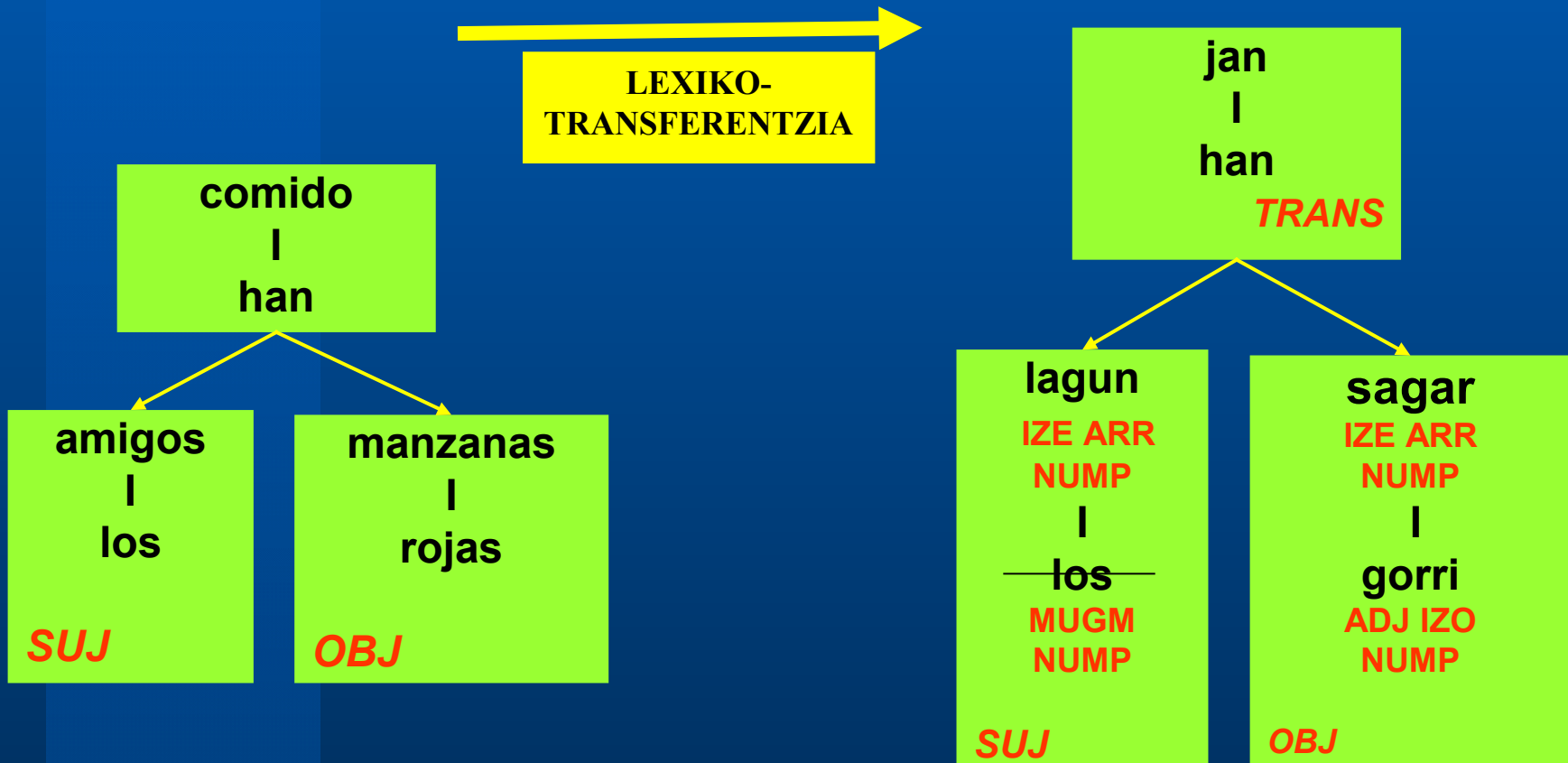




# Transferentzia

## Lexiko-transferentzia

“Los amigos han comido manzanas rojas ”





# Transferentzia

## Egitura-transferentzia (I)

“Los amigos han comido manzanas rojas ”



**EGITURA-  
TRANSFERENTZIA**  
Zatien barruan  
(aditz kateak ez)

jan  
|  
han  
**TRANS**

jan  
|  
han  
**TRANS**

lagun  
IZE ARR  
NUMP  
|  
~~los~~  
~~MUGM~~  
~~NUMP~~  
MUGM  
NUMP  
**SUJ**

sagar  
IZE ARR  
NUMP  
|  
gorri  
ADJ IZO  
NUMP  
**OBJ** **NUMP**

lagun  
IZE ARR  
NUMP  
|  
~~los~~  
MUGM  
NUMP  
**SUJ**

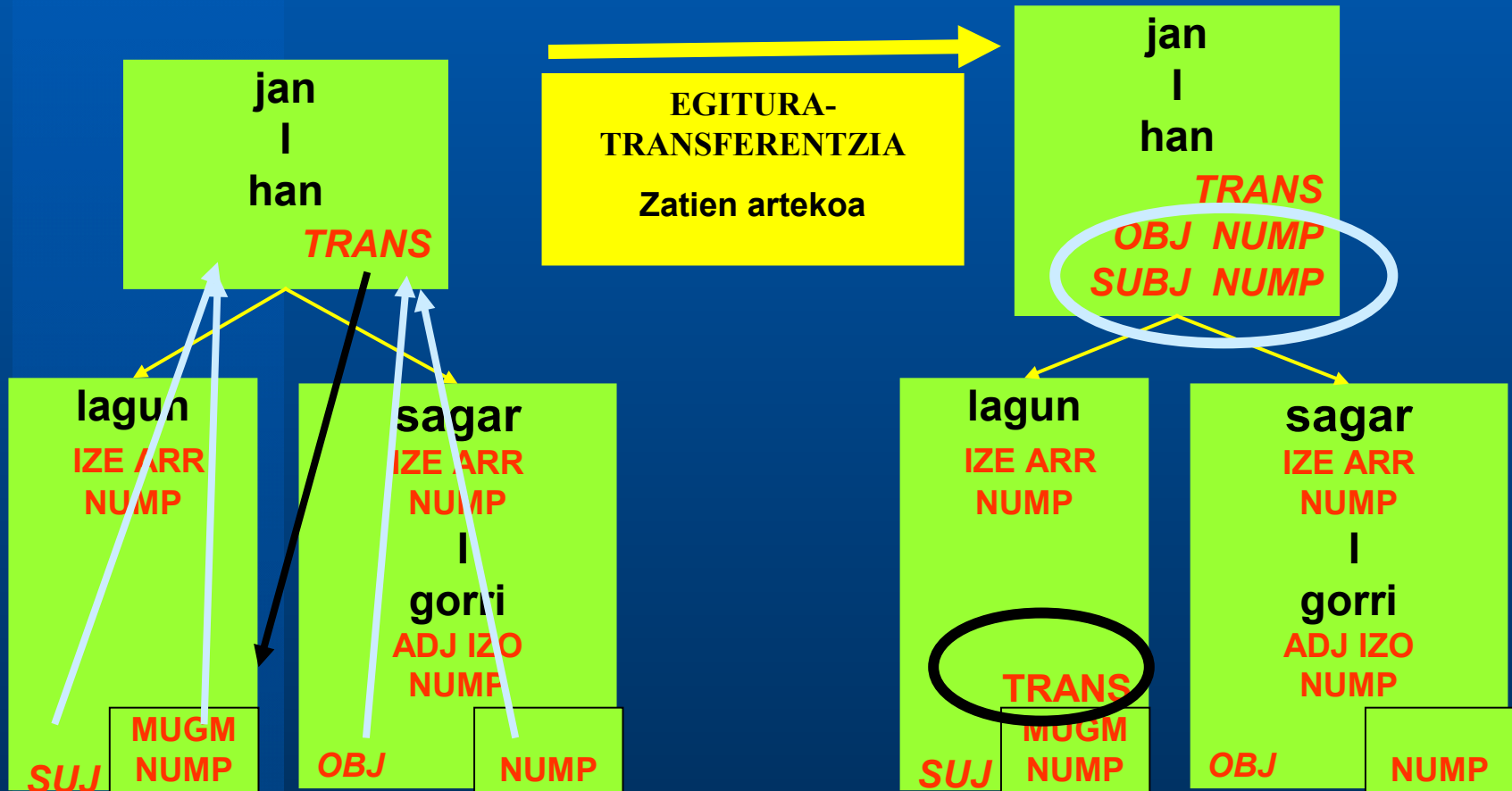
sagar  
IZE ARR  
NUMP  
|  
gorri  
ADJ IZO  
NUMP  
**OBJ**



# Transferentzia

## Egitura-transferentzia (II)

“Los amigos han comido manzanas rojas ”

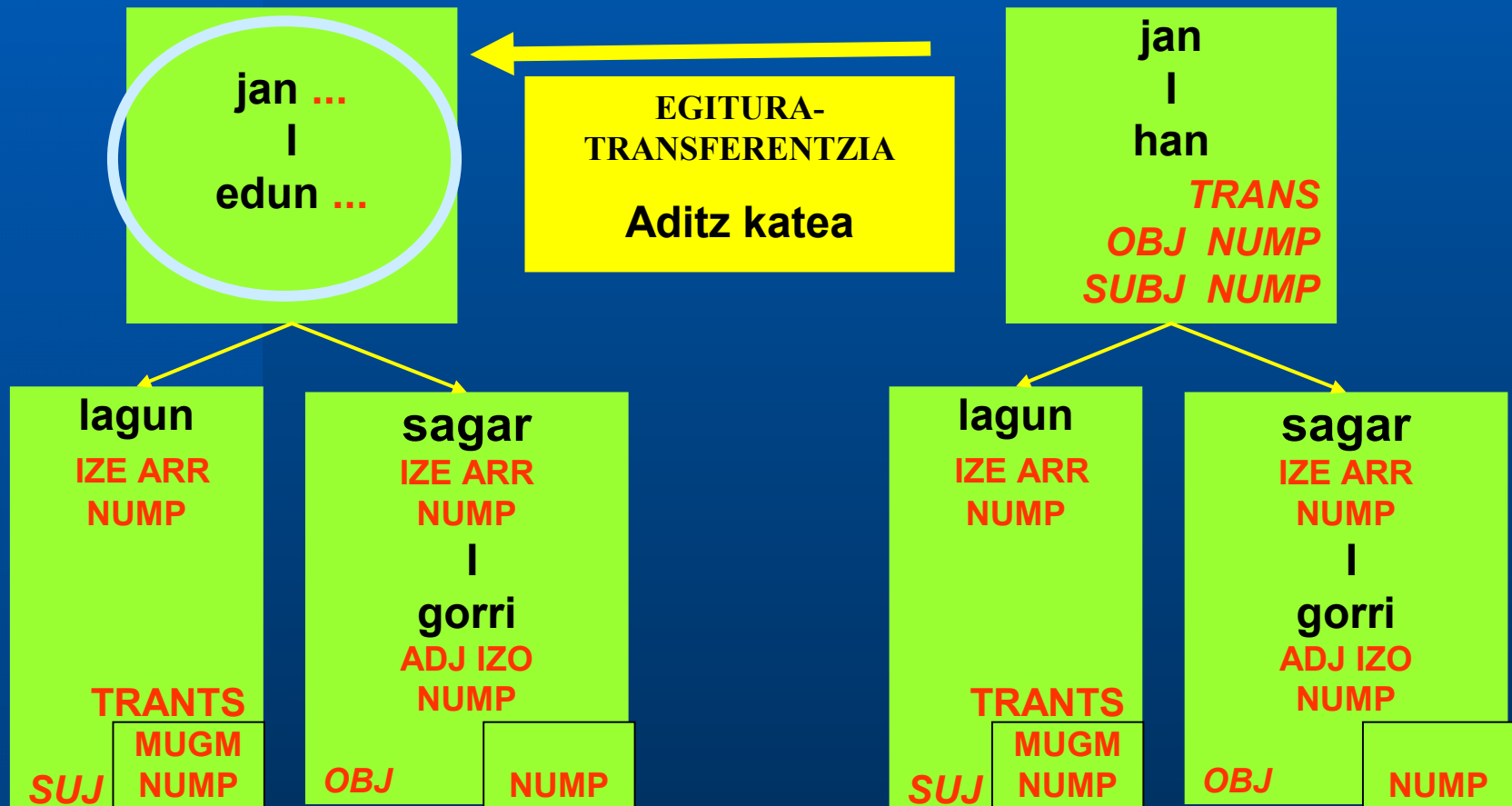




# Transferentzia

## Egitura-transferentzia (II)

“Los amigos han comido manzanas rojas ”

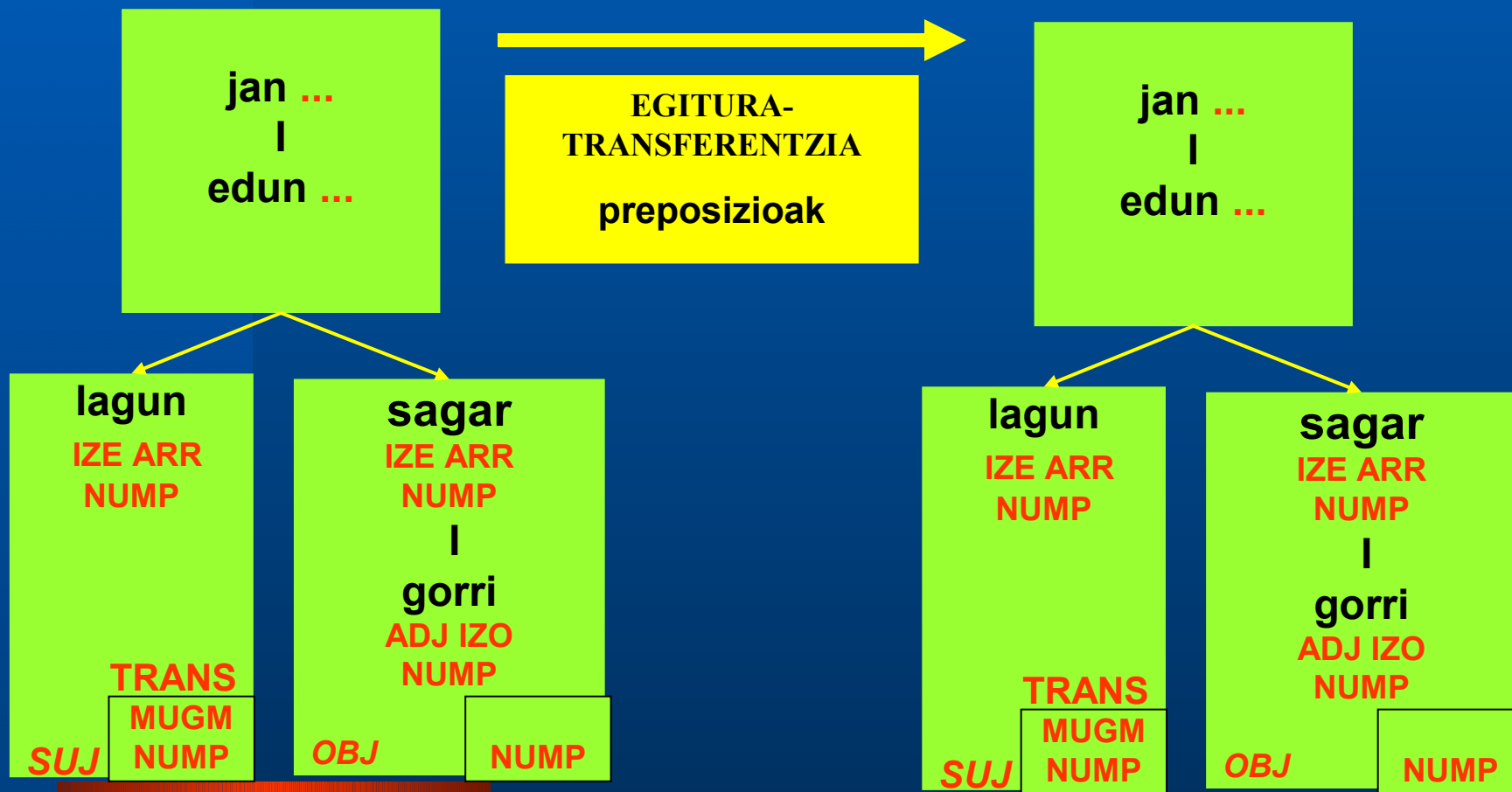




# Transferentzia

## Egitura-transferentzia (II)

“Los amigos han comido manzanas rojas ”

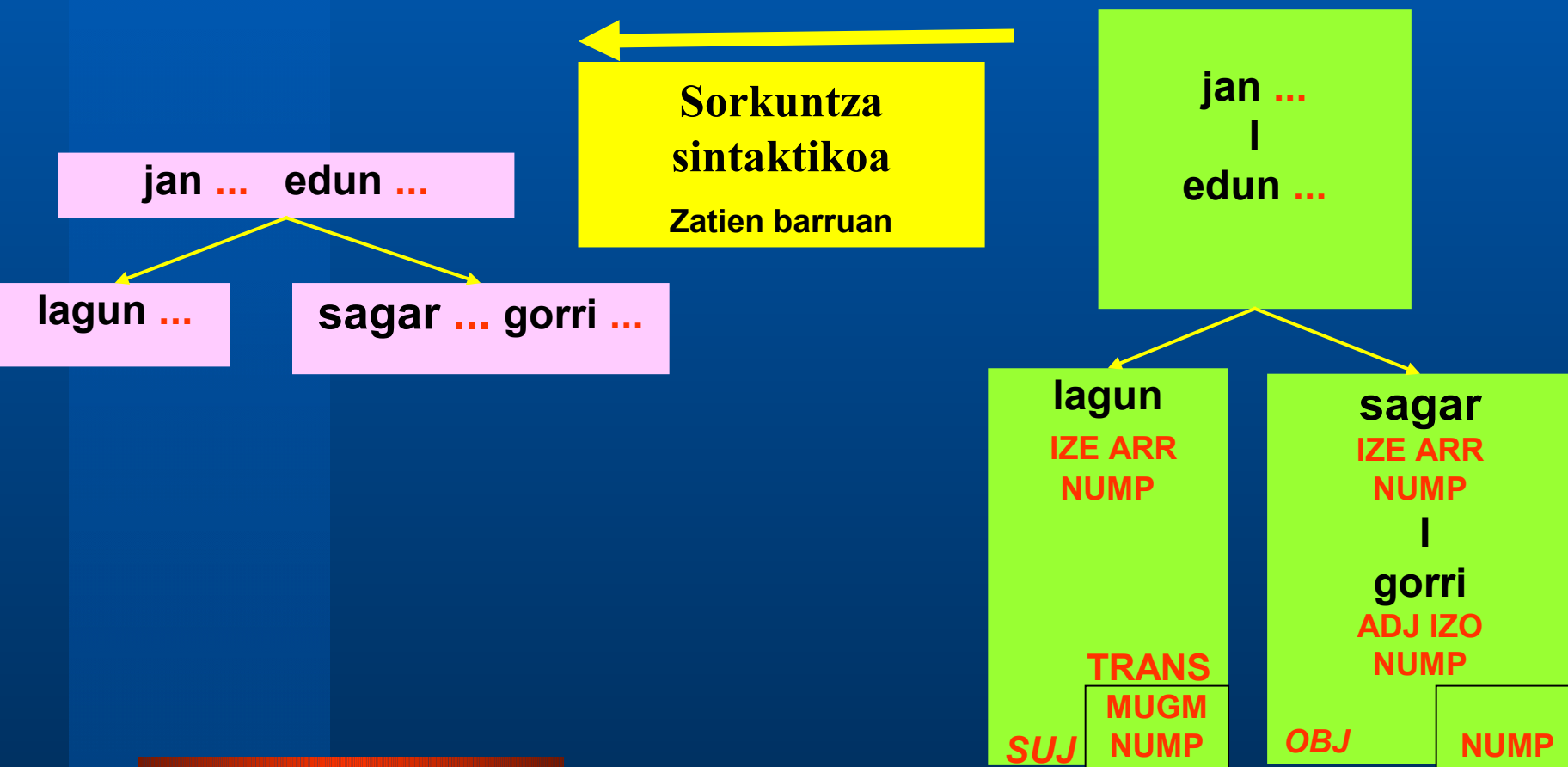




# Transferentzia

## Egitura-transferentzia (II)

“Los amigos han comido manzanas rojas ”

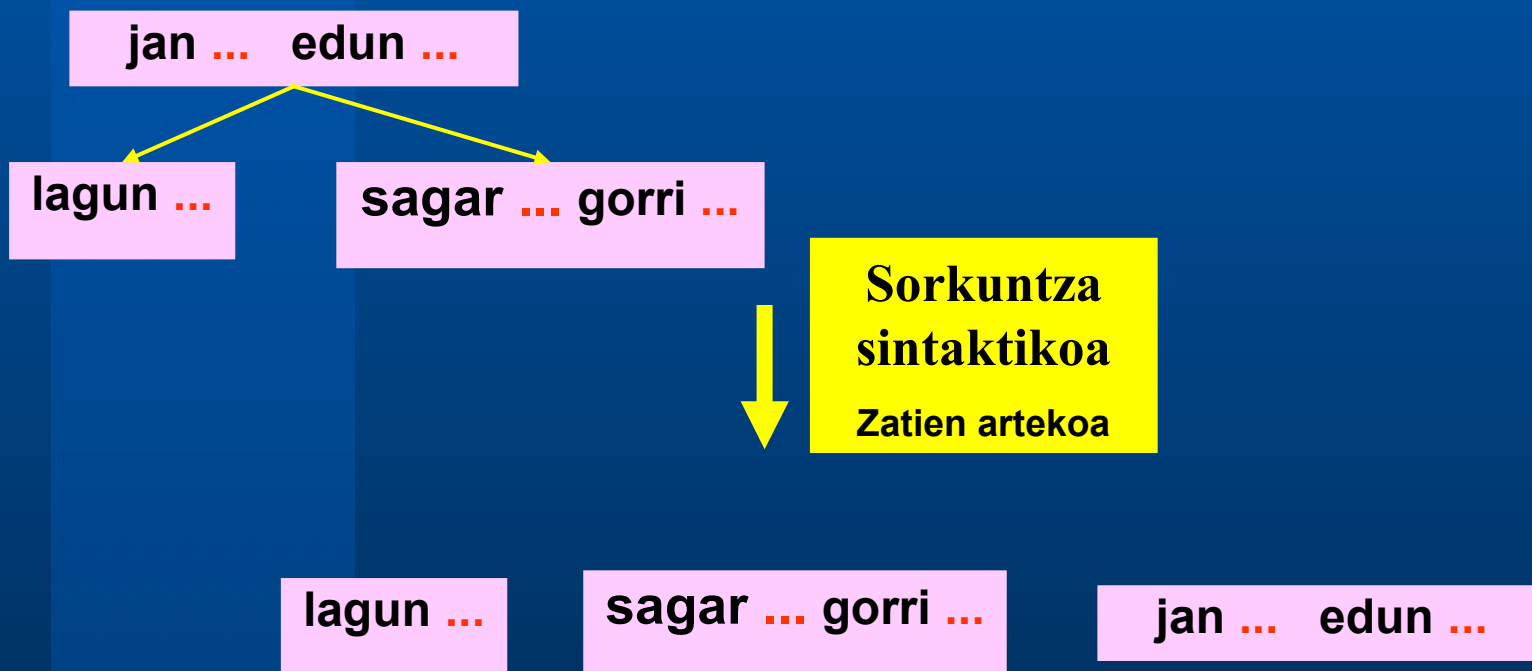




# Transferentzia

## Egitura-transferentzia (II)

“Los amigos han comido manzanas rojas ”





# Transferentzia

## Egitura-transferentzia (II)

“Los amigos han comido manzanas rojas ”

“Lagunek sagar gorriak jan dituzte”



Sorkuntza  
morfologikoa

lagun ...

sagar ... gorri ...

jan ... edun ...





# OpenTrad – Bi jokaera

## Hizkuntza latinoen artekoena:

- Testuaren analisi sakonik ez da behar  
=> Abiadura handia egoera finituko teknologiarri esker  
Itzulpen-prozesua nabigazioarena baina azkarragoa da  
=> nabigatzen den bitartean bisitatutako orrien itzulpena lor daiteke batere atzerapenik gabe
- Fidagarritasuna edo kalitatea ~ % 95

## Espainieratik euskarara

- Analisi sintaktiko sakona behar da  
Automatekin bakarrik ezin da => motelagoa
  - Hitz bat itzultzeko aukera anitz.  
Oraindik semantika ez => maiztasun handieneko adiera  
Euskararekin errore gehiago sortzen dira
- => Kalitate txikiagoa



# OpenTrad - Mugak

## Matxin esaldi sinpleak itzultzen:

- *Ellos ganan esta batalla, pero perderán la guerra.*  
***Beraiek bataila hau irabazten dute, baina gerra galduko dute.***
- *El traductor que estamos desarrollando.*  
***Garatzen ari garen itzultzailea.***
- *Además ahora come manzanas y peras.*  
***Gainera orain sagarrak eta udareak jaten ditu.***



# OpenTrad - Mugak

Matxin prentsako adibide bat itzultzen:

## **SUCESO**

*Un ciudadano vasco aparece ahogado con los pies atados en aguas del Puerto de Barcelona.*

*Dos personas han aparecido muertas en aguas del Puerto de Barcelona. Uno de ellos, al parecer un ciudadano vasco, tenía los tobillos atados entre sí con una corbata.*

## **GERTAERA**

*Euskal hiritar bat <ahogado> azaltzen da <Puerto\_de\_Barcelona> uretan lotutako oinekin.*

*Bi pertsonak hilak azaldu dira <Puerto\_de\_Barcelona> uretan. Beraien bat, euskal hiritar bat, artean lotutako orkatilak hartzen nituen gorbata batekin.*



# OpenTrad - Mugak

Matxin prentsako adibide bat itzultzen:

## SUCESO

*Un ciudadano vasco aparece ahogado con los pies atados en aguas del Puerto de Barcelona.*

*Dos personas han aparecido muertas en aguas del Puerto de Barcelona. Uno de ellos, al parecer un ciudadano vasco, tenía los tobillos atados entre sí con una corbata.*

## GERTAERA

*Euskal hiritar bat <ahogado> azaltzen da <Puerto\_de\_Barcelona> uretan lotutako oinekin.*

*Bi pertsonak hiriak azaidu dira <Puerto\_de\_Barcelona> uretan. Beraien bat, euskal hiritar bat, artean lotutako orkatilak hartzen nituen gorbata batekin.*



# OpenTrad - Mugak

Matxin prentsako adibide bat itzultzen:

## **SUCESO**

*Un ciudadano vasco aparece ahogado con los pies atados en aguas del Puerto de Barcelona.*

*Dos personas han aparecido muertas en aguas del Puerto de Barcelona. Uno de ellos, al parecer un ciudadano vasco, tenía los tobillos atados entre sí con una corbata.*

## **GERTAERA**

*Euskal hiritar bat <ahogado> azaltzen da <Puerto\_de\_Barcelona> uretan lotutako oinekin.*

*Bi pertsonak hilak azaldu dira <Puerto\_de\_Barcelona> uretan. Beraien bat, euskal hiritar bat, artean lotutako orkatilak hartzen nituen gorbata batekin.*



# OpenTrad - Mugak

Matxin prentsako adibide bat itzultzen:

## SUCESO

*Un ciudadano vasco aparece ahogado con los pies atados en aguas del Puerto de Barcelona.*

*Dos personas han aparecido muertas en aguas del Puerto de Barcelona. Uno de ellos, al parecer un ciudadano vasco, tenía los tobillos atados entre sí con una corbata.*

## GERTAERA

*Euskal hiritar bat <ahogado> azaltzen da <Puerto\_de\_Barcelona> uretan lotutako oinekin.*

*Bi pertsonak hilak azaldu dira <Puerto\_de\_Barcelona> uretan. Beraien bat, euskal hiritar bat, artean lotutako orkatilak hartzten nituen gorbata batekin.*



# OpenTrad - Mugak

- Oraingoaz, beti aukeratzen da maiztasun handieneko adieraren itzulpena

Hemos montado un tomate de narices.  
Sudurren tomate bat igo dugu. 😊



# OpenTrad - Mugak

- Analisi sintaktiko anbiguotan interpretazio bakarra aukeratu du analizatzaileak testua ulertu barik.

Beti ez du asmatzen:

- de → 1) -ko
- 2) -tik
- 3) -ren

Trae el pan de la tarde.

✓ **Arratsaldeko ogia dakar.**

Trae el pan de la panadería.

? **Okintzako ogia dakar.**

Trae el pan de su madre.

✓ **Bere amaren ogia dakar.**

Yo voy de casa al campo.

✓ **Ni etxetik noa mendiari.** 😊

Soy de ciudad.

😊 **Hiritik naiz.**





# OpenTrad - Mugak

- Hitz anitzeko unitate lexikalen tratamendua bideratzen:

red wine

**ardo beltza** (EZ *ardo gorria*)

sin embargo

**hala ere** (EZ *enbargorik gabe*)



# Ekinbideak

- **Matxin**
  - Erregeletan eta transferentzian oinarritutako prototipoa
  - es->eu eta en->eu
- **OpenTrad**
  - es->eu, es<->ca, es<->ga
- **Itzulpen-memoriak sortu eta erabiltzeko tresnak**
- **Ikerlerroak**



# Itzulpen-memoriak

- **Itzulpen memoriaren sorkuntza eta kudeaketa automatikorako tresnak**
  - Itzultzeko unitateen identifikazioa
  - Testu elebidunen parekatze-prozesu baten bitartez unitateen baliokideak bilatu.
  - Unitatea esaldia baino txikiagoa: Sintagma mailakoak.
- **Aurreko tresnak erabiliz sortu itzulpen-memoriak TMX formatuan**
- **Itzulpen-memoretan dagoen informazioa kontsultatzeko tresnak. Web serbitzua.**
- **Code & Syntax eta Eleka**



# Ekinbideak

- **Matxin**
  - Erregeletan eta transferentzian oinarritutako prototipoa
  - es->eu eta en->eu
- **OpenTrad**
  - es->eu, es<->ca, es<->ga
- **Itzulpen-memoriak sortu eta erabiltzeko tresnak**
- **Ikerlerroak**



# Ikerlerroak- Transferentzian

(transferentzian oinarritutako hurbilketaren barruan)

**Hobekuntza posibleak:**

- Terminologia-erauzketa automatikoa (ERAUZTERM...)
- Adiera-hautapenerako laguntza aurreratuak (Wordnet, WSD teknikak, ...)
- Aditzen azpikategorizazioari buruzko datuak
- Postedizioa



# Ikerlerroak- Beste hurbilketak

- **Metodo estatistikoak corpus parekatuen gainean**

Kevin Knight

A Statistical MT Tutorial Workbook, August 1999.

[www.isi.edu/natural-language/people/knight.html](http://www.isi.edu/natural-language/people/knight.html)

- **Adibideen Oinarritutako Itzulpen Automatikoa**

Díaz de Ilarraza A., Mayor A., Sarasola K. 2002

Adibideetan Oinarritutako Itzulpen Automatikoa (AOIA): azterketa bibliografikoa eta euskararekin lan egiteko proposamenak

Senez aldizkaria. 17. - 35. orr.

[ixa.si.ehu.es/lxa/Argitalpenak/Artikuluak](http://ixa.si.ehu.es/lxa/Argitalpenak/Artikuluak)

- **Metodo hibridoak**

Hutchins John

The State of Machine Translation in Europe and Future Prospects.

[www.kkhec.ac.ir/Linguistics%20articles%20index/state\\_of\\_machine\\_translation\\_in.htm](http://www.kkhec.ac.ir/Linguistics%20articles%20index/state_of_machine_translation_in.htm)



# Ikerlerroak- Beste aplikaziotan

*“Hemendik urte batzuetara, agian ez ditugu ikusiko itzulpen automatiko “huts”-eko sistemarik, baizik eta ordenagailu bidezko hainbat tresna eta aplikazio non Itzulpen Automatikoa osagaietako bat baino izango ez den. “ J. Hutchins*

- Denak batera, eta erabiltzaileak konturatu gabe:
  - informazio-bilaketa
  - informazio-erauzketa
  - laburbiltze-sistemak
  - itzulpen automatikoa
- Ildo horretatik doaz:
  - “cross-lingual information retrieval”
  - laburbiltze-sistema eleanitzak
  - testu-sorkuntza eleanitza datu-baseetatik



# Ikerlerroak: Itzulpen-laguntzak

## Itzultzailearentzako lan-estazioak (*workstation*)

- Itzulpen memoriak
- Hiztegi elektronikoen kontsulta aurreratua
- Terminologiaren erauzketa, bilaketa eta kudeaketa
- Testu-edizio eleanitza errazten duen hainbat leiho era koordinatuan kudeatzen dituen interfazea
- Softwarearen lokalizazio-lanak errazteko tresnak
- Bitestuak (corpus elebidunak) parekatzeko tresnak
- Eta nahi izanez gero, eta nahi den zatian... itzulpen automatiko osoa erabiltzeko aukera