CL2009 Liverpool, 20-23 July

NOCANDO: A multilingual annotated corpus for the study of Information Structure

Lisa Brunetti[1], Stefan Bott[2], Joan Costa[3], Enric Vallduví[3]
lisa.brunetti@lpl-aix.fr, sbott@lsi.upc.edu, joan.costa@upf.edu, enric.vallduvi@upf.edu

[1]Laboratoire Parole et Langage, Université de Provence I,
[2]Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya
[3]Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra

Outline of the presentation

- The project: theoretical goals
- The corpus:
  - quantitative information
  - participants
  - methodology
  - transcription and segmentation
  - annotation
- Corpus exploitation
- Future work

The project

The project

NOCANDO: Construcciones no-canónicas en el discurso oral : un estudio transversal y comparativo
(‘Non-canonical constructions in oral speech:
a crosslinguistic perspective’)

The project

NOCANDO: Construcciones no-canónicas en el discurso oral : un estudio transversal y comparativo
(‘Non-canonical constructions in oral speech:
a crosslinguistic perspective’)

Funded by the Ministerio de Educación y Ciencia de España
(I+D HUM2004-04463), 2004-2007.

Principal investigator: Enric Vallduví

Reserach group: Grup de Lingüística Formal (GLiF) http://parles.upf.edu/glif/
Dep. Traducció i Ciències de Llenguatge
Universitat Pompeu Fabra, Barcelona

Collaborators: Lisa Brunetti, Stefan Bott, Joan Costa, Estela Puig Waldmüller,
Teresa Suñol, Louise McNally, Josep Maria Fontana, Alex Alsina.

The project

Goal

NOCANDO seeks to establish a cross-linguistically valid taxonomy of **non-canonical constructions (NOCANs).**

The project

Goal

NOCANDO seeks to establish a cross-linguistically valid taxonomy of **non-canonical constructions (NOCANs).**

What is a NOCAN?

A morphologically, syntactically and/or prosodically marked construction from the point of view of the properties of a language.

The project

Goal

NOCANDO seeks to establish a cross-linguistically valid taxonomy of **non-canonical constructions (NOCANs).**

What is a NOCAN?

A morphologically, syntactically and/or prosodically marked construction from the point of view of the properties of a language.

What is the function of a NOCAN?

NOCANs otpimize the way the informational content of a sentence is conveyed (Vallduví 1992). NOCANs are explicit marks of the INFORMATION STRUCTURE of the sentence.

The project

Examples of NOCANs

(1)	Al hombre	se	le	cae	el café
	to-the man	RFL	to-him	falls	the coffee
	Ind. Object		obj.cl.	Verb	Subject

“The man drops the coffee”

The project

Examples of NOCANs

(1)	Al hombre	se	le	cae	el café
	to-the man	RFL	to-him	falls	the coffee
	Ind. Object		obj.cl.	Verb	Subject
	“The man	drops the coffee”			

Clitic Left Dislocation

The project

Examples of NOCANs

(1)	Al hombre	se	le	cae	el café
	to-the man	RFL	to-him	falls	the coffee
	Ind. Object		obj.cl.	Verb	Subject

“The man drops the coffee”

Topic **Comment**

Clitic Left Dislocation

The project

Examples of NOCANs

(1)

Al hombre	se	le	cae	el café
to-the man	RFL	to-him	falls	the coffee
Ind. Object		obj.cl.	Verb	Subject

“The man drops the coffee”

Topic **Comment**

Clitic Left Dislocation

(2)

Pure la LINGUACCIA,	gli	fa,	la rana.
even the tongue	to-him	he-puts-out	the frog
Dir. Object		Verb	Subject

“Even the tongue did the frog put out to him”

The project

Examples of NOCANs

(1)

Al hombre	se	le	cae	el café
to-the man	RFL	to-him	falls	the coffee
Ind. Object		obj.cl.	Verb	Subject

“The man drops the coffee”

Topic **Comment**

Clitic Left Dislocation

(2)

Pure la LINGUACCIA,	gli	fa,	la rana.
even the tongue	to-him	he-puts-out	the frog
Dir. Object		Verb	Subject

“Even the tongue did the frog put out to him”

Focus Fronting

The project

Examples of NOCANs

(1)

Al hombre	se	le	cae	el café
to-the man	RFL	to-him	falls	the coffee
Ind. Object		obj.cl.	Verb	Subject

"The man drops the coffee"

Topic **Comment**

Clitic Left Dislocation

(2)

Pure la LINGUACCIA,	gli	fa,	la rana.
even the tongue	to-him	he-puts-out	the frog
Dir. Object		Verb	Subject

"Even the tongue did the frog put out to him"

Focus **Background**

Focus Fronting

The project

Examples of NOCANs

(3) - I no se n'adona que la granota s'ha posat a davant
and not realizes that the frog herself has put in front

- i **és** ELLA **que** està a punt de prendre's el biberó
and is her who is about to take for-himself the baby-bottle

“And she does not realize that the frog placed himself before the baby and it’s him who is going to drink from the bottle”

The project

Examples of NOCANs

(3) - I no se n'adona que la granota s'ha posat a davant
and not realizes that the frog herself has put in front

- i **és** ELLA **que** està a punt de prendre's el biberó
and is her who is about to take for-himself the baby-bottle

Cleft

“And she does not realize that the frog placed himself before the baby and it’s him who is going to drink from the bottle”

The project

Examples of NOCANs

(3) - I no se n'adona que la granota s'ha posat a davant
and not realizes that the frog herself has put in front

- i **és** ELLA
and is her

que està a punt de prendre's el biberó
who is about to take for-herself the baby-bottle

Contrastive focus

Background

Cleft

“And she does not realize that the frog placed himself before the baby and it’s him who is going to drink from the bottle”

The corpus

- ❖ quantitative information
- ❖ participants
- ❖ methodology
- ❖ transcription and segmentation
- ❖ annotation

The corpus: general information

Spontaneous narrations in Catalan, Italian, Spanish, German, and English.

Total number of speakers: **68**
Total number of narrations: **222**
Total duration: ca **16 h** (2’-10’ per narration)

	Catalan	**Italian**	**Spanish**	**German**	**English**
Speakers	19	16	13	9	11
Recording time	4:02:43 h	4:04:32 h	2:35:20 h	2:09:13	2:32:20 h
Word count (linux/cygwin)	37555 w	27392 w	25077 w	15944 w	21970 w (es)
Segment count	5856 seg	4306 seg	3801 seg	2154 seg	3140 seg (es)

The corpus: participants

The corpus: participants

Mostly **students** at the **Universitat Pompeu Fabra** in **Barcelona**.
A smaller number from different working environments.

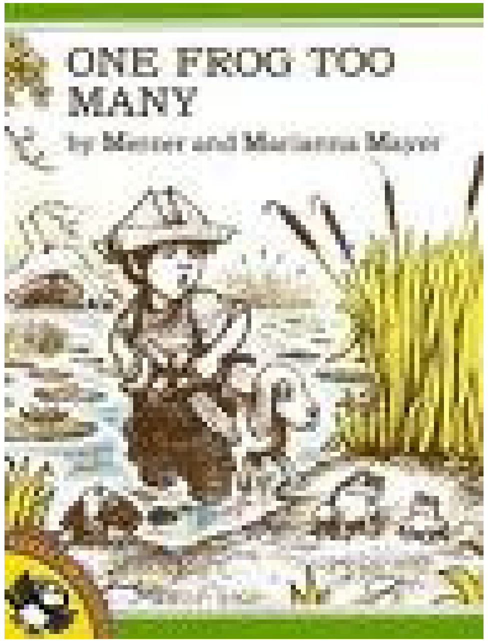
	Catalan	Spanish	Italian	German	English
Geo-graphical origin	Catalonia (except one from Comunitat Valenciana)	Catalonia (except one from Castilla y León)	Different parts of Italy	Different parts of Germany	Different parts of USA and UK
Mean age	**22** (18-30)	**20** (17-29)	**29** (20-56)	**34** (22-67)	**27** (20-41)

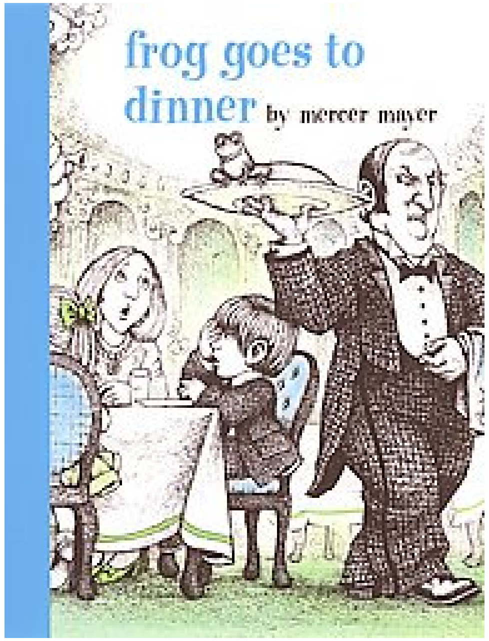
The corpus: methodology

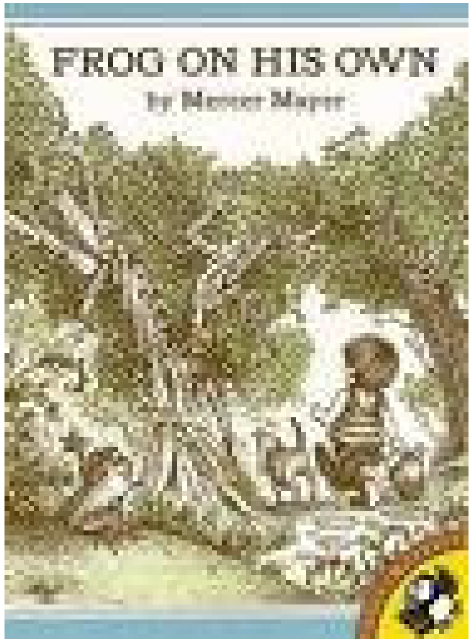
The corpus: methodology

Speakers narrated the stories of three text-less picture story books by Mercer Meyer: *Frog goes to dinner , Frog on his own, One frog too many.*

Cf. Berman and Slobin 1994, Strömqvist & Verhoven 2004







The corpus: methodology

Speakers told the experimenter the stories in a random order.
Speakers could browse the book before starting the narration.

A questionnaire was filled by speakers concerning age, geographical origin, personal language history.

The corpus: transcription & segmentation

The corpus: transcription & segmentation

Orthographic transcription based on the LIP corpus (De Mauro et al. 1993).

Truncated phrase or sentence	**la tortug-- la granota**
Pauses	**#, ##, ###**
Unintelligible words	**[?], [?], [???]**
Missing part	**[...]**
Reconstructed broken word	**sta[te]**
Truncated word	**-pe-**
Vocalic lengthening at the end of a word	**bueno_**
Extra-linguistic comment	**[LAUGHS]**
Phonetic symbol	**[fff]**
Hesitation	**hm**
Standard phrasing symbols:	**(,) (.) (?) (!)**

The corpus: transcription & segmentation

Orthographic transcription based on the LIP corpus (De Mauro et al. 1993).

Truncated phrase or sentence	**la tortug-- la granota**
Pauses	**#, ##, ###**
Unintelligible words	**[?], [?], [???]**
Missing part	**[...]**
Reconstructed broken word	**sta[te]**
Truncated word	**-pe-**
Vocalic lengthening at the end of a word	**bueno_**
Extra-linguistic comment	**[LAUGHS]**
Phonetic symbol	**[fff]**
Hesitation	**hm**
Standard phrasing symbols:	**(,) (.) (?) (!)**

Segmentation

One clause – one line

- subordinate clauses included (cf. CHILDES).

- Temporal/aspectual and modal verb periphrases excluded. (Criteria to identify periphrases: Gavarró and Laca 2002).

The corpus: Annotation

The corpus: Annotation

An annotation of NOCANs is realized for **Catalan, Italian, and Spanish**.

Cf. MULI corpus, *Baumann 2006.*

The corpus: Annotation

An annotation of NOCANs is realized for **Catalan, Italian, and Spanish**.

Cf. MULI corpus, *Baumann 2006.*

Why these languages?

- Similar linguistic properties (relatively free word order, null sbj, SVO, Obj Cl Pro..)
- Similar strategies to express informational notions:
  - large use of syntax
  - limited use of prosody (cf. English) Vallduví and Engdhal 1996
- **Similar or identical NOCANs**

The corpus: Annotation

An annotation of NOCANs is realized for **Catalan, Italian, and Spanish**.

Cf. MULI corpus, *Baumann 2006.*

Why these languages?

- ➢ Similar linguistic properties (relatively free word order, null sbj, SVO, Obj Cl Pro..)
- ➢ Similar strategies to express informational notions:
  - large use of syntax
  - limited use of prosody (cf. English) Vallduví and Engdhal 1996
- ➢ **Similar or identical NOCANs**

However, NOCANs of these languages may vary in terms of:

- frequency (e.g. clrd in Catalan vs Spanish, Villalba 2007), Leonetti 2008
- function (e.g. subject inversion in Spanish vs Italian)

The corpus: Annotation

NOCANs concerning the subject

The corpus: Annotation

NOCANs concerning the subject

sbjinv = subject inversion (the subject occurs after the verb)

(4) Els va acompanyar **el taxista**. Catalan
them PAST take the taxi-driver
'The TAXI-DRIVER drove them'

sbjinv_deacc = post-verbal deaccented subject in a declarative sentence.

(5) ...que està disfressat, **aquest nen.** Catalan
for is dressed-up this child
"...for this child is dressed up"

The corpus: Annotation

NOCANs concerning the subject

sbjinv = subject inversion (the subject occurs after the verb)

(4) Els va acompanyar **el taxista**. Catalan
them PAST take the taxi-driver
'The TAXI-DRIVER drove them'

sbjinv_deacc = post-verbal deaccented subject in a declarative sentence.

(5) ...que està disfressat, **aquest nen.** Catalan
for is dressed-up this child
"...for this child is dressed up"

nsbj = null subject (the subject is not expressed)

(6) Invece il bambino è molto contento, perché **ha salvato** la sua rana. Italian
instead the boy is very happy because has saved the his frog
'The boy on the contrary is very happy, because he saved his frog'

nsbj_c = null subject in an coordinate clause

The corpus: Annotation

NOCANs concerning the subject

arbnsbj = arbitrary subject (which is null in these languages, cf. Jaeggli 1986)

(7) Y un día a este niño **le regalaron** pues una caja muy grande Spanish
and one day to this boy to-him they-gave well a box very big
‘And one day this boy received a large box’

sbj-sep = separation of the (preverbal) subject from the verb by sentential
adverbials or other material that has scope over the entire clause

(8) però **la rana** come al solito riesce a infilarsi nella_ nella situazione
but the frog as the usual manages to sneak-herself into-the into-the situation
‘but the frog, as usual, manages to sneak into the the,,, situation’ Italian

The corpus: Annotation

NOCANs concerning all arguments

clld = clitic left dislocation   An argument dislocated to the left and resumed by a clitic pronoun (cf. Benincà 2001, Zubizarretta 1999, Cinque 1990)

(9) **Al hombre** se **le** cae el café.   Spanish
to-the man RFL to-him the coffee
"The man drops the coffee"

ld = left dislocation   Dislocation to the left WITHOUT clitic resumption

(10) **A un bambino** un giorno arriva un regalo.   Italian
to a boy one day arrives a present
"One day a boy receives a present"

ht = hanging topic   A left dislocated NP resumed by a pronoun expressing its grammatical function. (cf. Benincà 2001, Zubizarretta 1999)

(11) **La rana grande**, la situación no **le** gustaba mucho.   Spanish
the frog big   the situation not to-her pleases much
"As for the big frog, she didn't like the situation at all"

The corpus: Annotation

NOCANs concerning all arguments

clrd = clitic right dislocation   An argument dislocated to the right and resumed by a clitic pronoun. (cf. Benincà 2001, Zubizarretta 1999)

(12) el gat ja **l'**ha vist, **a la granota** Catalan
the cat already it has seen to the frog
"The cat already SAW the frog"

rd = right dislocation   Dislocation to the right WITHOUT resumption

(13) y **le_** muerde_ [...] el anca a la otra ranita, **la rana grande.** Spanish
and to-him he-bites the hip to the other little-frog, the frog big
"and the big frog bites the other frog's hip"

cldbl = clitic doubling It differs from clrd in that the doubled argument is in the same intonational contour as the verb (focus domain).

(15) Entonces la tortuga lo ve y **se** lo dice **al niño.** Spanish
so the turtle it sees and to-him it says to-the boy
"So the turtle sees what happened and tells the boy everything"

The corpus: Annotation

NOCANs concerning all arguments

obj-sep = separation of the (postverbal) direct object from the verb

(16) Y cogió en su mano **a la ranita pequeña** Spanish
and he-took in his hand to the frog little
“And he took the little frog into his hands”

narg = null argument

(17) i llavors en Jaume es va adonar que que, home, **era la seva granota** Catalan
and then the Jaume RFL PAST realizes that that well it-was the his frog
“and therefore Jaume ralizes that that, well, it was his frog”

The corpus: Annotation

NOCANs concerning all arguments

focfr = focus fronting: A left peripheral element with focal stress (Benincà 2001, Rizzi 1997, Zubizarreta 1998)

(18) veu una dona que està amb un cotxet; així **de LLUNY** la veu. Catalan
sees a woman who is with a baby-carriage like-that from far her he-sees
"He sees a woman with a baby carriage; she sees it from far away"

deacc = de-accenting

(19) ma Lara non è molto simpatica, **con questa rana.** Italian
but Lara not is very nice with this frog
'But Lara is NOT very nice, towards this frog'

The corpus: Annotation

NOCANs concerning clauses

pres = presentational sentences

(20) **C'era** una volta un bambino — Italian
there was one time a boy
"Once upon a time there was a boy"

pass = passive construction

(21) la familia de William **es expulsada** del restaurante. — Spanish
the family of William is expelled from-the restaurant
"William's family is expelled from the restaurant"

impers = impersonal construction

(22) e lui continua hm a indicare **non si sa** dove. — Italian
and he keeps hm to point not IMP knows where
'And he keeps pointing who knows where'

The corpus: Annotation

Types of clauses

cleft = cleft sentences
Construction: Verb 'to be' + XP + 'that' S without XP

(23) **Era** ese SAPO **que** les había querido hacer algún susto Spanish
"It was the frog who wanted to scare them"

pscleft = pseudo-cleft sentences
Construction: Dem. pron. + Rel. clause + verb 'to be' + 'that' S / NP

(24) y **lo que** pasa **es que** el barquito se hunde Spanish
"and what happens is that the boat sinks"

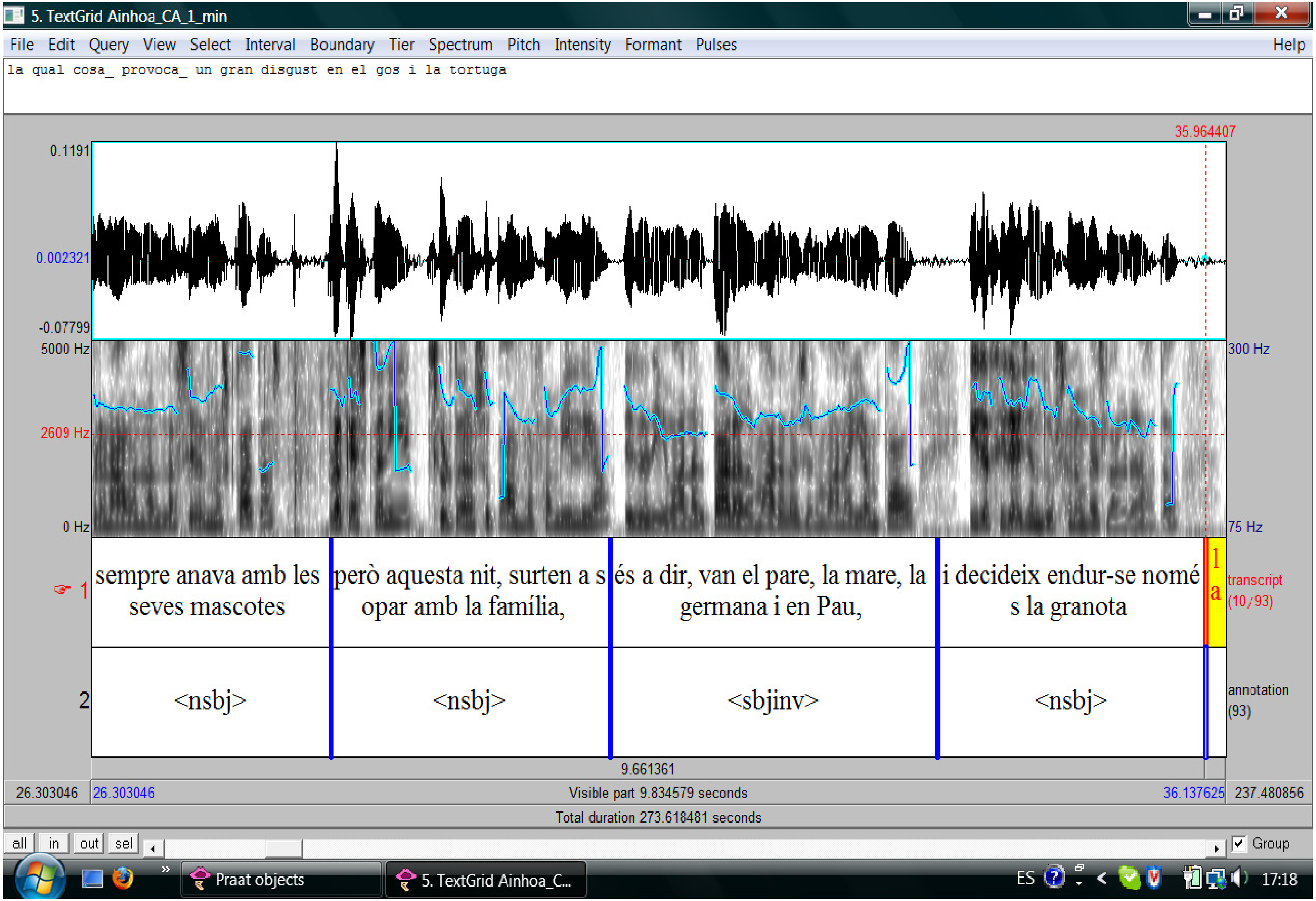
inv-pscleft = inverted pseudo-cleft sentences
Construction: NP + verb 'to be' + Dem. pron. + Rel. clause

(25) Y bueno, el niño **es el que** dirige la balsa Spanish
"and well the boy is the-one who leads the boat"

Example of transcription and annotation

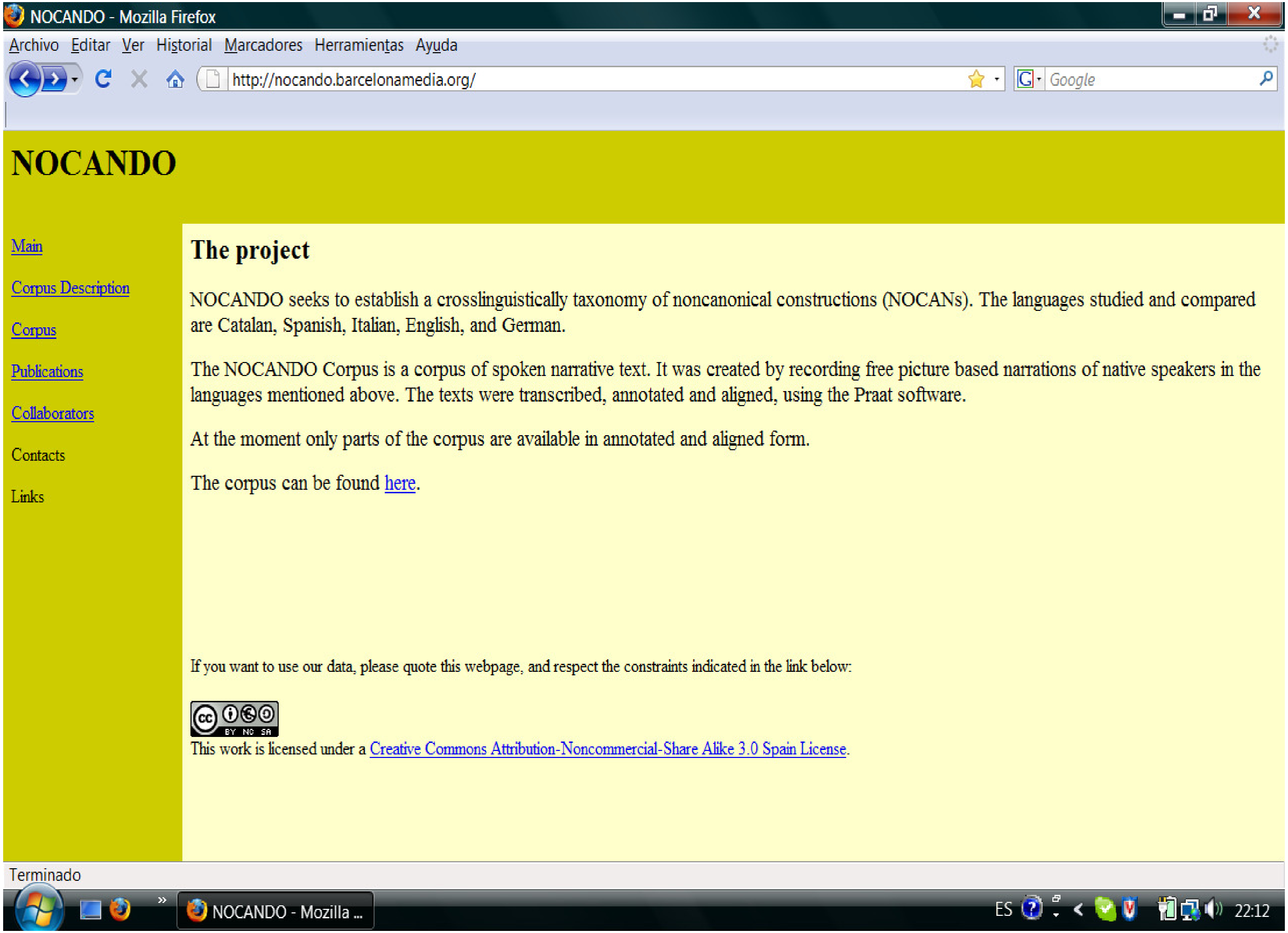
```
<segment id="IT_04_2_0004" nocans="nsbj">
                    Erano stanchi della città,  </segment>
<segment id="IT_04_2_0005" nocans="nsbj">
                    avevano bisogno di un poco di natura, di aria fresca,  </segment>
<segment id="IT_04_2_0006" nocans="">
                    di camminare in mezzo agli alberi...  </segment>
<segment id="IT_04_2_0007" nocans="nsbj">
                    E così, tutti contenti, uscirono di casa  </segment>
<segment id="IT_04_2_0008" nocans="nsbj_c">
                    e andarono verso il bosco.  </segment>
<segment id="IT_04_2_0009" nocans="">
                    Michelino aveva messo la rana e la tartaruga in un secchiello,
</segment>
<segment id="IT_04_2_0010" nocans="clld,ld">
                    il cane nel secchiello non c'entrava, ovviamente.  </segment>
<segment id="IT_04_2_0011" nocans="">
                    Cammina cammina,  </segment>
<segment id="IT_04_2_0012" nocans="cldbl">
                    a un certo punto la rana ne approfitta di un momento di distrazione,
di Michelino,  </segment>
```

Example of transcription and annotation on Praat



```
la qual cosa_ provoca_ un gran disgust en el gos i la tortuga
```

sempre anava amb les seves mascotes	però aquesta nit, surten a sopar amb la família,	és a dir, van el pare, la mare, la germana i en Pau,	i decideix endur-se només la granota	la
<nsbj>	<nsbj>	<sbjinv>	<nsbj>	



NOCANDO - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda
http://nocando.barcelonamedia.org/
Google

NOCANDO

Main
Corpus Description
Corpus
Publications
Collaborators
Contacts
Links

The project

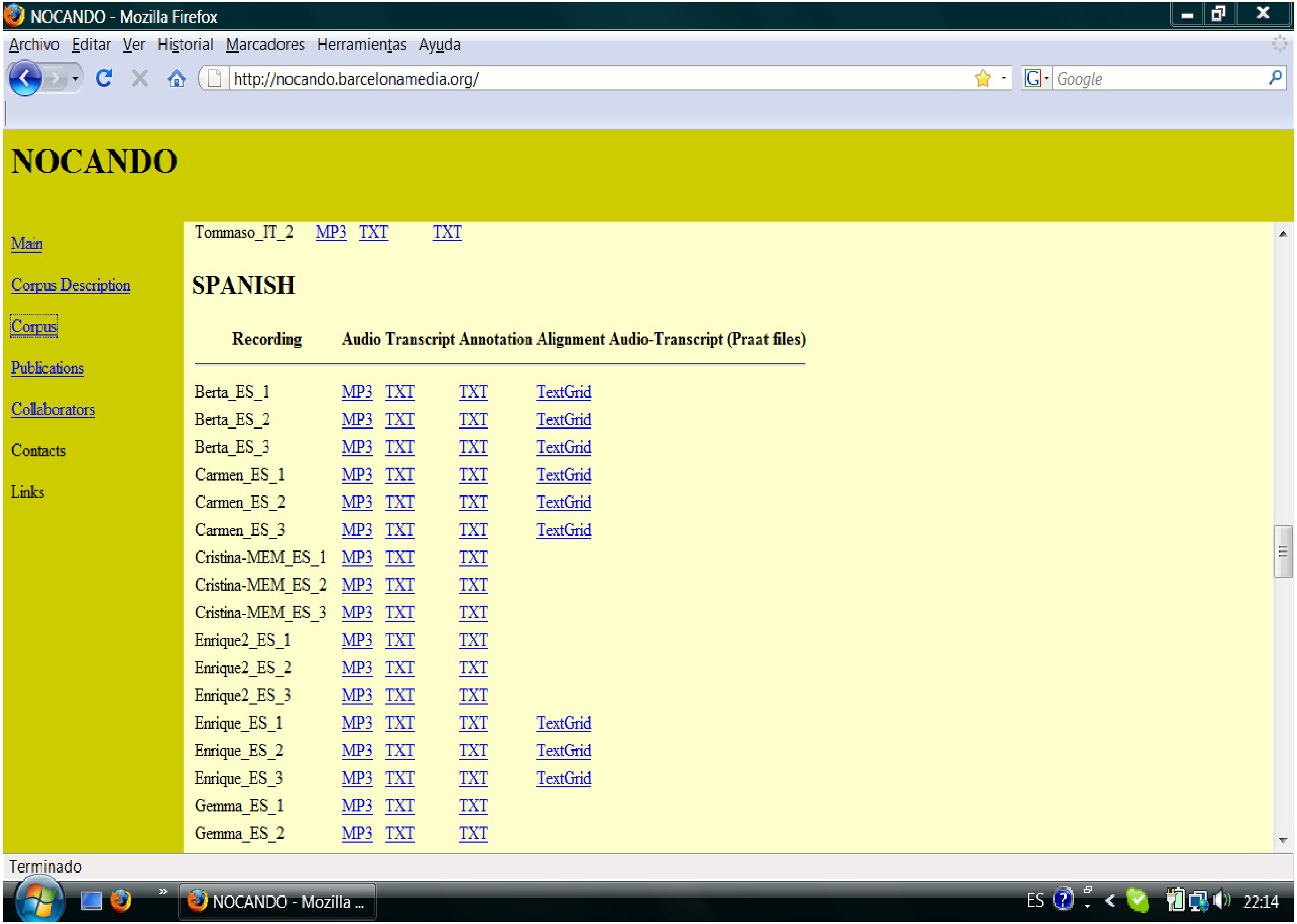
NOCANDO seeks to establish a crosslinguistically taxonomy of noncanonical constructions (NOCANs). The languages studied and compared are Catalan, Spanish, Italian, English, and German.

The NOCANDO Corpus is a corpus of spoken narrative text. It was created by recording free picture based narrations of native speakers in the languages mentioned above. The texts were transcribed, annotated and aligned, using the Praat software.

At the moment only parts of the corpus are available in annotated and aligned form.

The corpus can be found here.

If you want to use our data, please quote this webpage, and respect the constraints indicated in the link below:


NOCANDO - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda
http://nocando.barcelonamedia.org/
Google
Terminado
NOCANDO - Mozilla ...
ES
22:14

NOCANDO

Main

Corpus Description

Corpus

Publications

Collaborators

Contacts

Links

Tommaso_IT_2 MP3 TXT TXT

SPANISH

Recording	Audio	Transcript	Annotation	Alignment Audio-Transcript (Praat files)
Berta_ES_1	MP3	TXT	TXT	TextGrid
Berta_ES_2	MP3	TXT	TXT	TextGrid
Berta_ES_3	MP3	TXT	TXT	TextGrid
Carmen_ES_1	MP3	TXT	TXT	TextGrid
Carmen_ES_2	MP3	TXT	TXT	TextGrid
Carmen_ES_3	MP3	TXT	TXT	TextGrid
Cristina-MEM_ES_1	MP3	TXT	TXT	
Cristina-MEM_ES_2	MP3	TXT	TXT	
Cristina-MEM_ES_3	MP3	TXT	TXT	
Enrique2_ES_1	MP3	TXT	TXT	
Enrique2_ES_2	MP3	TXT	TXT	
Enrique2_ES_3	MP3	TXT	TXT	
Enrique_ES_1	MP3	TXT	TXT	TextGrid
Enrique_ES_2	MP3	TXT	TXT	TextGrid
Enrique_ES_3	MP3	TXT	TXT	TextGrid
Gemma_ES_1	MP3	TXT	TXT	
Gemma_ES_2	MP3	TXT	TXT	

Corpus exploitation

Corpus exploitation

❑ **Naturally occurring data** for theoretical studies on information structure and discourse (cf. Bott 2007, Brunetti 2009a,b, Mayol 2009).

Corpus exploitation

❑ **Naturally occurring data** for theoretical studies on information structure and discourse (cf. Bott 2007, Brunetti 2009a,b, Mayol 2009).

❑ Cooccurrence of NOCANs with particular linguistic environments or with other NOCANs (cf. Brunetti 2009a).

Corpus exploitation

❑ **Naturally occurring data** for theoretical studies on information structure and discourse (cf. Bott 2007, Brunetti 2009a,b, Mayol 2009).

❑ Cooccurrence of NOCANs with particular linguistic environments or with other NOCANs (cf. Brunetti 2009a).

❑ Comparison among Romance languages.

Corpus exploitation

	Catalan		Italian		Spanish	
overt sbj	1561	**35,7 %**	1262	**38, 9 %**	1027	**35,5 %**
nsbj	1665	**38,1 %**	1173	**36,1 %**	1084	**37,5 %**

Corpus exploitation

	Catalan		Italian		Spanish	
overt sbj	1561	**35,7 %**	1262	**38, 9 %**	1027	**35,5 %**
nsbj	1665	**38,1 %**	1173	**36,1 %**	1084	**37,5 %**
sbjinv	332	**7,6 %**	215	**6,6 %**	265	**9,1 %**

Corpus exploitation

	Catalan		Italian		Spanish	
overt sbj	1561	**35,7 %**	1262	**38, 9 %**	1027	**35,5 %**
nsbj	1665	**38,1 %**	1173	**36,1 %**	1084	**37,5 %**
sbjinv	332	**7,6 %**	215	**6,6 %**	265	**9,1 %**
Clld + ld	62	**1,4%**	44	**1,35%**	39	**1,35 %**
Clrd + rd	22	**0,5 %**	21	**0,64 %**	11	**0,38 %**
ht	10	**0,2%**	2	**0,06%**	9	**0,3 %**

Corpus exploitation

	Catalan		Italian		Spanish	
overt sbj	1561	**35,7 %**	1262	**38, 9 %**	1027	**35,5 %**
nsbj	1665	**38,1 %**	1173	**36,1 %**	1084	**37,5 %**
sbjinv	332	**7,6 %**	215	**6,6 %**	265	**9,1 %**
Clld + ld	62	**1,4%**	44	**1,35%**	39	**1,35 %**
Clrd + rd	22	**0,5 %**	21	**0,64 %**	11	**0,38 %**
ht	10	**0,2%**	2	**0,06%**	9	**0,3 %**
cldbl	92	**2,1 %**	7	**0,2 %**	61	**2,1 %**

Corpus exploitation

	Catalan		Italian		Spanish	
overt sbj	1561	**35,7 %**	1262	**38, 9 %**	1027	**35,5 %**
nsbj	1665	**38,1 %**	1173	**36,1 %**	1084	**37,5 %**
sbjinv	332	**7,6 %**	215	**6,6 %**	265	**9,1 %**
Clld + ld	62	**1,4%**	44	**1,35%**	39	**1,35 %**
Clrd + rd	22	**0,5 %**	21	**0,64 %**	11	**0,38 %**
ht	10	**0,2%**	2	**0,06%**	9	**0,3 %**
cldbl	92	**2,1 %**	7	**0,2 %**	61	**2,1 %**
pscleft **+ inv-psclef**	40	**0,9 %**	10	**0,3 %**	37	**1,28 %**

Corpus exploitation

	Catalan		Italian		Spanish	
overt sbj	1561	**35,7 %**	1262	**38, 9 %**	1027	**35,5 %**
nsbj	1665	**38,1 %**	1173	**36,1 %**	1084	**37,5 %**
sbjinv	332	**7,6 %**	215	**6,6 %**	265	**9,1 %**
Clld + ld	62	**1,4%**	44	**1,35%**	39	**1,35 %**
Clrd + rd	22	**0,5 %**	21	**0,64 %**	11	**0,38 %**
ht	10	**0,2%**	2	**0,06%**	9	**0,3 %**
cldbl	92	**2,1 %**	7	**0,2 %**	61	**2,1 %**
pscleft **+ inv-psclef**	40	**0,9 %**	10	**0,3 %**	37	**1,28 %**
pass	5	**0,1 %**	67	**2 %**	7	**0,24 %**

Future work

- ❖ Extension of the corpus
- ❖ Collection and annotation of a corpus of dialogues in the same languages
- ❖ Extension of the corpus to other languages
- ❖ Annotation of NOCANs in Germanic languages
- ❖ Extension of the annotation to informational categories (cf. Bauman 2006, Calhoun et al. 2005, a.o.), semantic categories (thematic roles, animacy...), and discourse properties.

Selected References

❑ Baumann, S. 2006, ‘Information Structure and Prosody: Linguistic Categories for Spoken Language Annotation’ in S. Sudhoff et al. (eds.), Methods in Empirical Prosody Research (Language, Context and Cognition 3), W. de Gruyter.

❑ Boersma, P. and Weenink, D. 2009, Praat: doing phonetics by computer (Version 5.0.47) [Computer program]. Retrieved January 21, 2009, from http://www.praat.org/

❑ Bott. S. 2007 Information Structure and Discourse Modelling. PhD.Diss, UPF.

❑ Brunetti, L. 2009a, ‘On the semantic and contextual factors that determine topic selection in Italian and Spanish’, in G. van Bergen et H. de Hoop (eds.), *Special issue on Topics Cross-linguistically, The Linguistic Review* 26, 2/3.

❑ Brunetti, L. 2009b, Discourse Functions of Fronted Foci in Italian and Spanish’, in A. Dufter et D. Jacob (eds.) *Focus and Background in Romance Languages, Studies in Language Companion Series,* Benjamins.

❑ Gavarró, A.; Laca, B. 2002, “Les perífrasis temporals, aspectuals i modals”. In: Solà, joan [et al.]. Gramàtica del català contemporani, Barcelona, Empúries, vol.3, pp. 2665-2774.

❑ Strömqvist, S. and L. T. Verhoven (eds.) 2004, Relating events in narrative, Vol.2: Typological and contextual perspectives. Mahwah, NJ: Lawrence Erlbaum Associate.

❑ Vallduví, E. 1992, *The informational component*, Garland.

❑ Vallduví, E., Engdhal, E.1996, The linguistic realization of information packaging, Linguistics 34.

❑ Villalba, X. 2007. La dislocació a la dreta en català i castellà, microvariació en la interfície sintaxi/pragmàtica, *Caplletra: revista internacional de filología* 42: 273-302.

Acknowledgments

We thank: Estela Puig Waldmüller, Teresa Suñol,
Josep Maria Fontana, Louise McNally,
Gemma Boleda, Alex Alsina