

NOCANDO

A multilingual annotated corpus for the study of Information Structure

Lisa Brunetti¹ Stefan Bott² Joan Costa³ Enric Vallduví³

¹Université de Provence, Aix-Marseille I

²Universitat Politècnica de Catalunya

³Universitat Pompeu Fabra

Mannheim, 24. September 2009

Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS
- 3 The NOCANDO corpus
- 4 Corpus exploitation
- 5 Conclusion and outlook

NOCANDO

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

- The project goal:
 - NOCANDO seeks to establish a cross-linguistically valid taxonomy of non-canonical constructions (NOCANS) and
 - study the relation between NOCANS and Information Structure (IS)
- 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 optimize the way the informational content of a sentence is conveyed (Vallduví 1992). NOCANS are explicit marks of the INFORMATION STRUCTURE of the sentence.

Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS**
- 3 The NOCANDO corpus
- 4 Corpus exploitation
- 5 Conclusion and outlook

Information Structure

Information Structure (a.k.a theme-rheme, focus-background, topic-comment, ...) is like a poltergeist:

- We all know that poltergeists and information structure exist
- but they are only observable through evidence, which is
 - indirect
and
 - sparse.
- Information-structural units (such as focus, topic, ...) cannot be directly observed.
- The *exact* inventory of these units cannot be determined by direct observation either (some categories are uncontroversial, others less clear).

Information Structure

What we know about Information Structure

- There is a long tradition of studies (which rely mainly on introspective data)
- Information-structural units (foci, topics, ...) are coded in natural language by means of prosody, syntax and morphology (NOCANS)
- IS highly depends on context
- Some registers (especially spoken text registers) are richer in IS marking than others

Information Structure

What we are interested in

- How do *different*, but related, languages mark IS units with different NOCANS (the crosslinguistic perspective)
- Study IS in spontaneously produced text
- IS marking in *context*

Information Structure

What we (at NOCANDO) assume about Information Structure

- Topic-comment distinction (Topic marks what the sentence is about)
- Focus-background distinction (The focus introduces new or contrastive information)
- There is also non-topic non-focus material (called *tails*, Vallduví 1992)
- Information-structural units are crosslinguistically stable, while their marking is variable

Examples of NOCANS and their function I

Clitic left dislocation

Al	hombre
to-the	man
Ind.	Object

"The man

TOPIC

(Spanish)

se	le	cae	el	café
RFL	to-him	falls	the	coffee
	obj.cl.	Verb	Subject	

drops the coffee"

COMMENT

Examples of NOCANS and their function II

Focus Fronting

Pure la LINGUACCIA,
even the tongue
 Dir. Object

"Even the tongue

FOCUS

(Italian)

gli fa, la rana
to-him he-puts-out the frog
 Verb Subject

did the frog put out to him"

BACKGROUND

Examples of NOCANs and their function III

Cleft sentences

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

"And she does not realize that the frog placed himself before the baby"

i és ELLA
and is her

"And it is her

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

who is about to take the baby bottle"

**CONTRASTIVE
FOCUS**

BACKGROUND

(Catalan)

Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS
- 3 The NOCANDO corpus**
 - Annotated NOCANS
 - Corpus distribution
- 4 Corpus exploitation
- 5 Conclusion and outlook

Quantitative 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)

Speakers

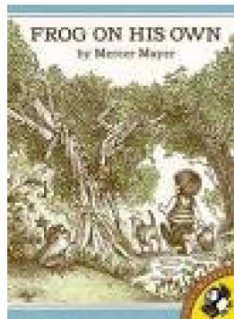
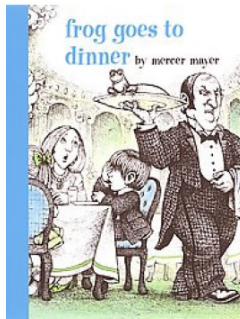
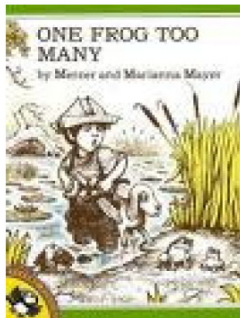
Mostly students at the Universitat Pompeu Fabra in Barcelona. A smaller number from different working environments.

	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)

Corpus compilation and 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, Strömquist & Verhoven 2004



a new frame

- 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.



Transcription and segmentation

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

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).

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:	(,) (.) (?) (!)

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 subj, SVO, Obj CI Pro..)
- Similar strategies to express informational notions:
 - large use of syntax
 - limited use of prosody (as opposed to e.g. 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)

Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS
- 3 The NOCANDO corpus
 - Annotated NOCANS
 - Corpus distribution
- 4 Corpus exploitation
- 5 Conclusion and outlook

Corpus annotation: Subject oriented NOCANS

- sbjinv = subject inversion (the subject occurs after the verb)
 - (1) Els va acompanyar el taxista.
them PAST take the taxi-driver
'The TAXI-DRIVER drove them' (Catalan)
- sbjinv_deacc = post-verbal deaccented subject in a declarative sentence
 - (2) ...que està disfressat, aquest nen.
for is dressed-up this child
'...for this child is dressed up' (Catalan)

Corpus annotation: Subject oriented NOCANS

- nsbj = null subject (the subject is not expressed)
 - (3) Invece il bambino è molto contento, perché ha salvato la sua rana.
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' (Italian)
- nsbj_c = null subject in a coordinate clause

Corpus annotation: Subject oriented NOCANS

- arbnsbj = arbitrary subject (which is null in these languages, cf. Jaeggli 1986)
 - (4) Y un día a este niño le regalaron pues una
and one day to this boy to-him they-gave well a
caja muy grande
box very big
'And one day this boy received a large box'
literally: 'And one day they gave the boy a large box
as a present' (Spanish)

Corpus annotation: Subject oriented NOCANS

- sbj-sep = separation of the (preverbal) subject from the verb by sentential adverbials or other material that has scope over the entire clause
- (5) però la rana come al solito riesce a
but the frog as the usual manages to
infilarsi nella_ nella situazione
sneak-herself into-the into-the situation
'but the frog, as usual, manages to sneak into the ...
the situation' (Italian)

Corpus annotation: Subject oriented NOCANS

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

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

- ld = left dislocation. Dislocation to the left WITHOUT clitic resumption

(7) A un bambino un giorno arriva un regalo.
to a boy one day arrives a present
'One day a boy receives a present'

Corpus annotation: Subject oriented NOCANS

- ht = hanging topic. A left dislocated NP resumed by a pronoun expressing its grammatical function. (cf. Chinque 1990, Benincà 2001, Zubizarretta 1999)
 - (8) La rana grande, la situación no le gustaba
the frog big the situation not to-her pleases
mucho.
much
'As for the big frog, she didn't like the situation at all'
(Spanish)

Corpus annotation: Subject oriented NOCANs

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

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

Corpus annotation: Subject oriented NOCANS

- rd = right dislocation. Dislocation to the right WITHOUT clitic resumption

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

Corpus annotation: Subject oriented NOCANS

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

(11) Entonces la tortuga lo ve y se lo dice
then the turtle it sees and to-him it says
al niño.

to-the boy

'So the turtle sees what happened and tells the boy everything' (Spanish)

Corpus annotation: NOCANS concerning all arguments

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

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

Corpus annotation: NOCANS concerning all arguments

- narg = null argument

(13) i llavors en Jaume es va adonar que
and then the Jaume RFL PAST realizes that
que, home, era la seva granota
that well it-was the his frog

Áand therefore Jaume ralizes that that, well, it was his
frog' (Catalan)

Corpus annotation: NOCANS concerning all arguments

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

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

Corpus annotation: NOCANS concerning all arguments

- deacc = de-accenting

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

Corpus annotation: NOCANS concerning sentence types

- pres = presentational sentences

(16) C'era una volta un bambino
there was one time a boy
'Once upon a time there was a boy' (Italian)

Corpus annotation: NOCANS concerning sentence types

- pass = passive construction

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

Corpus annotation: NOCANS concerning sentence types

- impers = impersonal construction

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

Corpus annotation: NOCANS concerning sentence types

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

(19) Era ese SAPO que les había querido hacer
was this frog that them had wanted make
algún susto
some scare

'It was the frog who wanted to scare them' (Spanish)

Corpus annotation: NOCANS concerning sentence types

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

(20) y lo que pasa es que el barquito se
and it that happens is that the boat REFL
hunde
sinks

'and what happens is that the boat sinks' (Spanish)

Corpus annotation: NOCANS concerning sentence types

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

(21) Y bueno, el niño es el que dirige la balsa
and well the boy is he that directs the raft
'and well the boy is the-one who leads the raft'
(Spanish)

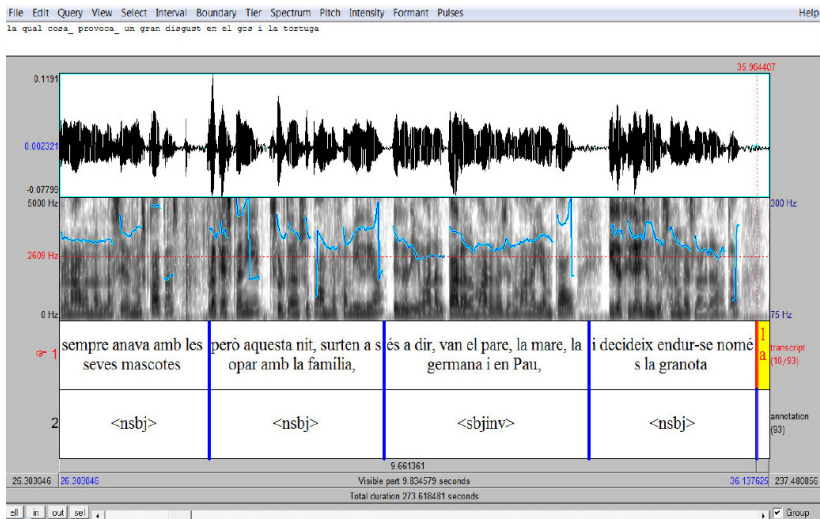
Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS
- 3 The NOCANDO corpus**
 - Annotated NOCANS
 - Corpus distribution**
- 4 Corpus exploitation
- 5 Conclusion and outlook

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="cldd,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>
```

Alignment audio-transcription-annotation



The corpus distribution

The NOCANDO corpus is available online under a Creative Commons license (<http://nocando.barcelonamedia.es>)



The screenshot shows a web browser window with the URL <http://nocando.barcelonamedia.org>. The page has a yellow header with the word "NOCANDO" in bold. On the left side, there is a vertical navigation menu with links for "Main", "Corpus", "Description", "Publications", "Collaborators", "Contacts", and "Links". The main content area is titled "The project" and contains the following text:

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](#).

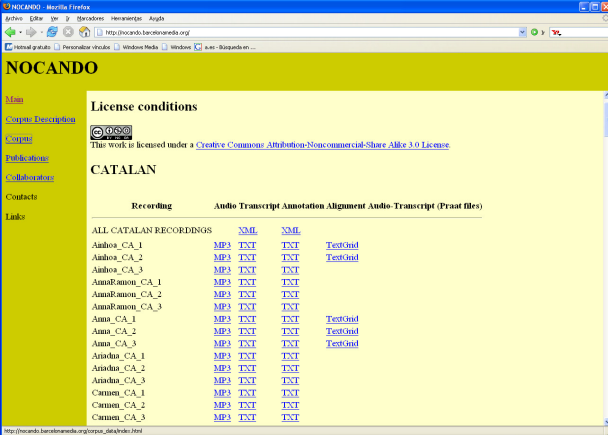
If you want to use our data, please quote this webpage, and respect the constraints indicated in the link below:

 This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 Spain License](#).

Terminado

The corpus distribution

The distribution consists of audio files (mp3), transcription, annotation and alignment files (when available)



The screenshot shows a web browser window displaying the NOCANDO website. The page title is "NOCANDO" and the URL is "http://nocando.babelionedia.org". The page content includes a navigation menu on the left, a "License conditions" section with a Creative Commons Attribution-NonCommercial-Share Alike 3.0 License logo, and a "CATALAN" section with a table of recordings.

License conditions

This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 3.0 License](#)

CATALAN

	Recording	Audio	Transcript	Annotation	Alignment	Audio-Transcript (Praat files)
ALL CATALAN RECORDINGS	XML	XML				
Ainhoa_CA_1	MP3	TXT	TXT			TextGrid
Ainhoa_CA_2	MP3	TXT	TXT			TextGrid
Ainhoa_CA_3	MP3	TXT	TXT			
AnnaRamon_CA_1	MP3	TXT	TXT			
AnnaRamon_CA_2	MP3	TXT	TXT			
AnnaRamon_CA_3	MP3	TXT	TXT			
Anna_CA_1	MP3	TXT	TXT			TextGrid
Anna_CA_2	MP3	TXT	TXT			TextGrid
Anna_CA_3	MP3	TXT	TXT			TextGrid
Ariadna_CA_1	MP3	TXT	TXT			
Ariadna_CA_2	MP3	TXT	TXT			
Ariadna_CA_3	MP3	TXT	TXT			
Carmen_CA_1	MP3	TXT	TXT			
Carmen_CA_2	MP3	TXT	TXT			
Carmen_CA_3	MP3	TXT	TXT			

Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS
- 3 The NOCANDO corpus
- 4 Corpus exploitation**
- 5 Conclusion and outlook

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 subj	1561	35,7 %	1262	38,9 %	1027	35,5 %
nsbj	1665	38,1 %	1173	36,1 %	1084	37,5 %
arbnsbj	22	0,5 %	7	0,2 %	32	1,1 %
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 %
cleft	3		4		2	
pscleft + inv-pscleft	40	0,9 %	10	0,3 %	37	1,28 %
pass	5	0,1 %	67	2 %	7	0,24 %

Comments

- 53,89% of all segments contain a NOCAN
- The overall frequencies for NOCANs in Romance languages are *surprisingly* similar

Comments

- There is a considerable variation in the use of different NOCANS among speakers of the same language. Some examples:
 - 3 out of 7 passives in Spanish (a marked construction) were produced by only 1 speaker
 - Some NOCANS, like pseudo-clefts and inverted pseudo-clefts were only produced by less than half of the speakers, but some speakers use them relatively frequent (in up to 0,65% of the segments in the case of psleft and inv-psleft).
 - The use of impersonal constructions ranges from 0 to 4,95% (all Romance languages)
 - Deaccenting in Italian is very rare in Italian, but one speaker used it very commonly (13 out of 38 deaccentuations were produced by this speaker)
- ⇒ Some NOCANS are subject to strong rhetorical variation. Others are more evenly distributed.

Comments

- Null subjects (nsbj) are nearly as frequent as overt subjects.
→ Null subjects are probably not non-canonical.

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

Comments

- Null subjects (nsbj) are nearly as frequent as overt subjects.
→ Null subjects are probably not non-canonical.
- Mayol (2009) concludes (on the basis of Catalan NOCANDo data) that the choice between null subjects and subject pronouns is partly dependent on IS, but partly the choice is dependent on purely syntactic factors (antecedent is subject or direct object).

Comments

- Subject inversion has a low percentage, but is still one of the most frequent NOCANS. The high frequency of subjinv in Spanish is partly due to syntactic effects. A postverbal subject cannot serve as a sentence topic (and is usually focal).

	Catalan	Italian	Spanish
--	---------	---------	---------

subjinv	332	7,6 %	215	6,6 %	265	9,1 %
----------------	-----	-------	-----	-------	-----	-------

Comments

- Subject inversion has a low percentage, but is still one of the most frequent NOCANS. The high frequency of subjinv in Spanish is partly due to syntactic effects. A postverbal subject can not serve as a topic and are usually focal.
- Mayol (2009) concludes that subjects are preferred as background material (mostly topics). She found that only 1,66% of the segments in the Catalan corpus are either postverbal or directly marked as focused. → There is a strong relation between subjects and topichood.
- 74,72% of the focal subjects in the Catalan corpus were discourse new (Mayol 2009).

Comments

- *Surprisingly*, and contra previous assumptions (e.g., Villalba 2007), Catalan and Spanish behave very similar with respect to right dislocations. This may be a consequence of the text genre.

	Catalan		Italian		Spanish	
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 %

Comments

- There is a clear difference between Catalan/Spanish vs Italian concerning clitic doubling: *cldbl* is considered as highly marked for colloquial speech in Italian. The difference between Italian and the other two languages is presumably more syntactic in nature than informational.

Catalan	Italian	Spanish
---------	---------	---------

cldbl	92	2,1 %	7	0,2 %	61	2,1 %
--------------	----	-------	---	-------	----	-------

Comments

- Passives are extremely more common in Italian than in Spanish or Catalan. In Spanish/Catalan passives are traditionally considered as a marked construction (recommended by some normative grammars only in restricted contexts). These languages use other constructions instead, like e.g. arbitrary subjects.

Catalan	Italian	Spanish
---------	---------	---------

pass	5	0,1 %	67	2 %	7	0,24 %
-------------	---	--------------	----	------------	---	---------------

Outline

- 1 The NOCANDO project
- 2 Information Structure and NOCANS
- 3 The NOCANDO corpus
- 4 Corpus exploitation
- 5 Conclusion and outlook

Conclusions

- The NOCANDO corpus is a valuable resource for the study of information structure in spoken register
- It is potentially valuable for a wider study of spoken register in the covered languages
- Some previous assumptions about IS and IS marking can be confirmed with quantitative data, some findings raise interesting questions
- The NOCANDO corpus is also a good source for spontaneous data for *qualitative* studies
- It allows to study IS marking in linguistic context
- The NOCANDO corpus can be both exploited and extended by third parties

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.

Acknowledgments

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

Thanks

Thank you very much!

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.

Selected references

- 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.
- Strömquist, S. and L. T. Verhoven (eds.) 2004, Relating events in narrative, Vol.2: Typological and contextual perspectives. Mahwah, NJ: Lawrence Erlbaum Associate.
- Mayol, Laia (2009) "Pronouns in Catalan: information, discourse and strategy" Ph.D. Dissertation. Advised by Robin Clark. University of Pennsylvania.
- 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 filologia 42: 273-302.