

A Romance perspective on subordination and fragments

Embedded Gapping in Spanish

Gabriela Bîlbîie¹ and Oscar García Marchena²

¹**University of Bucharest & LLF, Paris**

`gabriela.bilbiie@gmail.com`

²**Ecole Polytechnique & LLF, Paris**

Workshop *Experimental and corpus-based approaches to ellipsis*

LSA 2017 Linguistic Institute

29 July 2017, University of Kentucky

Embedding and ellipsis

- Most ellipsis studies are limited to coordination, considered as the preferred syntactic relation (if not the sole option) for many elliptical constructions.
- Some elliptical constructions seem to be preferred in coordination contexts (e.g. Gapping). Some others (VPE, RNR) may occur in subordination too, the target clause being embedded in the source.
 - (1)
 - a. Robert cooked the first course, **and** Mary the dessert.
 - b. *Robert cooked the first course, **because** Mary the dessert.
 - (2)
 - a. Joan write a novel, **and** Marvin did too.
 - b. Joan write a novel **after** Marvin did too.
 - (3)
 - a. You know a man who sells, **and** I know a man who buys, pictures of Elvis Presley.
 - b. It seemed likely to me, **though** it seemed unlikely to everyone else, that he would be impeached. (Chaves & Sag 2008)

Content

1 **Embedded Gapping**

2 Experimental data

3 Corpus data

4 Perspectives

The ban on embedded Gapping

- It is usually assumed (Hankamer 1979, Neijt 1979 and the subsequent literature) that Gapping is a root phenomenon : the gapped clause cannot be embedded within the conjunct it belongs to. Gapping \neq VPE (6).
 - (4)
 - a. *Alfonso stole the emeralds, and **I think that** Mugsy the pearls. (Hankamer 1979)
 - b. *Some had eaten mussels and **she claims that** others shrimp. (Johnson 2009)
 - (5)
 - a. *Bill went to Paris and **I think that** John to Rome.
 - b. Bill went to Paris and **I think that** John went to Rome.
 - (6) The adults had eaten mussels, and **she claims that** the children did too. (Farudi 2013)

- According to Johnson (2006, 2009, 2014), there would be a strong syntactic constraint on Gapping (and a diagnostic of this elliptical construction), called 'the No Embedding Constraint' (7). "Gapping is the only ellipsis process constrained by the No Embedding Constraint" (Johnson 2014 : 8).

(7) **The No Embedding Constraint** (Johnson 2014)

Let A and B be conjoined or disjoined phrases, and β be the string elided in B whose antecedent is α in A. Then α and β must contain the highest verb in A and B.

- The contrast in (8-a)-(8-b) is considered as strong evidence for a low (subclausal) coordination analysis of Gapping (i.e. a coordination of vPs and not a coordination of clauses) : either as a case of VP-ellipsis (Coppock 2001) or as a case of across-the-board movement of the shared verb out of each conjunct (Johnson 2009).

- (8) a. Alfonso stole the emeralds, and Mugsy the pearls.
b. *Alfonso stole the emeralds, and **I think that** Mugsy the pearls. (Hankamer 1979)

Counter-example 1 : Farsi data

- The first reported counter-examples come from Farsi (the variety of Persian spoken in Iran). Farudi (2013) observes that in Farsi gaps are possible under a wide range of embedding verbs and argues that the subordinating heads are not parenthetical, but syntactically integrated.
- (9) Māmā chāi xord va **fekr mi-kon-am** bābā qahve.
 mother tea ate.3SG and think IMPFV-do-1SG father coffee
 'Mother drank tea and I think Father (drank) coffee.'
- (10) Jiān be Sārā gol dād va **fekr mi-kon-am** ke Ārtur
 Jian to Sarah flower gave.3SG and think IMPFV-do-1SG that Arthur
 be Giti ketāb.
 to Giti book
 'Jian gave flowers to Sarah and I think that Arthur (gave) books to Giti.'
- (11) Mahsā in ketāb-ro dust dār-e va **Minu mi-dun-e** ke
 Mahsa this book-OBJ like have-3SG and Minu IMPFV-know-3SG that
 māmān-esh un ketāb-ro.
 mother-3SG that book-OBJ
 'Mahsa likes this book and Minu knows that her mother (likes) that book.'

Counter-example 2 : Romance data

- Furthermore, attested data from two Romance languages (Spanish (12) and Romanian (13)) show that 'the No Embedding Constraint' is too strong crosslinguistically.

(12) Pedro le regaló flores a María y **creo**
 Pedro DAT.3SG offered flowers to Maria and think.1SG
que Alicia unos libros.
 that Alicia some books
 'Pedro offered flowers to Maria and I think Alicia some books.'

(13) Nu EU îl urăsc pe el, ci **cred**
 NEG I ACC.3MASC.SG hate.1SG DOM him but think.1SG
că EL pe mine.
 that he DOM me
 'I don't hate him, but I think he hates me.'

- Novel data : empirical evidence from an acceptability judgment task and a corpus study on Spanish that embedded gapping is acceptable.

Not all predicates embed : a semantic constraint at work ?

- Previous work on Spanish embedded fragments (de Cuba & MacDonald 2013, Fernández-Sánchez 2016) insists on the crosslinguistic relevance of the semantic distinction (factive vs. non factive predicates) as a very strong constraint for fragment embedding : only non-factive verbs can embed fragments.
- Two lists of predicates (from de Cuba & MacDonald 2013) :
 - Predicates allowing embedding : *suppose, imagine, suspect, think, say...*
 - Predicates disallowing embedding : *regret, hate, love, know, find out...*

- (14) ¿Quién robó las joyas? 'Who stole the jewels?'
- a. – {Creo | supongo | me imagino | pienso} que tu hijo.
'I {think | suppose | imagine} that your son.'
 - b. – *{Lamento | sé | me sorprende | me desagrada} que tu hijo.
'I {regret | know | am surprised | dislike} that your son.'

- Fernández-Sánchez (2016) : embedded gapping cannot occur under factive predicates.

- (15)
- a. Alfonso robó las esmeraldas y {creo | imagino |
Alfonso stole the emeralds and {think | imagine |
supongo ...} que Mugsy las perlas.
suppose ...} that Mugsy the pearls
- b. *Alfonso robó las esmeraldas y {lamento |
Alfonso stole the emeralds and {regret |
me encanta | odio ...} que Mugsy las perlas.
love | hate ...} that Mugsy the pearls

- Factive vs. non-factive predicates (cf. Kiparsky & Kiparsky 1970, Karttunen 1971) :
 - Factive predicates presuppose the truth of their complement. They assign the status of an established fact to their object.
 - Non-factive predicates are not accompanied by a similar presupposition (they leave room for doubt and uncertainty).

- Two problems with the previous constructed data :
 - Most factive verbs on the previous lists are emotive factives (*lamentar* 'regret', *odiar* 'hate', *encantar* 'love', *desagradar* 'displease'...), which require the subjunctive mood in Spanish. So, we a priori don't know if their unacceptability with Gapping is due to embedding or rather to a mood mismatch (indicative/subjunctive).

(16) Alfonso robó las esmeraldas y **lamento que** Mugsy
 Alfonso stole the emeralds and regret that Mugsy
 { ?robó | **robara** } las perlas.
 { stole.IND | stole.SBJV } the pearls

- The remaining factive verbs (knowledge factives : *saber* 'know', *descubrir* 'discover', *observar* 'notice', *ver* 'see' ...) are attested in some embedded fragments (web and corpus data).

(17) A : ¿Fumas? B : Ya **sabes que** yo muy poquito.
 'A : Do you smoke? B : You know that I (smoke) just a little bit.'

- Because of these possible confounds, previous constructed data seem to be not very reliable.

Content

1 Embedded Gapping

2 Experimental data

3 Corpus data

4 Perspectives

- Causal relationship of interest :
 - Is there any interaction between Gapping and embedding? Hypothesis : Embedded Gapping should be acceptable (at least with some verbs).
 - Is there any interaction between Gapping and factivity? Hypothesis : only non-factive verbs can embed gapped clauses.
- Controlling for possible confounds :
 - Eliminate factive verbs which are incompatible with the indicative mood in Spanish.
 - Pay attention to the heterogeneous behaviour of factive verbs (Karttunen 1971, Hooper & Thompson 1973, Hooper 1974), and take into account a more fine-grained distinction (true factives vs. semifactives).
 - Facilitate the acceptability of Gapping by paying attention to the context factor. Many scholars (Kuno 1976, Prince 1986, Steedman 2000, etc.) observe that the acceptability of Gapping seems to be heavily dependent on the context ("Gappings are felicitous just in case they can be taken to instantiate an open proposition", cf. Prince 1986 : 212). ⇒ Each experimental item is introduced by a context sentence, which is a *wh*-question, giving rise to two contrastive pairs.
 - More importantly, taking into account the non-Gapping counterparts (full clauses, with verb repetition in the second clause).

Acceptability judgment test

- In order to check the two hypotheses presented above, we set up an experiment on acceptability judgments in Spanish, by using the Ilex Farm platform. Rating acceptability by using a 7-point Likert scale.
- Participants : 29 Spanish native speakers from age 19 to 49 (mean : 36) were retained (4 bilingual participants eliminated from the study).
- A within-subjects design : each participant sees every condition.
- A within-items comparison : each item set realizes all conditions.
- 4 types of items : instruction items (3), practice items (9), experimental items (24 sets), filler items (24 sets).
- A crossed factorial design (2x3 design) with two factors (GAPPING and EMBEDDABILITY), with two and respectively three levels, giving rise to 6 conditions.

Values for GAPPING factor	Values for EMBEDDABILITY factor
[+gapping]	[-embedding]
[-gapping]	[+embedding&non-factive]
	[+embedding&factive]

Experimental items

- (18) *¿Qué instrumento aprendieron estos dos niños en la escuela de música ?*
'What instruments did these two children learn in the music school ?'
- a. **[+gapping, -embedding]**
 María aprendió el piano y Angel la guitarra.
 'Maria learnt to play the piano and Angel the guitarre.'
- b. **[+gapping, +embedding&nonfactive]**
 María aprendió el piano y **me parece que** Angel la guitarra.
 'Maria learnt to play the piano and it seems that Angel the guitarre.'
- c. **[+gapping, +embedding&factive]**
 María aprendió el piano y **me sorprende que** Angel la guitarra.
 'Maria learnt to play the piano and I am surprised that Angel the guitarre.'
- d. **[-gapping, -embedding]**
 María aprendió el piano y Angel aprendió la guitarra.
- e. **[-gapping, +embedding&nonfactive]**
 María aprendió el piano y **me parece que** Angel aprendió la guitarra.
- f. **[-gapping, +embedding&factive]**
 María aprendió el piano y **me sorprende que** Angel aprendió la guitarra.

- Context sentences (*wh*-questions) before target conditions.
- Experimental items are created in lexically matched sets.
- Only transitive verbs in past tense forms ; the subject is an animate definite NP, the object is an inanimate theme (in order to avoid Differential Object Marking).
- 6 conditions ; 4 observations per participant per condition \Rightarrow 24 experimental items.

(19) *¿Qué indemnización obtuvo la pareja por el accidente de coche ?*
 'What compensation did the couple get for the car accident ?'
 La mujer recibió 35.000€ y (**{se rumorea que | veo que}**) el marido
 200.000€.
 'The woman received 35.000€ and (**{it is rumoured that | I see that}**)
 her husband 200.000€.'

(20) *¿Qué sorpresas prepararon los padres para el cumpleaños de la hija ?*
 'What surprises did the parents prepare for her daughter's birthday ?'
 La madre encargó un collar grabado y (**{me imagino que | me gusta que}**) el padre un reloj de marca.
 'The mother ordered a engraved necklace and (**{I imagine that | I like that}**) the father a big-brand watch.'

- Embedding predicates (Karttunen 1971) :
 - Non-factive verbs : epistemic and communication verbs (*creer* 'believe', *pensar* 'think', *suponer* 'suppose', *imaginarse* 'imagine', *parecer* 'seem', *sospechar* 'suspect', *decir* 'say', *rumorearse* 'report')
 - 'Factive' verbs :
 - True factives : emotion verbs (*sorprender* 'be surprised', *impresionar* 'be impressed', *gustar* 'like', *molestar* 'be bothered', *horrizar* 'be horrified')
 - Semi-factives : knowledge verbs (*saber* 'know', *ver* 'see', *observar* 'observe')

Filler items

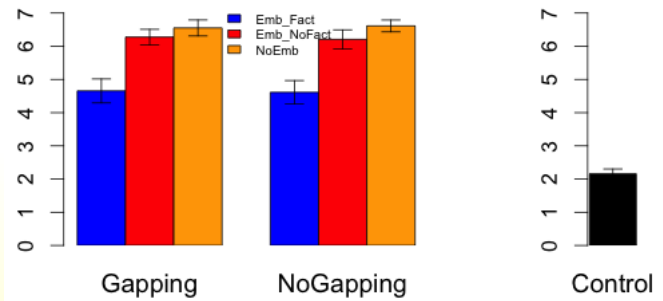
- 24 items (2x2 design)
- We've slightly modified another experiment in order to serve as filler for our experiment. We modified two of the four conditions, in order to have as control some ungrammatical items (agreement errors, tense errors, etc.).

(21) *¿Cuándo golpeó Pedro a Juan ?*
'When did Pedro hit Juan ?'

- a. Pedro golpeó a Juan cuando era joven.
'Pedro hit Juan when (he) was young.'
- b. Pedro golpeó a Juan cuando **seremos** joven.
- c. Fue Pedro quien golpeó a Juan cuando era joven.
- d. Fue Pedro quien golpeó a Juan cuando **seremos** joven.

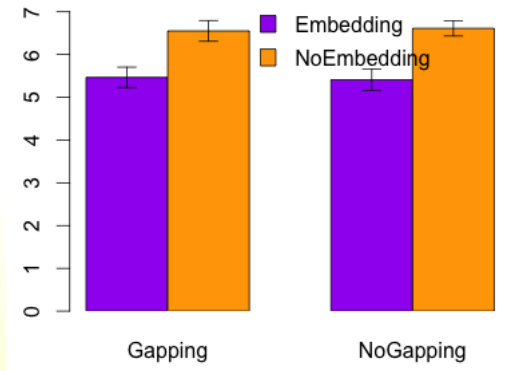
Results : general overview

- Embedded Gappings are ok and they are acceptable in the same way as their embedded full counterparts. Clear contrast with ungrammatical control items.
- Sensitivity to the semantic type of the embedding predicate in both Gapping and non-Gapping contexts.
- No difference between the use of Gapping and the use of the full clause.



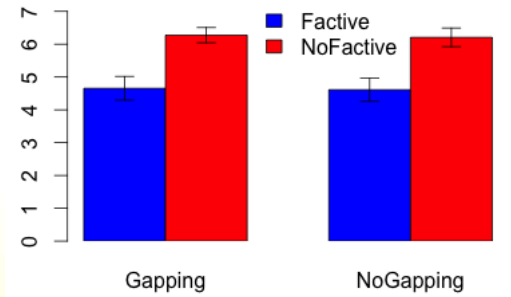
Results : effect of embedding

- Embedded configurations (mean rate 5.4) are less acceptable than non-embedded ones (mean rate 6.5), but much more acceptable than ungrammatical controls (mean rate 2).
- There is a penalty for embedding (in particular, with factive predicates), but it is not related to Gapping.



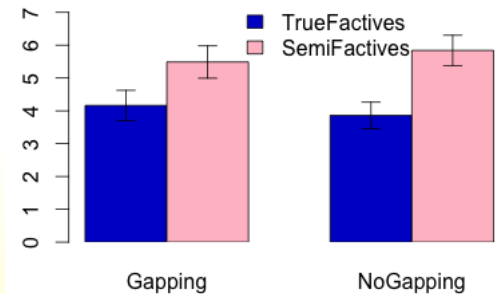
Results : effect of factivity

- Embedded clauses under a factive verb (mean rate 4.6) are less acceptable than embedded clauses under a non-factive verb (mean rate 6.2).
- There is a penalty for embedding under factive predicates, but it is not related to Gapping.



Results : zoom into factive verbs

- Semi-factives and true factives don't give rise to the same acceptability judgments!
- Embedded clauses under a true factive verb (mean rates : 4.1 Gap and 3.8 NoGap) are less acceptable than embedded clauses under a semi-factive verb (mean rates : 5.5 Gap and 5.8 NoGap).
- There is a penalty for embedding under true factive predicates, but again it is not related to Gapping.



- We analysed the general results by using a mixed-effects linear regression and by specifying the "maximal" random effects.
- The dependent variable is the acceptability judgment, predictor variables are \pm Gapping and Embeddability (no embedding, embedding factive, embedding non-factive).
- There is no significant difference between Gapping and non-Gapping cases.
- We see a significant effect of the Embeddability (non-embedding vs. embedding cases) and of the Factivity (factive vs. non-factive).
- There is no interaction between Gapping and Embeddability. There is no interaction either between Gapping and Factivity.

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	5.821733	0.215437	34.470000	27.023	< 2e-16 ***
Gap_NoGap	-0.005535	0.104146	36.750000	-0.053	0.958
NoEmb_Emb	0.822235	0.109311	29.330000	7.522	2.53e-08 ***
NoFact_Fact	0.370437	0.048238	56.890000	7.679	2.37e-10 ***
Gapp_NoGap : NoEmb_Emb	-0.013896	0.118389	39.820000	-0.117	0.907
Gapp_NoGap : NoFact_Fact	0.011357	0.067707	52.310000	0.168	0.867

- When doing experimental work on ellipsis, we have to compare elliptical occurrences with non-elliptical ones (full clauses), in order to better control our factors and to rule out other explanations (confounds) for any treatment effect that we see.
- What has been claimed to be specific to Gapping is in fact more general.
- Embedding is acceptable in Gapping and it follows some general constraints, applying outside ellipsis too.
- Factive predicates don't behave the same, confirming the dichotomy proposed by Hooper (1974) between semi-factive verbs (e.g. *find out*, *know*, *see*, etc.) and true factive ones (e.g. *regret*, *forget*, *amuse*, etc.) : the semi-factive verbs are much closer to non-factives than to true factive verbs.
- These classes of predicates were taken into account to explain Main Clause Phenomena in English (Hooper & Thompson 1973) : factives are bad, semi-factives are good, and non-factives are good too.

Content

1 Embedded Gapping

2 Experimental data

3 Corpus data

4 Perspectives

Spanish CORLEC Corpus

- In order to better understand the speakers' preferences between conditions (a) and (b), we compare our experimental findings with the naturally occurring data from the CORLEC Spanish Corpus (*Corpus Oral de Referencia del Español Contemporáneo*, Marcos-Marín 1992).
- The 1,078,780 words and 63,291 utterances of the CORLEC corpus are organized in 17 genres (8 monologic and 9 dialogic) and cover a variety of contexts from informal conversation to university lessons, which assures a diversity of registers, situations and speakers. It is also annotated for dysfluency, which allows us to distinguish ellipsis from dysfluency cases.
- 734 embedded verbless utterances, divided in :
 - 153 embedded fragments (most of them, with one remnant ; very few cases with two remnants)
 - 581 embedded verbless clauses

- CORLEC Spanish Corpus furnishes a few data where Gapping appears indeed in embedding contexts (García-Marchena 2015), with both non-factive and semi-factive predicates, but not with factive ones.
- Interestingly, most of our corpus data with embedded gapping have a high frequency of true/false and epistemic predicates, but also they involve some symmetric/reciprocal relation between the source clause and the gapped clause (22-a) and/or a reinforced semantic contrast (marked by 'adversative' conjunctions such as *pero* 'but' (22-b) or by negative remnants (22-c)).

- (22) a. Pero el chico la ama y **dicen que** ella a él. (CONV 033A)
 'But the boy loves her and they say she him.'
- b. Ella se lo va a comer todo pero **me parece que** yo solo un poco. (CONV 012A)
 'She is going to eat everything but I think I only a bit.'
- c. Luisa ha estado en ese club muchas veces pero **por supuesto que** yo nunca. (DEB 026A)
 'Luisa has been in that club many times but of course me never.'

- Preference for symmetrical relations and reinforced semantic contrast in Romanian too :

- (23) a. Nu EU îl urăsc pe el, ci **cred** **că**
 NEG I ACC.3MASC.SG hate.1SG DOM him but think.1SG that
 EL pe mine.
 he DOM me
 'I don't hate him, but I think he (hates) me.'
- b. Ion o iubește pe Ana și **văd** **că** și ea
 Ion ACC.3FEM.SG loves DOM Ana and see.1SG that too she
 pe el.
 DOM him
 'Ion loves Ana and I see that she (loves) him too.'
- c. Ion este îndrăgostit de Ana, **nu știu însă dacă** și ea de el.
 'Ion is in love with Ana, but I don't know if she (is in love) with him too.'

- These usage preferences are quite expected, since gapping, unlike other fragments, imposes very strong semantic and discursive constraints (Hartmann 2000, Kehler 2002), namely there must be a semantic contrast and a discursive symmetric relation between the source and the gapped clause.

Content

- 1 Embedded Gapping
- 2 Experimental data
- 3 Corpus data
- 4 Perspectives**

Conclusions

- The 'No Embedding Constraint' postulated by Hankamer (1979), Neijt (1979), Johnson (2009, 2014), etc. and considered to be a strong syntactic constraint specific to Gapping constructions must be released.
 - The 'No Embedding Constraint' can no longer be used as a diagnostic of Gapping.
- The claim made on embedded gapping (Fernández-Sánchez 2016) needs to be reformulated : "embedded gaps are sensitive to the semantic kind of predicate under which they are embedded" ⇒ **embedded clauses in general** are sensitive to the semantic type of embedding predicate.
 - Some predicates embed clauses better than others. ⇒ An overlap with the findings made by Hooper & Thompson (1973) wrt Main Clause Phenomena in English.

- Possible explanations for the gradience observed with embedding wrt the semantic type of predicate :
 - ⇒ A **pragmatic** principle (cf. Green 1976 for MCP in English) : embedding is more acceptable with predicates indicating the speaker's agreement with the content of the embedded clause. In other words, embedding is facilitated when the speaker desires to be understood as committed to the truth of the subordinate clause.
 - ⇒ A **semantic** principle (cf. Hooper & Thompson 1973) : embedded assertions are more acceptable than embedded presupposed clauses (factive verbs are presupposition triggers). It is inappropriate in language to emphasize backgrounded or already known material.

Analysis of embedded Gapping

- Given that embedded Gapping is acceptable, a low (subclausal) coordination analysis of Gapping (Coppock 2001, Johnson 2009) should be rejected.
- In order to deal with embedded Gapping in Spanish, Fernández-Sánchez (2016) posits two different syntactic analyses :
 - A low coordination analysis à la Johnson for 'genuine' (non-embedded) Gapping.
 - An analysis in terms of clausal ellipsis for embedded Gapping.
- However, these new data are not problematic for a fragment-based analysis (with semantic reconstruction of ellipsis), as proposed by Abeillé et al. (2014) for Gapping in Romance, and by Ginzburg & Sag (2000) for fragments in general.

- Syntax-pragmatics-semantics interface :
 - Syntactically, a fragmentary clause is embedded under a main clause (so, no weak verbs!).
 - Pragmatically, the embedding predicate in the main clause has a discursive function, i.e. an evidential marker (indicating the speaker's grounds for asserting the complement, cf. Simons 2007).
 - Cf. high frequency of the 1st person corroborated with the high frequency of epistemic predicates : the speaker may take on some degree of commitment to the truth of the embedded fragment, leaving room for doubt and uncertainty.
 - Semantically, it is not clear if at the semantic level the main clause has a parenthetical use, i.e. if it is the content of the embedded clause that is the main content of the utterance (cf. Hooper 1974).

Key take-away points

- We expected that embedded Gapping might be acceptable. Indeed, it is acceptable.
 - We expected that there would be an embedding sensitivity wrt the semantic type of the embedding predicate. Indeed, there is an effect giving rise to a gradience in acceptability judgments.
 - We expected that these effects would be related to Gapping. However, they are not correlated with ellipsis, but they reflect more general preferences.
- ⇒ Further experimental work must be done for other languages too in order to show that the existence or non-existence of embedded Gapping is not related to this specific ellipsis type, but rather to a more general behaviour of that language concerning embedding.
- ⇒ In order to better understand the constraints applying to ellipsis phenomena, it is time to devote more attention to experimental and corpus studies than pure theoretical ones.

Selective references

- de Cuba, C. & J. MacDonald** 2013. On the Referential Status of Embedded Polarity Answers in Spanish. *Selected Proceedings of the 16th Hispanic Linguistics Symposium*. Somerville, MA : Cascadilla Proceedings Project.
- Farudi, A.** 2013. Gapping in Farsi : a crosslinguistic investigation. PhD thesis. Univ. of Massachusetts at Amherst.
- Fernández-Sánchez, J.** 2016. Embedded gapping isn't gapping. Manuscript, Universitat Autònoma de Barcelona.
- Hankamer, J.** 1979. *Deletion in coordinate structures*. New York : Garland Pub.
- Johnson, K.** 2006. Gapping. In M. Everaert & H. van Riemsdijk (eds.), *The Blackwell Companion to Syntax*. Oxford : Blackwell. 407-435.
- Johnson, K.** 2009. Gapping is not (VP-) ellipsis. *Linguistic Inquiry* 40, 289–328.
- Johnson, K.** 2014. Gapping. Manuscript, Univ. of Massachusetts at Amherst.
- Karttunen, L.** 1971. Some Observations on Factivity. *Papers in Linguistics* 4, 55-69.
- Hooper, J. & S. Thompson** 1973. On the Applicability of Root Transformations. *Linguistic Inquiry* 4(4), 465–497.
- Marcos-Marín, F.** 1992. Corpus de referencia de la lengua española contemporánea : Corpus oral peninsular. Technical report, Universidad Autónoma de Madrid.

Thank you !

We are very grateful to :

- Anne Abeillé (useful insights)
- Aixiu An (data analysis)
- Brian Dillon (R issues) : our experiment developed along with the *Experimental Syntax* LSA course
- Israel de la Fuente (fillers for our experiment)

This research is conducted within the LabEx EFL project, strand 2 *Experimental syntax*, operation SA3-IS2 *Full vs. elliptic clause alternation* (coord. A. Abeillé, F. Mouret).