On Intervention effects in weak islands. A self-paced reading experiment.

VINCENZO MOSCATI
Macquarie University, ARC Centre of Excellence in Cognition and its Disorders (CCD)
vincenzo.moscati@mq.edu.au

Filler-gap dependencies are created when a constituent is dislocated from its base position. A prototypical example is given by A'-movement in *wh-* questions. In this case, a clause-initial *wh-* element has to be linked to its corresponding *gap* in a lower, c-commanded position. Different factors might influence the human parser in resolving filler-gap dependencies, as i. the properties of the filler and ii. the presence of an *intervener*. In this paper, I present the result of a new self-paced reading experiment in which a particular kind of intervention effect, i.e. the one created by the sentential negative maker, will be observed in relation to different types of wh-constituents, distinguished in accordance to the oppositions +/- *argument* and +/- *referential*.

1. Introduction

A'-movement typically creates a filler-gap dependency, in which the *filler* is the dislocated constituent and the corresponding *gap* is the position where the same constituent would appear in absence of movement. Echo questions can be used to help us to detect the position of the *gap* (1), whereas a constituent has been dislocated (2):

(1) *What* did you say <t> to John ?
(2) You said *what* to John ?

For what concerns the possible span of a filler-gap relation, A'-movement is potentially unbounded in length and it may extends over several clause boundaries, as (3) below shows:

(3) What did Harry say that Tom thought that Mary was hiding <t> ?

However, this movement is not unconstrained and long distance dependencies become impossible across a range of interveners. One example is given by adjunct clauses:

(4) a. You skip the class because you needed to do *what*?
b. **What* did you skip the class because you needed to do <t> ?

While (4)a is grammatical when the interrogative pronoun is left in its base position, the attempt to move it outside the adjunct clause (4)b generates ungrammaticality. Violations as the one in (4) are usually labeled as *strong islands*, for the reason that they are insensitive to the properties of the *filler*, i.e. the kind of extracted *wh-* element:

(5) a. **Where* did you skip the class because you needed to go <t> ?
b. **How much* did you skip the class because you needed to earn <t> ?

Strong islands constraints have been extensively studied both from the theoretical and the psycholinguistic point of view, for the reason that they could provide important information about the functioning of the human parser. In the last twenty-five years, various real-time measures as self-pace reading (Stowe 1986, Pickering et al. 1994, Phillips 2006) and event-related potentials (Kluender & Kutas 1993, McKinnon & Osterhout 1996, Neville et al. 1991) have been employed, all of them focusing on filler-gap effects in strong islands.

There are, however, other kinds of island-effects that have been kept distinct from the previous ones since they are modulated in accordance to the referential properties of the filler. For this reason, they are usually referred to as *weak islands*. I illustrate this point by considering a particular type of weak islands, the ones in which the presence of a negative operator generates an effect of ungrammaticality (Ross 1967/83; Obenauer 1984). On a par with strong islands, the extraction of a wh-constituents from (6) and (7) is grammatically marked:

(6) **How* did you not behave?
(7) **How much beer* did you not drink <t> ?
(8) *Which beer* did you not drink <t> ?

What is interesting about weak islands is the fact that this kind of violation is not rigid but it varies as a function of the extracted *wh-* constituent. Speakers usually find (8) more acceptable than (7). This asymmetry has been accounted for either in syntactic (Rizzi 1990) or semantic terms (Szabolczi & Zwarts 1993, Fox and Hackl 2007, Abrusan & Spector 2011) capitalizing on the interaction between the negative operator and the reference of the extracted constituent.

However, in spite of the great attention that weak islands received in the theoretical literature and the existence of many psycholinguistic studies on strong islands, weak islands still relatively unexplored from a psycholinguistic point of view. In particular, the effects of the intervener on parsing have not been substantiated by real time studies. In this paper, I'll present the results of a new self-pace reading experiment. My goal, here, is to establish when (and if) intervention effects are detectable during the processing of filler-gap dependencies.

2. A syntactic account of negative islands

As I already pointed out, the core facts related to the difference in acceptability between (7) and (8) have been captured in various ways, invoking either a syntactic or a semantic-based explanation. Disentangling and testing the predictions of the two families of accounts is beyond the purposes of this work and I'll frame negative islands phenomena in their early syntactic formulation (Rizzi 1990). Consider again (6), (7) and (8) repeated below:

(9) *Which beer*$_i$ did you not drink <t_i> ?
(10) **How* did you not behave <t>?
(11) **How much beer* did you not drink <t> ?

According to Rizzi's analysis, the crucial difference between (9) on one side and (10)-(11) on the other, relies on the availability of *binding*: while in (9) the *wh*-element and its trace can be co-indexed, the same mechanism is excluded in (10) and (11). This follows by assuming the existence of two conditions active on indexing:

(12) *Condition 1 on indexing*: a referential index must be licensed by a referential theta role.

(13) *Condition 2 on indexing:* the assignment of referential indices is limited to (wh-) phrases which "refer to specific members of a pre-established set"

Sentence (9) satisfies *condition 1*, given that the trace might have an index since it is the internal argument of the verb. Moreover, it is quite easy to imagine (and accommodate) a situation in which a certain set of beer brands is given (*condition 2*). Sentence (10), instead, violates *condition 1*, given that the trace is not theta-governed. This condition is instead satisfied in (11). However, in (11), the *wh*- can hardly[1] pass the requirement imposed by *condition 2*, for the reason that a pre-established set is not easily available for interrogative pronouns denoting quantities. The only other available mechanism able to license the traces in (10) and (11) is antecedent-government, where the restriction in X α-governs Y only if there is no Z such that: applies:

(14) X α-governs Y only if there is no Z such that:
(i) Z is a typical potential antecedent governor for Y
(ii) Z c-commands Y but does not c-command X

However, negation (Z) constitutes a potential antecedent, blocking the link between the *wh*- elements and their traces. For this reason, antecedent government is also excluded and sentence (11) is ungrammatical.

A question which we may want to ask, at this point, is whether this mechanism of intervention has any psychological reality. In particular, an effect should be visible on the position where the filler-gap link is hindered. In our terms, on the position of the intervener Z - the negative operator. In addition, an effect could be also visible at the gap site, where the filler stored in memory has to be integrated in the structure. These issues will by addressed in a new experiment

[1] Accordingly, the sentence is acceptable if discrete quantities are salient in the discourse.

based on reading times, which will be presented in section 4. In the next section, I'll review some of the relevant data on island effects, as reported in previous studies based on the self-paced reading paradigm.

3. Reading times: trace integration and island effects

Before turning to islands, a preliminary question concerns the functioning of the parser when a filler is encountered. Broadly speaking, a *filler-gap* dependency is created every time a dislocated *wh-* element has to be integrated in the syntagmatic structure. This means that the *wh-* needs to be stored in memory, at least until the corresponding gap position is met. Thus, the parsing algorithm should first, keep track of the wh- element and second, be able to generate gaps in the appropriate structural positions.

This relation can be thought as imposing processing costs on the computation, since the filler has to remain active until a suitable gap is generated. However, different strategies could be employed in order to minimize these costs. One of them is the early integration of the filler. According to this idea, the parser will try to integrate the filler as soon as possible. Evidence in support of a parsing strategy of this sort comes from several reading times studies (Crain & Fodor 1985, Stowe 1986). In particular, reaction times seem to increase whenever an overt constituent is encountered in a position where a trace could instead be generated. This phenomenon has been interpreted by assuming that the parser always tries to generate traces, in order to minimize the processing costs. However, if an overt constituent is encountered, instead of a trace, the parser has to revise its strategy and reanalyze the structure. This would result in an increase in reaction times.

Now let's move one step further and consider the syntactic factors that influence the parser's decisions. More specifically, we want to know whether the syntactic (or semantic) constraints which generate islandhood could affect the parsing strategy. If the parser is sensitive to island constraints, it should not hypothesize traces within islands and, by reflex, no increase in reading times should be observed. This seems to be the case, as documented in Stowe (1986), Pickering et al. (1994) and in Phillips (2007). To illustrate, consider the two sentences (15) and (16), both presenting a subject island:

(15) The teacher asked what ^{SUBJ}[the silly story about (*<t>) Greg's older brother] was supposed to mean <t>
[Stowe 1986]

(16) The school superintendent learned which schools ^{SUBJ}[the proposal that expanded (*<t>) drastically and innovatively upon the current curriculum] would overburden <t> during the following semester
[Phillips 2006}

In both sentences, the first potentially available gap site is within an island. If the parser initially is insensitive to this kind of grammatical constraints, it is expected to generate traces within the subject island. As a result, a slow-down in reaction times should be observable. However, the aforementioned studies showed that there is no evidence of a slow down, suggesting that the parser does not hypothesize gap positions within islands. This supports the idea that the islands constraints are directly encoded within the parser's syntagmatic rules.

Let us now look at negative islands. Is this kind of violation also encoded in the parsing algorithm? Following the previous logic, if the syntactic constraints underlying negative islands are built-in, we expect that the parser will not hypothesize traces within islands of this sort. However, it is also possible that negative islands are quite different from strong islands and that the structure is filtered out at later stages of processing. In the experiment presented in the next section, we address this issue by timing the subjects' reactions at possible gap sites within negative islands.

One last point worth to be mentioned is the fact that, for what concerns the intermediate positions between the filler and the gap, namely potential sites for intervention, an increase in reading times has never been attested, neither in strong islands nor in weak islands. For this reason, it is worthwhile to look also at intermediate positions: if an account along the lines sketched in the previous section (RM, Rizzi 1990) is on the right track, the interaction between a non-indexed *wh*-element and sentential negation should result in a measurable increase in reaction times.

4. Wh- types and intervention: a self pace reading experiment

On-line experimental procedures, as reading times, can be employed in the study of the intervention effects previously discussed. In particular, we are interested in measuring these effects in relation to at least three different types of *wh*- elements:

(17) a. *why* [- argument, - index]
b. *how much*[2] N [+ argument, - index]
c. *which* N [+ argument, + index]

On the basis of the discussion in section 2, the three interrogative pronouns in (17) are expected to show a different behavior in positive and negative sentences. Let us consider first w*hy* in (17)a. This element is plausibly base-generated in the left-periphery of the clause (Rizzi 2001) and, for this reason, the filler-gap dependency is at best very short-lived. This means that no gap is expected after the lexical verb:

(18) *why* <t> John is(n't) drinking at the party?

In (19), instead, *how much* and *which* serve as the direct objects of the verb. This means that they can create a proper filler-gap dependency between the clause initial position and the internal argument position:

(19) *how much*/*which* juice John is(n't) drinking <t> at the party?

[2] I'll refer to these elements as non-indexed. This label is purely descriptive and it only expresses the fact that, in absence of a discursive context, it is more demanding for the reader to accommodate a plausible referent-set.

In positive sentence, no difference is expected between *how much* and *which*, since no intervener is present. However, in negative sentences, a weak island is created by negation. In this case, the different referential properties of the two *wh-* elements in (17)a-b might play a role and a grater slow down at the intervener site is expected in the case of *how much*. Moreover, if the intervener blocks the filler-gap chain, traces should not be generated in the direct object position. The prediction, in this case, is that no difference in reaction times should be observed between *how much* and *why* on the lexical verb.

In order to verify these predictions, the interaction between the sentence polarity (positive, negative) and the different kinds of *wh-* elements in (17) will be investigated with a self paced reading procedure based on the *stationary-window* paradigm (Just et. al 1982).

Materials and procedure.

Participants sat in front of a computer screen and were asked to read a series of sentences. Words unfolded stepwise at the center of the monitor and, at the end of each sentence, participants were asked to rate it on a scale from 1 to 7, according to the perceived naturalness of the sentence.

Since we were interested in observing the effect of negation in relation to the extracted *wh-* element, the stimuli were grouped into six conditions, in a 2 (Polarity) X 3 (Wh_Type) factorial design (table 1). Subjects heard 20 sentences per condition, for a total of 120 sentences.

The *wh-* elements appeared at the beginning of an embedded clause, in order to avoid the sentence-initial position. Embedded clauses were half of the times positive and half of the times negative. Four different verbs were chosen (*wonder*, *find out*, *want to know* and *discovered*) for the matrix clause and 20 different lexical verbs appeared in the embedded. All of them were optional transitive verbs. This made both wh- adjuncts (*why*) and arguments (*which/how much*) natural in the context.

Sixty fillers (tab 2) were interspersed within the test sentences and, in total, subject read and rated 180 sentences (120 test + 60 fillers), divided in three blocks of 60 sentences each. Within each block, the presentation was randomized.

Table 1. Conditions used in the self-paced reading experiment.

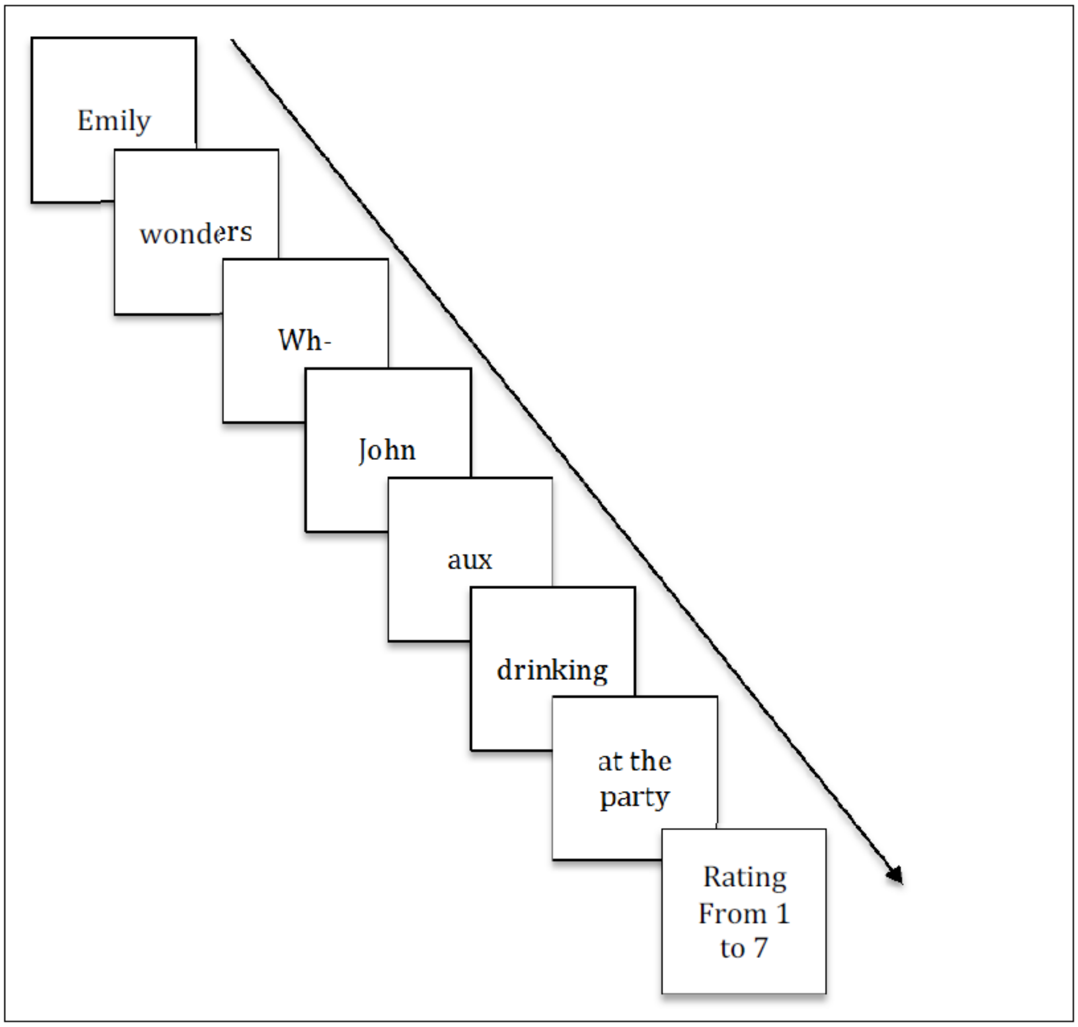
Conditions	Wh-Type	Polarity	Example
(1)	why	pos.	… why John is drinking at the party
(2)	which	pos	… which beer John is drinking <t> at the party
(3)	how much	pos.	… how much beer John is drinking <t> at the party
(4)	why	neg.	… why John isn’t drinking at the party
(5)	which	neg.	… which beer John isn’t drinking <t> at the party
(6)	how much	neg.	… how much beer John isn’t drinking <t> at the party

Table 2. Fillers.

Conditions	Example
(a)	Adams knows that the lawyer is helping the defendant
(b)	Frank thinks that the politician isn't keeping the promise
(c)	Ted is away while the manager is firing the employee

Figure 1 represents the timeline of the stimuli. Three different constituents could appear in the *wh*-position (*why, which N, how much N*) and two different auxiliary forms alternate in the aux position (*is, isn't*).

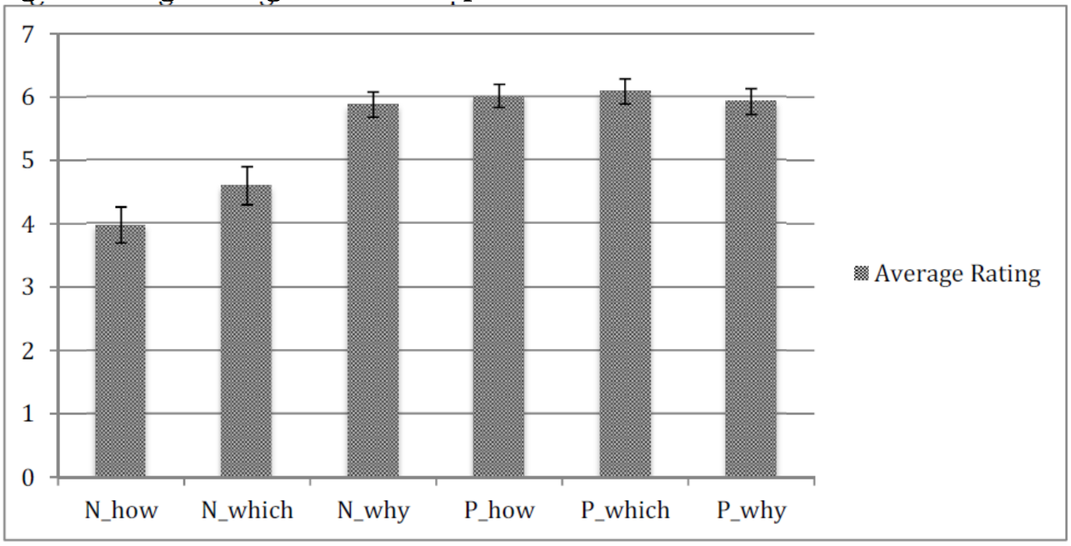
Fig. 1. Self paced reading: sequence of presentation.



Subjects.
21 adult subjects, native Australian English speakers, took part to the experiment. They were all undergraduate students at Macquarie University.

Results.
Let us consider first the off-line judgments given at the end of each sentence. Results are visually reported in figure 2, where ratings are averaged across all the subjects. As expected, the lowest ratings are obtained with non-indexed *wh*-elements (*how much*) in negative sentences. Also in the negative *which* condition, subjects found the sentences marginally acceptable. In all the other conditions, acceptability judgments raised at almost the same level, including *why* in negative sentences.

Fig.2. Average ratings in the six experimental conditions.



Data were analyzed with a 2(polarity) x 3(wh_type) repeated measure ANOVA. The analysis revealed a significant main effect of Polarity (F(20)=66.158, p<.001) and Wh_Type (F(40)=27.050, p<.001). The interaction between Polarity and Wh_Type was also significant (F(40)=44.614, p<.001). Post-hoc (Bonferroni) comparisons revealed that the difference between *which* and *how*, in negative sentences, was also significant (p<.01). These results are consistent with the judgments predicted by the analysis presented in section 2.

We turn now to the on-line data. Reading times were normalized by filtering out the outliers. Individual value greater than the *mean + 2 standard deviations* were substituted by the value *mean + 2sd*. The average reading times, for each segment in each condition, are reported in Table 3.

Tab. 3. Reaction times per conditions

Position	N-how		N-which		N-why		P-how		P-which		P-why	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
RT1	.5181	.0164	.5423	.0191	.5214	.0184	.5341	.0182	.5225	.0175	.5574	.0195
RT2	.5409	.0116	.5383	.0110	.5374	.0116	.5559	.0132	.5194	.0117	.5327	.0127
RT3	.6586	.0172	.5852	.0138	.4847	.0091	.6320	.0146	.5634	.0121	.4947	.0119
RT4	.5845	.0143	.5616	.0117	.5078	.0108	.5612	.0116	.5521	.0108	.5058	.0120
RT5	.5869	.0155	.5476	.0120	.5234	.0130	.5078	.0101	.5066	.0101	.4889	.0088
RT6	.7107	.0215	.6792	.0212	.5697	.0148	.5904	.0172	.6199	.0194	.5502	.0131
RT7	.8362	.0255	.8007	.0233	.8036	.0243	.7342	.0213	.7405	.0227	.7818	.0233

Let us consider separately the reading times for positive (figure 3) and negative (figure 4) sentences, looking at each distinct position of the embedded clause.

Fig. 3. Reading times in the positive conditions

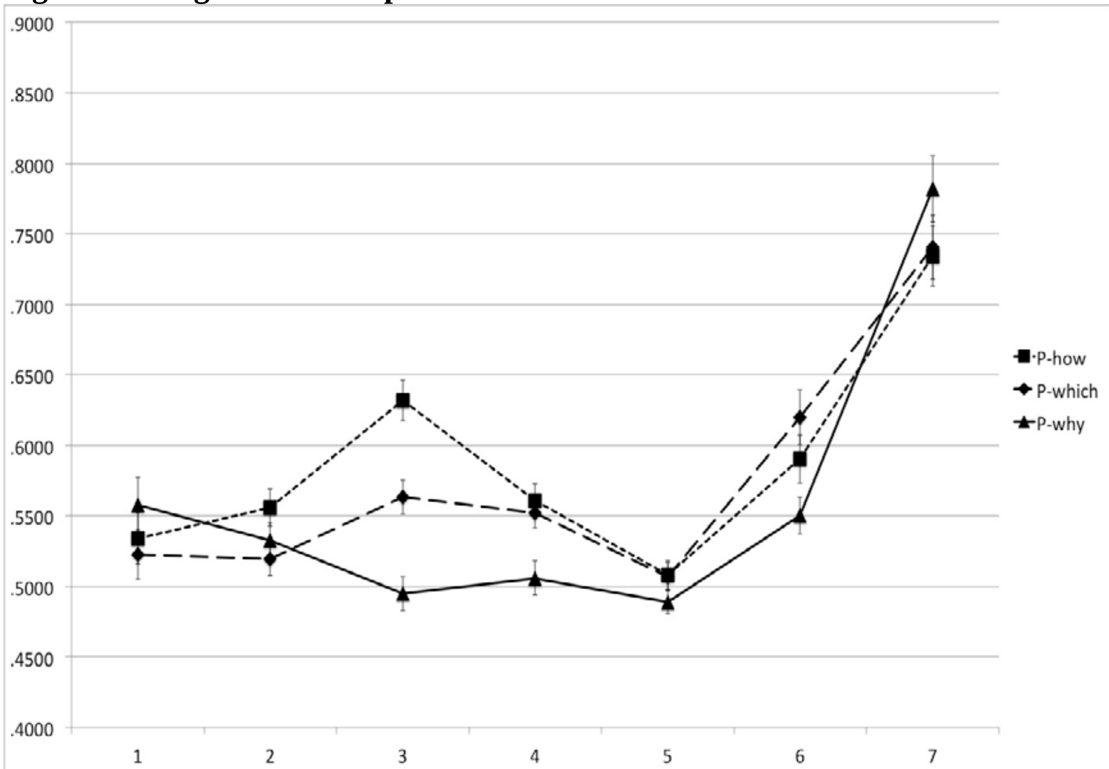
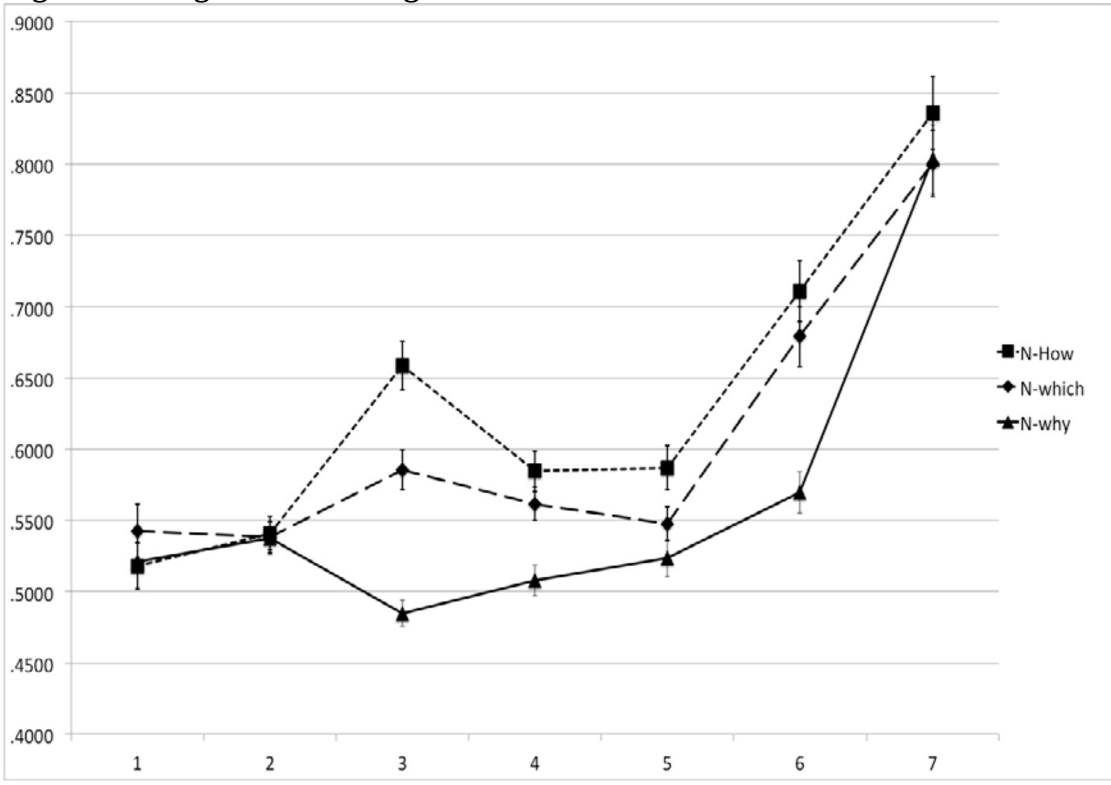


Fig. 4. Reading time in the negative conditions



In the first two positions (POSITION 1 and 2), reading times are relative to the subject and to the matrix verb. There is no visible difference between the various conditions and they are of little interest here. For this reason, we move directly to POSITION 3: where the different *wh-* elements appear and the embedded clause begins. For each constituent of the embedded clause, data were submitted to a 2(Polarity: *positive, negative*) x 3(Wh_type: *how-much, which, why*) repeated Measure ANOVA. Results are reported separately below:

a. *Position 3. Wh-element.*
The analysis revealed here only a main effect of Wh-type (F1(2, 40)=57.552, p<.0001). Post-hoc comparisons between each level of the Wh-type factor show significant differences between all the level of the variable Wh-type (p<.05).

This result could be readily explained by the variation in the number of characters between the various *wh-* elements (*how much* X > *which* X > *why* X).

b. *Position 4. Embedded subject.*
A main effect of Wh-type (F1(2, 40)=19.766, p<.0001) reached statistical significance. Planned comparisons between each level of the Wh-type factor show significant difference between *why* and *how much* N (p<.001) and *why* and *which* N (P<.001).

This result is consistent with the assumption that filler-gap dependencies impose a memory cost on sentence processing. In fact, *why* has a much faster reading time, when compared with *how much* and *which*. This difference is unlikely to be related to the lenght-effect found in position 3. In fact, now there is no difference left between *how much* and *which* and subjects process the two at the same speed. No other length effect would affect the reaction times, here, since the word length of the sentential subjects was counterbalanced across all the conditions.

c. *Position 5. Auxiliary.*
At the position where the distinction between positive and negative sentences is introduced, the ANOVA revealed a main effect of Polarity (F1(1,20)=26.001, p<.0001) and Wh-type (F1(2, 40)=7.019, p<.005. The interaction between Polarity x Wh-type (F1(2,40)=4.047, p<.05) was also significant. Post-hoc comparison between each level of the Wh-type factor showed only a significant increase in reaction times with *how much*, when compared to *which* (p<.05) and *why* (p<.005).

In positive sentences, no difference is observable between the three different kinds of *wh-* elements (figure 3), while the introduction of negation has two interesting effects. The first is a general increase in reaction times for each kind of wh-element. The second, is the significant interaction between polarity and wh-type. As the post-hoc comparison shows, this effect it due to an increase in reading time in the case of *how much*, the non-indexed wh- element.

d. *Position 6. Lexical verb.*
Main effects of Polarity (F1(1,20)=15.374, p<.001) and Wh-type (F1(2, 40)=12.332, p<.0001) where found also in this position, as well a significant interaction between Polarity and Wh-type (F1(2,40)=6.860, p<.005). Interestingly, post hoc analysis revealed no significant difference between *which* and *how much* (p>0.5), while they both differ from *why* (p<.005).

This result could be interpreted if we assume that a slow-down is associated with the filler integration. In this case, higher reaction times are expected for both *which* and *how much*, but not for *why*. Interestingly, an increase in reaction times for the two argument *wh-* is found in the positive and in the negative conditions. This suggests that the presence of an intervener won't prevent the parser from placing a gap site after the lexical verb.

e. *Position 7. Prepositional phrase*
After the verb, differences associated with the Wh-type disappear and only a main effect of Polarity reaches significance (F1(1,20)=6.962, p<.05).

This result shows that, at this point, the trace has been integrated in the constituent structure and the filler-gap dependency has been solved. Only the generic cost associated with negation is still having an effect in sentence final position.

5. General Discussion

The experimental results confirm that reaction times in sentence processing are influenced by both the referential and the argumental properties of the filler. For what concerns the argument/adjunct distinction, we compared the reaction times after the lexical verb between adjunct and argument *wh-*, finding a significant slow-down at the gap site for argument *wh*- elements. This effect has been found in both positive and negative sentences and it is consistent with previous findings on filler-gap dependencies (Crain & Fodor 1985, Stowe 1986, Phillips 2006).

The referential properties of the *wh*- element also seem to play a role in resolving filler-gap dependencies. In particular, in negative sentences, reaction times increase as soon as the sentential negative marker is encountered. Interestingly, this effect is higher for *wh*- elements denoting quantities i.e. *how much*. This effect is predicted by both semantic theories, as the one in Szabolczi & Zwart (1993) or syntactic theories (Rizzi 1990). Therefore, reaction times do not permit us to discriminate between these two families of accounts. However, other real time measures, sensitive to syntactic or semantic violations (ERP, MEG) could be helpful and this could be a viable direction for future research.

One last remark concerns the increase of reaction times after the verb, found in the negative condition for *wh*- arguments. This result could be interpreted by saying that *which* and *how much* are integrated in the syntagmatic structure after the verb and that negation doesn't block the generation of gap positions after the verb. This is consistent with the fact that negative island structures could be saved when the appropriate semantic or discursive conditions are met (Fox and Hackl 2006, Abrusan & Spector 2011).

References.

Abrusán, M and B. Spector. 2011. An Interval-Based Semantics for Degree Questions: Negative Islands and their Obviation, *Journal of Semantics* 28(1): 107-147.

Cinque, G. 1990. Types of A'-dependencies. Cambridge: MIT Press.

Crain, S., and J. D. Fodor. 1985. How can grammars help parsers? *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives,* ed. by David Dowty, Lauri Karttunen, and Arnold M. Zwicky, 94–128. Cambridge: Cambridge University Press.

Fox, D., and M. Hackl. 2007. The Universal Density of Measurement. *Linguistics and Philosophy* 29, 537-586.

Just, M.A., P. A. Carpenter and J. D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 11.228–38.

Kluender, R. & Marta Kutas. 1993a. Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. Journal of Cognitive Neuroscience 5.196–214.

Kluender, R.& M. Kutas. 1993b. Subjacency as a processing phenomenon. Language and Cognitive Processes 8.573–633.

McKinnon, R. and L. Osterhout. 1996. Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognitive Processes* 11.495–523.

Neville H., J. Nicol, A. Barss, K. Forster and M. Garrett. 1991. Syntactically-based sentence processing classes: Evidence from event-relatedbrain potentials. *Journal of Cognitive Neuroscience* 3.151–65.

Obenauer, H. 1984. On the Identification of Empty Categories. *The Linguistic Review* 4:153-202.

Phillips, C. 2007. The real-time status of island phenomena. Language 82.795–823.

Pickering, M., S. Barton, & R. Shillcock. 1994. Unbounded dependencies, island constraints and processing complexity. Perspectives on Sentence Processing, ed. by Chuck Clifton Jr., Lyn Frazier, & Keith Rayner, 199–224. London: Lawrence Erlbaum.

Rizzi, Luigi. 2001. On the Position Int(errogative) in the Left Periphery of the Clause. In Guglielmo Cinque and Giampaolo Salvi (eds.), Current Studies in Italian Syntax: Essays Offered to Lorenzo Rizzi, 287-296.

Rizzi, L. 1990. Relativized Minimality: MIT Press Cambridge, Mass.

Ross, J. R., 1967. Constraints on Variables in Syntax . Cambridge, MA: MIT dissertation. Published in 1986 as Infinite Syntax! by Ablex, Norwood, N. J.

Stowe, L. A. 1986. Evidence for on-line gap-location. *Language and Cognitive Processes* 1.227–45.

Szabolcsi, A., and F. Zwarts. 1993. Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics* 1:235-284.