



**HAL**  
open science

# Does Structural Complexity Necessarily Imply Processing Difficulty?

Frédérique Gayraud, Bruno Martinie

► **To cite this version:**

Frédérique Gayraud, Bruno Martinie. Does Structural Complexity Necessarily Imply Processing Difficulty?. Journal of Psycholinguistic Research, 2007, 37 (1), pp.21 - 31. <10.1007/s10936-007-9057-8>. <hal-04087196>

**HAL Id: hal-04087196**

**<https://hal.science/hal-04087196v1>**

Submitted on 4 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Does Structural Complexity Necessarily Imply Processing Difficulty?	
Article Sub-Title		
Journal Name	Journal of Psycholinguistic Research	
Corresponding Author	Family Name	<b>Gayraud</b>
	Particle	
	Given Name	<b>Frédérique</b>
	Suffix	
	Division	Laboratoire Dynamique du langage, ISH
	Organization	CNRS & Universié Lyon2
	Address	14, Avenue Berthelot, 69363, Lyon, France
	Email	Frederique.Gayraud@univ-lyon2.fr
Author	Family Name	<b>Martinie</b>
	Particle	
	Given Name	<b>Bruno</b>
	Suffix	
	Division	Laboratoire Dynamique du langage, ISH
	Organization	CNRS & Universié Lyon2
	Address	14, Avenue Berthelot, 69363, Lyon, France
	Email	b.martinie@wanadoo.fr
Schedule	Received	
	Revised	
	Accepted	
Abstract	<p>Our goal is to establish a link between the time needed to plan a sentence containing an embedded clause and the structure of this sentence. Contrary to a traditional monolithic conception of subordination, three types of embeddings were considered, depending on their degree of syntactic integration: subcategorized, modifier and pseudo-embedded clauses. We hypothesized that in the case of subcategorization, fewer pauses should occur between the matrix and the subordinate clause since the latter is required by the lexical properties of verbs. By contrast, pseudo-embedded clauses are the less integrated. Hence, they should exhibit planning characteristics similar to the ones of simple sentences, the matrix clause and the subordinate clauses being planned in two steps. Twenty texts produced by French speaking adults were recorded. Pauses were characterized according to their duration and position. Globally, both predictions were confirmed. We conclude that supposedly complex sentences are not necessarily difficult to process.</p>	
Keywords (separated by '-')	Complexity - Pause - Processing - Subordination - Syntax	
Footnote Information		

Journal: JOPR  
Article: 10936-9057

## Author Query Form

Dear Author,

During the preparation of your manuscript for typesetting, some questions have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to [spr\\_corrections2@sps.co.in](mailto:spr_corrections2@sps.co.in)

### Disk use

In some instances we may be unable to process the electronic file of your article and/or artwork. In that case we have, for efficiency reasons, proceeded by using the hard copy of your manuscript. If this is the case the reasons are indicated below:

- Disk damaged       Incompatible file format       LaTeX file for non-LaTeX journal  
 Virus infected       Discrepancies between electronic file and (peer-reviewed, therefore definitive) hard copy  
 Other: .....

We have proceeded as follows:

- Manuscript scanned     Manuscript keyed in       Artwork scanned  
 Files only partly used (parts processed differently: .....

### Bibliography

If discrepancies were noted between the literature list and the text references, the following may apply:

- The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.  
 *Uncited references*: This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.

### Queries and/or remarks

Section/paragraph	Details required	Author's response
	Please check the citation of Tables 3, 4 and Graphs 1–4.	
	References Caplan et al. (1998), Ford and Holmes (1978), Holmes (1995), Koch (1995) are not given in the list. Please check.	

# Does Structural Complexity Necessarily Imply Processing Difficulty?

Frédérique Gayraud · Bruno Martinie

A  
U  
T  
H  
O  
R  
  
P  
R  
O  
O  
F

1 **Abstract** Our goal is to establish a link between the time needed to plan a sen-  
2 tence containing an embedded clause and the structure of this sentence. Contrary  
3 to a traditional monolithic conception of subordination, three types of embeddings  
4 were considered, depending on their degree of syntactic integration: subcategorized,  
5 modifier and pseudo-embedded clauses. We hypothesized that in the case of sub-  
6 categorization, fewer pauses should occur between the matrix and the subordinate  
7 clause since the latter is required by the lexical properties of verbs. By contrast,  
8 pseudo-embedded clauses are the less integrated. Hence, they should exhibit plan-  
9 ning characteristics similar to the ones of simple sentences, the matrix clause and the  
10 subordinate clauses being planned in two steps. Twenty texts produced by French  
11 speaking adults were recorded. Pauses were characterized according to their duration  
12 and position. Globally, both predictions were confirmed. We conclude that supposedly  
13 complex sentences are not necessarily difficult to process.

14 **Keywords** Complexity · Pause · Processing · Subordination · Syntax

## 15 **The Notion of Complexity versus Difficulty**

### 16 Complexity

17 In the literature, it has commonly been accepted that subordination equals syntactic  
18 complexity, and that correlatively, simple independent clauses have to be regarded  
19 as a sign of syntactic simplicity or poverty. This explains why authors like O'Donnell

---

F. Gayraud (✉) · B. Martinie  
Laboratoire Dynamique du langage, ISH, CNRS & Université Lyon2, 14, Avenue Berthelot,  
Lyon, 69363, France,  
e-mail: Frederique.Gayraud@univ-lyon2.fr

B. Martinie  
e-mail: b.martinie@wanadoo.fr

(1974), who found a greater number of subordinate clauses in oral than in written data, concluded that the former was syntactically simpler than the latter. And this also explains why Kroll (1977) and Chafe (1982) drew similar conclusions after having found more coordinate sentences in spoken language, while written texts should rely more on clause embedding. As for Halliday (1985) and Beaman (1984), they contested such a dichotomy, the crucial argument being that the equation “oral = syntactic simplicity” and its counterpart “written = syntactic complexity” relies on false comparisons, in which “*Differences (. . .) often turn out to result from differences in the formality and purpose or register of the discourse rather than true differences between spoken and written language*” (Beaman 1984, p. 79). However, even if Beaman concluded that “*the basic assumption of many linguists (. . .) i.e. that subordination implies complexity, is false*” (Beaman 1984, p. 80), she paradoxically validated—at least implicitly—this assumption, since all along her survey, clause embedding was regarded as a highly relevant characteristic for measuring syntactic complexity in general, even if she claimed that other criteria have to be taken into consideration. Hence, after having observed that clause embeddings are globally equally frequent in the two modalities for the same register and the same degree of formality, she concluded that “*Based on the assumption that subordination implies complexity, the results show that, contrary to many previous assumptions, spoken narrative is on the whole just as complex as, if not more complex in some respects, than written narrative*” (Beaman, 1984 p. 78).

To recapitulate, all the surveys above-mentioned are corpus-based studies, which typically analyzed spoken and written language or/and planned and unplanned discourses, and they all made the assumption that the more complex a discourse is, the more embedded clauses it should contain. Some of these surveys (Beaman, 1984; Halliday 1985; Poole and Field 1976. . .) claimed that oral is syntactically complex, and highlighted the need to refine the notion of syntactic complexity. But none of them has specified which syntactic features should be linked to syntactic complexity in the domain of clause embedding, yet such a point is necessary to maintain the assumption that there is something between clause embedding and syntactic complexity. And finally none of these studies either claimed that there is no relation between them.

In order to refine the study of subordination from a complexity point of view we propose to consider it as a continuum reflecting various degrees of syntactic integration (Creissels to appear; Koch 1995). Thus examples (1–3) show an increasing degree of integration.

- (1) Although it is raining, there are twenty people here already
- (2) Snails are happy when it rains
- (3) Paul wants you to come

Strictly speaking, the linguistic notion of ‘subordination’ involves the notion of syntactic dependency. Hence if B is subordinated to A, then it can be shown that B verifies properties according to which it is syntactically dependent on A. Syntactic tests can be used to evaluate this dependency. The negation test is fairly reliable. It is based on the following principle: if a subordinate clause is genuinely syntactically embedded in a main clause, then the negation of the main clause verb should operate on the subordinate verb, that is to say, the verb of the subordinate clause should be negated instead of the main verb clause. From that point of view, in (1), (2) and (3) above, only (2) and (3) contain clauses which can be said to be subordinate:

65 (1')\* It is not raining although there are twenty people here already (but although  
66 this is the dry season)<sup>1</sup>

67 (2') Snails are not happy when it snows (but when it rains)

68 (3') Paul does not want you to come (but to phone him)

69 In (1), the negation of the event TO RAIN cannot 'reach' the event TO BE HERE  
70 ALREADY—so it can be said that TO BE HERE ALREADY is not in the 'scope'  
71 of the negation —, just because syntactically, there is no hierarchical structure where  
72 *to rain* would dominate *to be here already*, as shown by the impossibility to contrast  
73 with *but*-sentences. In (1), only semantic relations are implied between the 'main'  
74 clause and the 'subordinate' one, which amounts to saying that *although* must not be  
75 analyzed as a subordinating conjunction, but rather as a coordinating conjunction,  
76 since it does not set any hierarchy between the two clauses that it links. By contrast,  
77 in (2) and (3), as shown in (2') and (3'), the negation of the events BE HAPPY and  
78 WANT does not negate these events: in both cases, it remains true that snails are  
79 happy (but only when it rains) and that Paul wants something (not your coming, but  
80 your phone call).

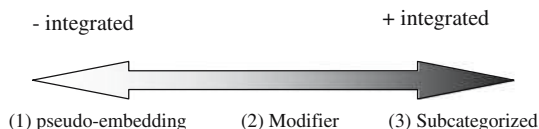
81 As a conclusion, *although*-clauses will be considered as 'pseudo-embeddings'. This  
82 category also includes *since*-clauses, (*why don't you buy it since you are so rich?*), *as*  
83 *for*-clauses (*As for Paul, I have met him once*) etc. The category of 'pseudo-embed-  
84 ding' excludes the adverbial clause (2) and the infinitive clause (3), which have been  
85 shown to be syntactically subordinate to the main clause.

86 Now, the difference between (2) and (3) is that only the infinitive clause (3) is  
87 included in the lexical properties of the main verb: in that case, the degree of syntactic  
88 integration must be considered as higher, and will be referred to as "subcategorized  
89 embedding". Thus, by definition, if V is a verb and S' a subordinate clause, S' is sub-  
90 categorized by V only if V is part of a subclass of verbs that can select subordinate  
91 clauses such as S'. In other words, any verb can select *when*-clauses (*to eat, to rain, to*  
92 *invite, to want...*) but not all verbs can select infinitive clauses (*\*I see you to come*). In  
93 the same vein, *whether*-clauses, *that*-clauses etc. are to be regarded as subcategorized  
94 embeddings: not all verbs can select a *whether*-clause (*to wonder can, to drive cannot*),  
95 and not all verbs can select *that*-clauses (*to say can, to invite cannot*) etc.: *he says*  
96 *that he is happy vs. \*he invites that people come early*. By contrast, cases illustrated  
97 by (2), where the subordinate clause does not depend on a specific verb class, but  
98 where it is a property of the verb class in its totality, will be referred to as "modifiers".  
99 The Modifier category also includes *while*-clauses (*Paul read the paper while he was*  
100 *waiting*), *if*-clauses (*I will take an umbrella if it rains*) etc. Relative clauses, which are  
101 generally embedded under a noun, are also modifiers, since all nouns can be modified  
102 by a relative clause.

103 We can rank the three types of clause on a continuum ranging from less to more  
104 integrated as illustrated in Fig. 1.

105 If globally speaking a complex entity is defined as composed of different inter-  
106 related elements, then, sentences (1), (2) and (3) are equally complex since they  
107 are composed of at least two sentential elements, in conformity with a traditional  
108 grammatical conception of subordination. Paradoxically, if the degree of syntactic  
109 integration is taken into consideration, (3) should be less complex than the other  
110 ones since high integration results in the creation of a cohesive planning unit. If this

1 \* indicates ungrammatical sentences.



**Fig. 1** Scale of syntactic integration

111 reasoning is correct, then it should be corroborated by behavioural measures such as  
 112 time of processing.

113 **Difficulty: Pauses and Planning**

114 The pioneer study by **Maclay and Osgood (1959)** hypothesized a nonrandom distri-  
 115 bution of pauses. They showed that pauses precede more often lexical words than  
 116 grammatical words, but if silent and filled pauses are distinguished, filled pauses occur  
 117 more often before grammatical words. Furthermore, they are found more often at  
 118 phrase boundaries while silent pauses tend to appear inside phrases. Another finding  
 119 is that phrase internal pauses do not reflect grammatical but lexical planning.

120 The well-known survey of **Goldman-Eisler (1972)** aiming at assessing the psycho-  
 121 logical validity of syntactic units established the link between syntactic properties and  
 122 temporal organization of speech production. It was shown that the higher the degree  
 123 of integration between clauses, the fewer pauses are encountered at the junction.  
 124 Hence the following hierarchy from the shorter to the longer between-clause transi-  
 125 tion: relative clauses < other subordinate clauses < coordinate clauses < sentences.  
 126 Similarly, **Hawkins (1971)** claimed that the clause is an important encoding unit of  
 127 speech on the basis of the finding that 2/3 of pauses were localized at clause boundaries  
 128 and accounted for 3/4 of total pause time.

129 **Hypotheses**

130 **H1.** In the case of subcategorization, fewer pauses should occur between the matrix  
 131 and the subordinate clause since the latter is required by the lexical properties of  
 132 verbs. Hence we can predict that such subcategorized clauses are planned early, that  
 133 is to say with the matrix, as a single entity. By contrast, pseudo-embedded clauses  
 134 are predicted to display more and longer pauses around the grammatical delimiter.  
 135 Sentential modifiers should exhibit intermediate planning characteristics since they  
 136 are intermediate as regard to their degree of syntactic integration. In other words if  
 137 we admit that chronometric data reflect complexity we should observe the following  
 138 scale: subcategorized < modifier < pseudo-embedding.

139 **H2.** If H1 stands, then, pseudo-embedding clauses should exhibit planning charac-  
 140 teristics similar to the ones of simple sentences, while subcategorized clauses should  
 141 be clearly different. We predict an intermediate state for modifiers.

142 **Method**

143 **Participants**

144 Twenty French speaking students (19 females, 1 male, age range: 17–21-years-old)  
 145 participated in the study. All were students in linguistics at Lyon2 University and  
 146 were given extra course credit for their participation.

---

147 Procedure

148 Each subject was asked (1) to expose orally his/her knowledge about violence, and  
149 (2) to relate a personal narrative in which s/he experienced violence or interpersonal  
150 conflict (However, this study did not take into consideration this text type variable).  
151 Each subject was recorded individually. The subjects were free to produce texts as  
152 long as they wanted and the mean length of texts was 35 clauses long ( $SD = 16.9$ ), the  
153 mean duration was 135 seconds ( $SD = 39$ ).

154 Apparatus

155 The signal was digitalized and the software of segmentation and pause detection  
156 STK-DDL Sound Toolkit<sup>2</sup> was used. Any silence exceeding 250 ms was detected  
157 as a pause. The recordings were transcribed using the software Transcriber®, a tool  
158 for segmenting, labeling and transcribing speech. During this operation, filled pauses  
159 were identified and measured. Each pause was coded according to its position.

160 Coding

161 Three pause locations were identified (main clause initial [xS]: see (4), before: see  
162 (5) or after: see (6) the grammatical delimiter [xQx]). In (4), the pause—indicated  
163 in bold—occurs at the beginning of the sentence, in (5), it occurs just before *que*  
164 ‘that’ and in (6), the pause immediately follows *que* ‘that’. Note that in this study, no  
165 difference is made between filled and silent pauses.

166 (4) <**FilledPause = 1.322s + SilentPause = 0.232s + FilledPause = 0.518s**> je pense que  
167 dans la vie de tous les jours nous avons nécessairement des rapports avec autrui  
168 (‘I think that in everyday life, we necessarily have relationships with others’)

169 (5) **nous avons du mal à accepter la différence** <**FilledPause = 0.370s + SilentPause**  
170 **= 0.237s**> que peut représenter l’autre (‘it is difficult to accept the difference that  
171 the other can represent’)

172 (6) mais qu’en même temps il dit que <**FilledPause = 0.231s**> ça m’apportera des  
173 libertés (‘but that at the same time, he claims that it will provide me with some  
174 freedom’)

175 Furthermore the degree of syntactic embedding was characterized on the basis of  
176 the syntactic criteria exposed above, which allowed to reach easily agreement between  
177 the two coders. We thus classified the data into three categories: Subcategorization  
178 (1), Modification (2), and Pseudo-embedding (3) above.

179 **Results**

180 Variety

181 A first remark concerns the variety of subordinate clauses encountered in our corpus.

---

<sup>2</sup> Laboratoire:ILaboratoire Dynamique du Langage, Université Lumière-Lyon2 & CNRS- UMR 5596.

**Table 1** Percentage (and number) of different types of subordinate clauses

Circumstantial	Complement clause	Infinitive	Relative
29% (63)	33% (71)	20% (43)	18% (39)

**Table 2** Percentage (and number) of different degrees of syntactic integration

Subcategorized	Modifier	Pseudo-embedding
64% (138)	24.5% (53)	11.5% (25)

**Table 3** Mean and standard deviation for duration and number of pauses according to the degree of syntactic integration

Subordination degree	Duration xS		Number xS		Duration xQx		Nbr xQx	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Subcategorized	.634	.911	.525	.641	.252	.531	.216	.623
Modifier	.480	.619	.472	.696	.326	.498	.302	.575
Pseudo-embedding	.661	.814	.640	.757	.775	.757	.920	.954

182 Regarding the distribution of sentence types, we found only 17% of ‘simple sen-  
183 tences’ versus 83% of ‘complex’ sentences among which we found the distribution  
184 presented in Table 1.

185 As can be seen from Table 1, the main types of subordinate clauses were signifi-  
186 cantly represented in the corpus. Table 2 shows how these categories were distributed  
187 according to the criteria of syntactic integration.

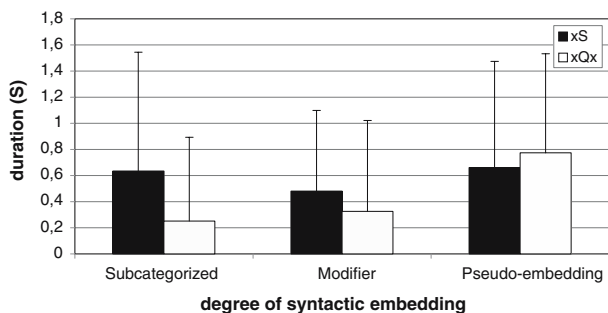
188 Similar to what was observed above, all the degrees of syntactic integration were  
189 represented with a dominance of subcategorization. We can even underline that the  
190 proportion decreased as a function of the degree of syntactic integration (Table 3).

191 The next section considers chronometric values in relation to these criteria.

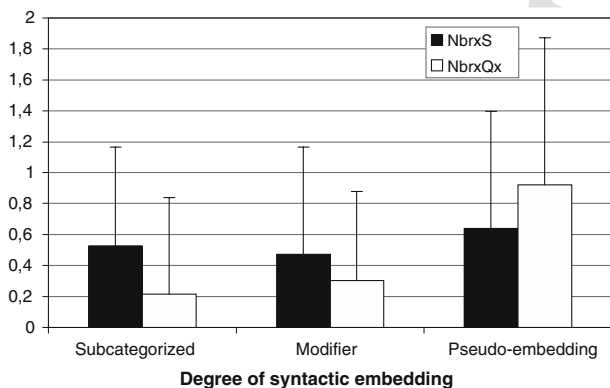
## 192 Cost of processing

193 Our first prediction was that the more subcategorized a clause is, the fewer and shorter  
194 should the pauses be at the grammatical delimiter location (Graphs 1 and 2).

195 The multivariate analysis of variance (MANOVA) for the three sentence types  
196 (subcategorized, modifier, pseudo-embedding) and two locations (sentence initial,  
197 around the grammatical delimiter) yielded a main effect of degree of syntactic embed-  
198 ding ( $F(2,154) = 10.120, p < .0001$ ), of pause location ( $F(2,154) = 62.108, p < .0001$ ) as  
199 well as an interaction of degree of syntactic embedding by pause location ( $F(4,154) =$   
200  $3.524, p = .0088$ ) on pause duration, such as pseudo-embeddings showed more pauses  
201 around the delimiter. *Post-hoc* testing (Fischer PLSD) at the 0.05 level revealed that  
202 pseudo-embedding showed longer pauses than modifier and subcategorization. For  
203 the number of pauses, we found a main effect of the degree of syntactic embed-  
204 ding ( $F(2,154) = 13.165, p < .0001$ ), an effect of pause location ( $F(2,154) = 194.196,$   
205  $p < .0001$ ) as well as an interaction of degree of syntactic embedding by pause loca-  
206 tion ( $F(4,15) = 7.113, p < .0001$ ), indicating more pauses around the delimiter for



**Graph 1** Pause duration according to the degree of syntactic embedding and location



**Graph 2** Number of pauses according to the degree of syntactic embedding and location

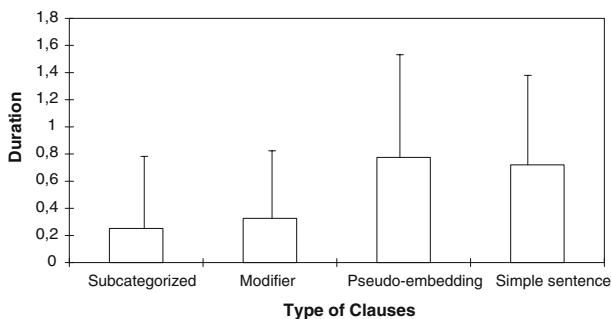
**Table 4** Mean and standard deviation for duration and number of pauses at initial location

Subordination degree	Duration		Number	
	Mean	SD	Mean	SD
Subcategorized	.252	.531	.216	.623
Modifier	.326	.498	.302	.575
Pseudo-embedding	.775	.757	.920	.954
Simple sentence	.720	.660	1	.854

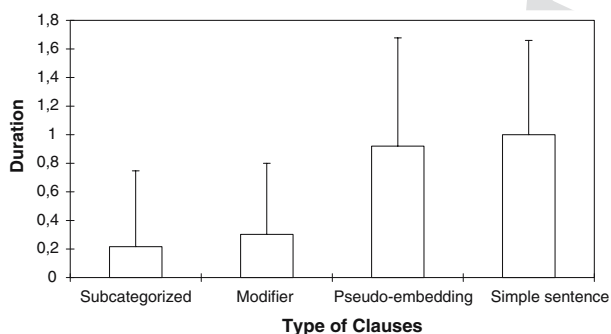
207 pseudo-embedding. *Post-hoc* comparisons revealed that pseudo-embeddings showed  
 208 more pauses than subcategorization and modifier.

209 Those results required the comparison of subordinate clause initial location and  
 210 simple sentence initial location (Table 4, Graphs 3 and 4). As a matter of fact, we  
 211 wanted to know whether pseudo-embeddings were processed the same way simple  
 212 sentences were.

213 We observed an overall effect of the type of sentence on the number and duration  
 214 of pauses ( $F(3,277) = 13.217, p < .0001$  for duration and  $F(3,277) = 22.338, p < .0001$   
 215 for the number). *Post-hoc* comparisons showed no significant difference between  
 216 subcategorized and modifiers, nor between pseudo-embedding and simple sentences  
 217 for both duration and number of pauses. However, subcategorized displayed shorter



**Graph 3** Duration of pauses at initial location



**Graph 4** Number of pauses at initial location

218 pauses than pseudo-embedding ( $p < .0001$ ) and than simple sentences ( $p < .0001$ ).  
 219 Similarly, they showed fewer pauses than pseudo-embeddings ( $p < .0001$ ) and than  
 220 simple sentences ( $p < .0001$ ). Furthermore, modifiers displayed shorter pauses than  
 221 pseudo-embedding ( $p = .0016$ ) and than simple sentences ( $p = .0003$ ) and fewer  
 222 pauses than pseudo-embeddings ( $p = .0004$ ) and than simple sentences ( $p = .0001$ ).  
 223 In other words, pseudo-embedded clauses and simple sentences behaved the same.

224 **Discussion**

225 The analysis of distribution shows that all the types of clauses were represented  
 226 even though our corpus is a spontaneous spoken speech corpus. All types were rep-  
 227 resented apart from traditionally very difficult to process structures such as object  
 228 center-embedded relative clauses (7), genitive relative clauses (8) indicated in bold  
 229 characters.

230 (7) The senator **that the reporter attacked** is reading the newspaper

231 (8) The senator **whose son is a lawyer** is reading the newspaper

232 It is well known that these sentences are very difficult to process (Blake et al. 1994;  
 233 Caplan et al. 1998; Fodor and Garrett 1967; Ford 1983; Frazier 1985; Gibson 1998;  
 234 Jurafsky and Martin 2000; Just and Carpenter 1992; Just et al. 1996; Kaan et al. 2000;  
 235 King and Just 1991; Stromswold et al. 1996). Considering the distribution of syntactic

236 integration, we observed that the less integrated a clause is, the less frequent it is and  
237 conversely.

238 As far as chronometric data are concerned, a first result is the partial validation  
239 of the first prediction: we indeed found more and longer pauses around the gram-  
240 matical delimiter as a function of integration. This result is globally compatible with  
241 Holmes (1995) whose survey found out that combined clauses (nonrestrictive rela-  
242 tive, independent, adverbial) are more likely to be preceded by a silent pause or a  
243 vocalized hesitation than embedded clauses (complement and restrictive relatives) in  
244 both French and English.

245 As far as sentence initial position is concerned, we found no difference: the absence  
246 of pauses between the main and the subordinate clause in the case of subcategori-  
247 zation was not compensated by sentence initial pauses as was predicted. We can  
248 hypothesize that in the case of subcategorized clauses, planning activity takes place at  
249 other sites. An explanation is provided by a study by Ford and Holmes (1978) using  
250 a different method measuring reaction time (RT) to tone signals while producing  
251 speech. The results showed that RTs at the beginning of clauses were longer for the  
252 more integrated clauses like relative and complement clauses while non embedded  
253 clauses showed significantly shorter RTs. These results suggest that in the case of  
254 highly integrated clauses, planning takes place during their beginning and not before,  
255 which is compatible with our study. This also shows limitations of surveys exclusively  
256 based on pause time since planning also occurs while speaking.

257 The second result confirms globally H2 : pseudo-embedding indeed resembles sim-  
258 ple sentences, which reflects the lack of syntactic integration in both cases. We had  
259 predicted an intermediate status for modifiers which is not confirmed : in reality, they  
260 do not differ from subcategorized clauses with which they share the characteristic of  
261 being syntactically integrated. Thus chronometric data reflect integration in a binary  
262 way (integrated *versus* nonintegrated).

263 If the notion of complex sentence was intrinsically related to the difficulty of pro-  
264 cessing, things would have been different. Indeed, if sentential complexity equaled  
265 difficulty of processing, fewer subordinations would be encountered; some would be  
266 avoided. Very complex sentences such as (9) , where a subordinate clause contained  
267 another subordinate clause which contained itself another subordinate clause etc.,  
268 were rare but not completely avoided.

269 (9) I think that we were a bit upset to realize that it was the young pupils and that  
270 we were being attacked by younger than us who were laughing at us when a friend of  
271 mine slipped after receiving a snowball in her face.

272 Supposedly complex sentences are not necessarily complex to process. Subcate-  
273 gorized clauses give rise to fewer and shorter pauses. According to our criteria, it  
274 means that they are less complex to process. Our interpretation is that on the linguis-  
275 tic ground, it is a consequence of valence which specifies both syntactic and semantic  
276 selections. The cognitive implication is that this choice reduction lightens planning  
277 work.

278 Our study showed that complexity can be tackled under several perspectives:

- 279 – a complex sentence is not necessarily difficult to process. Planning difficulty is influ-  
280 enced by different factors such as memory cost, the number of choices to make,  
281 words or syntactic structures frequency . . .
- 282 – sentential embedding does not constitute a homogeneous category with respect to  
283 the notion of linguistic complexity and subsequent planning cost.

284 From a syntactic point of view, an embedded clause deserves this label if it is possible  
285 to identify a mechanism of syntactic integration. If we relate difficulty of processing  
286 to complex sentence, then, it is predicted that the more integrated a clause is, the  
287 more difficult it should be to process. However, it is precisely the opposite that we  
288 found. Two possible conclusions can be drawn from this paradoxical finding: (i) the  
289 planning cost operates at the sentential and not the clausal level, which is reflected by  
290 the concentration of pauses in sentence initial position, (ii) pauses are not indices of  
291 processing difficulty.

## 292 References

- 293 Beaman, K. (1984). Coordination and subordination revisited: syntactic complexity in spoken and  
294 written narrative discourse. In D. Tannen (Ed.), *Coherence in spoken and written discourse*  
295 (pp. 45–80). Ablex: Norwood, NJ.
- 296 Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between mem-  
297 ory span and measures of imitative and spontaneous language complexity in preschool children.  
298 *International Journal of Behaviour Development, 17*(1), 91–107.
- 299 Chafe, W. (1982). Integration and involvement in speaking, writing and oral literature. In D. Tannen  
300 (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35–53). Ablex: Norwood,  
301 NJ.
- 302 Creissels, D. (2006). *Syntaxe générale : une introduction typologique*, Paris: Hermès (to appear).
- 303 Fodor, J.-A. & Garrett, M. (1967). Some syntactic determinants of sentential complexity. *Perception*  
304 *and Psychophysics, 2*, 289–296.
- 305 Ford, M. (1983). A method of obtaining measures of local parsing complexity throughout sentences.  
306 *Journal of Verbal Learning and Verbal Behaviour, 22*, 203–218.
- 307 Frazier, L. (1985). Syntactic complexity. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural*  
308 *language parsing* (pp. 129–189). Cambridge: Cambridge University Press.
- 309 Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.
- 310 Gibson, E., & Thomas, J. (1996). The processing complexity of English center-embedded and  
311 self-embedded structures. In C. Schutze (Ed.), *Proceedings of the NELS 26 workshop on language*  
312 *processing: MIT working papers in linguistics* (pp. 45–71). Cambridge: MA.
- 313 Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15*, 103–113.
- 314 Halliday, M. A. K. (1985). *Spoken and written language*. Oxford: Oxford University Press.
- 315 Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14*, 277–288.
- 316 Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing. An introduction to natural*  
317 *language processing, computational linguistics, and speech recognition*. New Jersey: Prentice Hall.
- 318 Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in  
319 working memory. *Psychological Review, 99*(1), 122–149.
- 320 Just, M.-A., Carpenter, P.-A., Keller, T.-A., Eddy, W.-F., & Thulborn, K.-R. (1996). Brain activation  
321 modulated by sentence comprehension. *Science, 274*, 114–116.
- 322 Kaan, E., Harris, A., Gibson, E., & Holcomb, P.-J. (2000). The P600 as an index of syntactic integration  
323 difficulty. *Language and Cognitive Processes, 15*(2), 159–201.
- 324 King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working  
325 memory. *Journal of Memory and Language, 2*, 580–602.
- 326 Kroll, B. (1977). Combining ideas in written and spoken English: A look at subordination and coord-  
327 ination. In E. Keenan & T. Bennett (Eds.), *Discourse across time and space* (pp. 69–108). Los  
328 Angeles, University of Southern California: Southern California Occasional Papers in Linguistics.
- 329 Maclay, H., & Osgood C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*,  
330 19–44.
- 331 O'Donnell, R. C. (1974). Syntactic difference between speech and writing. *American Speech, 49*, 102–  
332 110.
- 333 Poole, M. E., & Field, T. W. (1976). A comparison of oral and written code elaboration. *Language and*  
334 *Speech, 19*, 305–311.
- 335 Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension  
336 by positron emission tomography. *Brain and Language, 52*(3), 452–73.