

The impact of shared knowledge on speakers' prosody

Amandine Michelas*, Cécile Cau and Maud Champagne-Lavau*

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

Running head: Shared knowledge and speaker's prosody

Keywords: Prosody; Speech production; Shared knowledge; Perspective taking; Audience design; French

* Corresponding authors:

Amandine Michelas & Maud Champagne-Lavau

Laboratoire Parole et Langage

Aix-Marseille Université, CNRS

5 avenue Pasteur

13604 Aix-en-Provence

France

Email: michelas@lpl-aix.fr, maud.champagne-lavau@univ-amu.fr

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

Abstract:

How does the knowledge shared by interlocutors during interaction modify the way speakers speak? Specifically, how does prosody change when speakers know that their addressees do not share the same knowledge as them? We studied these effects in an interactive paradigm in which French speakers gave instructions to addressees about where to place a cross between different objects (e.g., You put the cross between the red mouse and the red house). We manipulated (i) whether the two interlocutors shared or did not necessarily share the same objects and (ii) the informational status of referents. We were interested in two types of prosodic variations: global prosodic variations that affect entire utterances (i.e., pitch range and speech rate variations) and more local prosodic variations that encode informational status of referents (i.e., prosodic phrasing for French). We found that participants spoke more slowly and with larger pitch excursions in the not-shared knowledge condition than in the shared knowledge condition while they did not prosodically encode the informational status of referents regardless of the knowledge condition. Results demonstrated that speakers kept track of what the addressee knew, and that they adapted their global prosody to their interlocutors. This made the task too cognitively demanding to allow the prosodic encoding of the informational status of referents. Our findings are in line with the idea that complex reasoning usually implicated in constructing a model of the addressee co-exists with speaker-internal constraints such as cognitive load to affect speaker's prosody during interaction.

51

52

1. Introduction

53

54

55 The way speakers share or do not share knowledge with their interlocutors during interaction
56 has consequences for words they produce but also for the prosody associated with the words.

57 The term *prosody* refers to variations of suprasegmental features of speech including pitch,

58 duration and intensity. When these three acoustic parameters vary, it induces different kinds of

59 prosodic variations. First, it creates local prosodic variations that affect words or small groups

60 of words through prosodic prominence attribution (i.e., through the degree to which these

61 words perceptually stand out from surrounding words). Second, it provides more global

62 prosodic variation that affects entire utterances such as speech rate and pitch range variations.

63 The present article deals with the possible impact of shared knowledge manipulation between

64 interlocutors on one or both of these two kinds of prosodic variations in French.

65

66 The first evidence of the impact of shared knowledge on speakers' prosody comes from

67 the fact that speakers encode the already-mentioned vs. contrastive status of words through

68 prosodic prominence by consulting information that is mutually shared with their interlocutors

69 (Halliday, 1967). For instance, in most Germanic languages such as English or Dutch, speakers

70 select deaccenting forms for already-mentioned words and prosodic prominent forms for

71 contrastive words (Chafe, 1976; Clark & Haviland 1977; Prince, 1986; Cruttenden, 1997;

72 Halliday, 1967). In contrast to Germanic languages, it appears that the correspondence between

73 deaccenting vs. prosodic prominence and information status is less systematic in Romance

74 languages. For instance, it appears that Catalan and Spanish make less use of prosodic

75 prominence to mark information status but also rely on syntactic strategies affecting word order

76 (see Swerts, Krahmern & Avesani, 2002, for a discussion of the distinction between the
77 prosodic marking of information status in Germanic and in Romance languages). Similar to
78 what happens in Catalan and Spanish, at least two strategies are allowed in French to encode
79 the new vs. given status of words: syntactic strategies such as dislocations (see Fery, 2001) or
80 just prosodic marking. In French, due to the absence of lexical stress, the prosodic encoding of
81 focus is different from that one of many other Romance languages, and mainly relies on
82 prosodic phrasing (i.e., the grouping of words into prosodic units of different sizes).
83 Specifically, a number of studies have reported a tendency for contrastive words to be parsed
84 in a separate prosodic unit from the rest of the utterance (Féry, 2001; Dohen & Loevenbruck,
85 2004; Beyssade et al., 2009; Chen & Destruel, 2010; Michelas et al., 2014). For instance, using
86 a paradigm in which participants had to correct sentences produced by a fictitious interlocutor
87 (a prompt), Dohen & Loevenbruck (2004) showed that participants produced utterances with
88 the verb phrase and the object noun phrase grouped together in the same prosodic constituent
89 (the Accentual Phrase or AP) in contexts in which no element of the utterance contrasted with
90 a previously-mentioned discourse element (e.g., [*Les loups*]_{AP} [*suivaient Marie-Lou*]_{AP} ‘The
91 wolves were following Marie-Lou’). However, they produced the element that contrasted with
92 a previously mispronounced element in a separate AP when they knew that the prompt
93 conveyed false-belief about this element (e.g., [*Les loups*]_{AP} [*suivaient*]_{AP} [*Marie-Lou*]_{AP} when
94 ‘Marie-Lou’ was mispronounced). More recently, in an interactive paradigm in which a
95 director had to indicate noun-adjective pairs of items to an addressee, Michelas et al. (2014)
96 showed that French participants parsed the target noun in the same AP as the following
97 adjective when it was identical to the noun in the preceding fragment (e.g., *bonbons marron*
98 ‘brown candies’ followed by [*bonbons violets*]_{AP} ‘purple candies’) while they parsed it in a
99 separate AP from the following adjective when it contrasted with it to warn their interlocutor
100 that this noun constituted a contrastive entity (e.g., *bougies violettes* ‘purple candles’ followed

101 by [*BONBONS*]_{AP}[*violets*]_{AP} ‘purple candies’). It is not clear whether this phenomenon resulted
102 from a speakers’ willingness to help their addressee (as predicted by ‘the audience design
103 hypothesis’; see Smiljanic & Bradlow, 2009) or whether it is mediated by production-internal
104 constraints on the speakers’ part that favor speech that is easy to produce (Arnold et al., 2012;
105 Bard et al., 2000; Bard & Aylett, 2004; Kaland et al., 2014; Fraundorf et al., 2015).
106 Nevertheless, it appears that when speakers share a set of alternatives with their partner, they
107 prosodically encode the new vs. given status of word.

108
109 Moreover, it is possible that the way in which interlocutors shared knowledge during
110 the interaction affects not only the prominent words but also all other regions of speakers’
111 utterance. This idea is related to the fact that when speakers are aware of a speech perception
112 difficulty on the part of addressees, they modify their global prosody (see Smiljanic & Bradlow,
113 2009, for a review). For instance, a lot of studies focusing on ‘clear speech’ (a distinct listener-
114 oriented speaking style that is more intelligible than conversational speech) converge on the
115 idea that speakers decrease their speech rate and increase their wider dynamic pitch range in
116 order to accommodate addressees in contexts in which their listener experience background
117 noise, hearing impairments or a different native language, (Picheny et al., 1986, 1989; Matthies
118 et al., 2001; Perkell et al., 2002; Ferguson & Kewley-Port, 2002; Bradlow et al., 2003; Krause
119 & Braida, 2004; Liu et al., 2004; Smiljanic & Bradlow, 2005). Following this idea, it is
120 reasonable to hypothesize that when speakers are aware they do not share the same knowledge
121 as their interlocutors, they produce clear speech thus modifying their global prosody to prevent
122 their listeners from comprehension difficulties.

123
124 In the present study, we collected speech samples in a controlled but interactive
125 paradigm that enabled us to measure prosodic variations on speech fragments that refer either

126 to objects shared by the two interlocutors (*shared knowledge condition*) or that refer to objects
127 that might differ between the two interlocutors (*not-shared knowledge condition*). In the
128 paradigm we built, participants played an interactive game. They were seated in a quiet room
129 with a confederate, facing each other, both with a computer screen. The participants had to
130 give instructions to the confederate about where to place a cross between the different objects
131 displayed on the screen (e.g., *Tu mets la croix entre la souris bordeau et la maison bordeau*;
132 ‘You put the cross between the red mouse and the red house’). The participants knew that when
133 objects appeared in white boxes, the confederate shared exactly the same objects as them on
134 her screen while this was not necessarily the case when objects appeared in black boxes. In this
135 case, participants knew that objects displayed on black boxes could differ either in their type
136 (e.g., mouse vs. car) or in their color (e.g., red vs. purple) on the confederate’s screen. We had
137 two goals. Our first goal was to test whether the fact that the speakers knew their addressee
138 shared or did not necessarily share the same set of objects/colors as them, affected the global
139 prosodic variations they produce. Following the idea that speakers adapt their global prosody
140 to the difficulties encountered by their addressee during spoken interaction, we hypothesized
141 that, in the absence of shared knowledge between participants, speakers would modify their
142 global prosodic features such as speech rate and pitch range to facilitate addressees’
143 comprehension as it is the case in clear speech. Specifically, we predicted that speakers would
144 speak more slowly and with larger pitch incursions in a condition in which they knew their
145 interlocutor did not necessarily share the same knowledge (*not-shared knowledge condition*)
146 compared to a condition in which they knew s/he shared it (*shared-knowledge condition*). Our
147 second goal was to test whether shared knowledge between interlocutors affected prosodic
148 variations that encoded informational status of referents. To do so, we manipulated the
149 contrastive status of target fragments that always appeared in second position within pairs
150 produced by the speaker (e.g., *maison bordeau* ‘red house’ in *Tu mets la croix entre la souris*

151 *bordeau et la maison bordeau*; ‘You put the cross between the red mouse and the red house’).
152 Within these target fragments, (1) the target noun was identical to the noun of the 1st fragment
153 while the target adjective contrasted to the one in the 1st fragment (e.g., *maison violette* vs.
154 *maison BORDEAU*) **contrastive adjective**, (2) the target noun in the 2nd fragment contrasted to
155 the noun in the 1st fragment while the adjective was identical (e.g., *souris bordeau* vs. *MAISON*
156 *bordeau*; **contrastive noun**) (3) both the target noun and the target adjective contrasted to their
157 counterparts in the 1st fragment (e.g., *souris violette* vs. *MAISON BORDEAU*; **contrastive**
158 **fragment**). Following Michelas et al. (2014) and others, we hypothesized that the presence of
159 shared knowledge between participants would affect the prosodic encoding of informational
160 status of referents and thus the prosodic phrasing produced by French participants. Specifically,
161 we predicted that speakers would isolate contrastive nouns in separate APs when they knew
162 their interlocutor shared exactly the same set of alternatives/objects as them (**shared knowledge**
163 **condition**) but not when they knew their interlocutor did not necessarily share it (**not-shared**
164 **knowledge condition**). Thus, we expected participants to produce more 2 AP phrasing in the
165 contrastive noun condition (*souris bordeau* vs. *MAISON bordeau*, ‘red mouse vs. red house’)
166 compared to the contrastive adjective condition (*maison violette* vs. *maison BORDEAU*
167 ‘purple house vs. red house’) in shared knowledge circumstances but not in not-shared
168 knowledge circumstances. In this last case, it would not be relevant for speakers to highlight
169 one particular alternative to their addressee since they knew their interlocutor did not
170 necessarily share the same set of alternatives as them. Finally, we tested an additional condition
171 in which both the target noun and the following adjective were contrastive (*souris violette* vs.
172 *MAISON BORDEAU* ‘purple mouse vs. red house’) to see whether our chance to obtain a 2 AP
173 phrasing is increased when the entire fragment (and not only the noun) is contrastive. This
174 hypothesis is based on Michelas et al. 2014’s study who obtained more 2 APs phrasing (72%
175 of cases) than 1 AP phrasing (28% of cases) on filler items showing this kind of “double

176 contrast”. If this double contrast actually favors the parsing of the noun in a separate AP from
177 the following adjective as it was the case in Michelas et al.’s study, speakers would produce
178 more 2 APs phrasing in our contrastive fragment condition compared to our contrastive noun
179 condition in shared knowledge circumstances.

180

181

182 **2. Method**

183

184

185 2.1. Participants

186

187 Twenty-two participants (16 women) between 18 and 47 years old (mean age = 21.4,
188 SD = 5.9; mean education: 13.8 years, SD = 0.8) participated in the experiment. All participants
189 gave written informed consent. All reported having no neurological, psychiatric, visual or
190 hearing impairment. They were all native speakers of French, the language of the experiment.
191 The study was approved by the university's ethics review board (*Comité d'évaluation éthique*
192 *d'Aix-Marseille université*) and conducted according to the principles expressed in the
193 Declaration of Helsinki. The experiment was classified as purely behavioral, and the testing
194 involved no discomfort or distress to the participants.

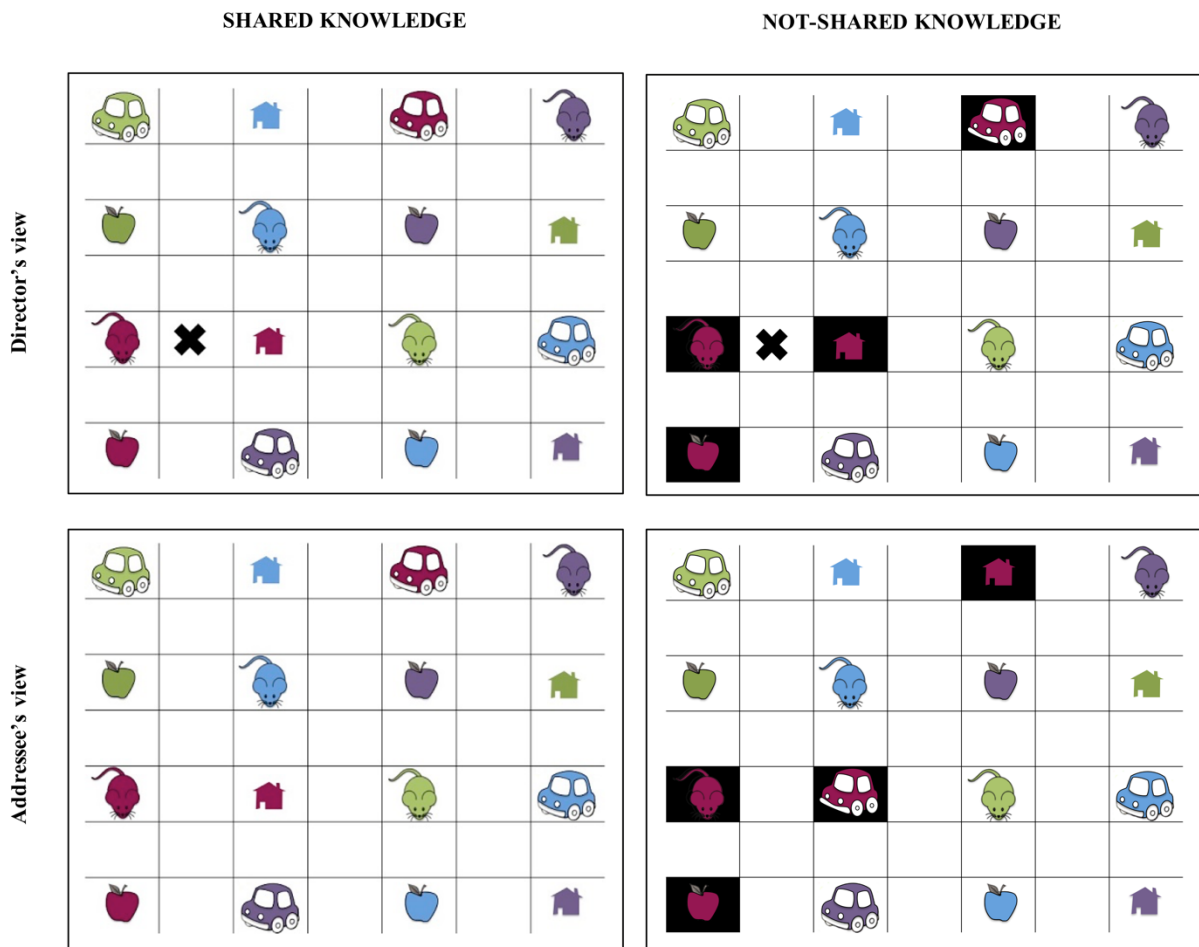
195

196 2.2. Materials

197

198 An interactive task involving cooperation between a director (the participant) and an
199 addressee (a confederate) was developed for the present study. During this task, the director

200 had to give instructions to the addressee about where to place a cross between different objects
201 on a grid by naming the type and the color of two critical objects (e.g., *Tu mets la croix entre*
202 *la souris bordeau et la maison bordeau*; ‘You put the cross between the red mouse and the red
203 house’). Each grid was composed of 16 objects corresponding to the combination of 4 types of
204 objects and 4 colors. For instance Figure 1 illustrates the director’s and addressee’s grid built
205 around the 4 following colors : *bordeau* ‘red’ (critical color), *violet* ‘purple’ (critical color),
206 *vert* ‘green’ (filler color), *bleu* ‘blue’ (filler color) and the 4 types of object: *souris* ‘mouse’
207 (critical object), *maison* ‘house’ (critical object), *pomme* ‘apple’ (filler object), *voiture* ‘car’
208 (filler object). A total of 56 different types of object were used to build the 84 grids of the
209 experiment. These grids were created from 14 target fragments (2 knowledge conditions x 3
210 informational status conditions x 14 target noun-adjective fragments). The objects and the
211 colors of the critical pair were chosen so that we controlled for the words produced by the
212 participants (e.g., the number of syllables, the syllable structure and the phonemic
213 characteristics of words).



215

216 Figure 1. Director and addressee's views for the *maison bordeau* 'red house' fragment in the
 217 contrastive noun shared knowledge condition (left panel) and not-shared knowledge
 218 (right panel). The director's view includes the cross (top panel) while the addressee's view
 219 does not (bottom panel). This figure is similar but not identical to the original image published in
 220 Michelas et al. (2017) and is therefore for illustrative purposes only.

221

222 In half of the grids, the cross has to be placed horizontally (as it is the case in Figure 1)
 223 whereas in the other half, it has to be placed vertically. The location of the cross was stable for
 224 the same target fragment (e.g. *maison bordeau* 'red house') that appeared six times in the six
 225 different conditions (see details of the 6 experimental conditions below) but was never the same
 226 across the 14 target fragments so that participants cannot anticipate the location of the cross. In

227 the immediate proximity of the two target items (i.e., below the two target items when the cross
228 has to be placed horizontally and on the left of the two target items when the cross has to be
229 placed vertically), a competitor object and a filler object were always visible. The competitor
230 object consisted of a different object of the same color for the *contrastive noun condition* (see
231 the list of the experimental conditions below and in Table 1), the same object but in a different
232 color for the *contrastive adjective condition* and a different object with a different color for the
233 *contrastive fragment condition*. The filler object was always a different object than the objects
234 used for the target pair and the competitor and had always a different color. The other objects
235 on the grid were randomly arranged. Each pair of target items gave rise to six different grids
236 (one for each experimental conditions). The arrangement of objects was the same for each
237 sextuplet of grids except for the white/black boxes manipulations as explained in the following
238 paragraph.

239
240 We developed two experimental manipulations. First, to control for the presence vs.
241 absence of shared knowledge between participants, we manipulated the speaker's knowledge
242 about whether his/her addressee shared or did not necessarily share the same objects as him/her.
243 To do so, in the *not-shared knowledge condition*, target items appeared in black boxes and
244 participants knew that they could have different objects or different colors for objects in those
245 boxes (see Figure 1 bottom panel) while in the *shared knowledge condition*, items appeared in
246 white boxes meaning for the participants that their interlocutor shared exactly the same objects
247 as him/her (see Figure 1 top panel). The black boxes were always four in number and their
248 distribution was conditioned by the type of contrast for the current pair of objects. For instance,
249 for the 'red mouse vs. red house' pair (contrastive noun), the four black boxes affected the four
250 red objects. In the same manner, for the 'purple house vs. red house' pair (contrastive
251 adjective), the four black boxes affected the four houses. For the contrastive fragment pair, the

252 four black boxes affected the four objects of the same color as the target item for half of the
253 target pairs (e.g., the four red objects for the ‘purple mouse vs. red house’ pair) or the four
254 objects of the same type as the target item for the other half of the target pairs (e.g., the four
255 mills for the ‘orange sun vs. brown mill’ pair). To obtain the addressee’s grid in the not-shared
256 knowledge condition, the second object of pairs of objects was always replaced by one of the
257 two other possible alternatives on the grid. For instance, in Figure 1, the red house in the
258 director’s grid was replaced by the red car in the addressee’s grid. This manipulation allowed
259 us to create a situation in which participants always shared the same set of objects for a given
260 grid but knew that, sometimes, critical objects or colors might differ at a given location on the
261 addressee’s screen and sometimes they might actually be the same even though they appeared
262 in black boxes. The location of black boxes was stable for the same target fragment that
263 appeared six times in the six different conditions but was different across the 14 target
264 fragments so that participants cannot anticipate their location. The complete instructions given
265 to participants are shown in Appendix I.

266
267 Second, in order to control for the informational status of referents, we manipulated
268 whether the noun or the adjective in the 2nd fragment was the same or contrasted with the noun
269 or the adjective in the 1st fragment. We obtained three types of informational status: (1) the
270 noun in the 2nd fragment was kept constant while the adjective in the 2nd fragment contrasted
271 to the adjective in the 1st fragment (e.g., *maison violette* vs. *maison BORDEAU*) **contrastive**
272 **adjective** (2) the noun in the 2nd fragment contrasted to the noun in the 1st fragment while the
273 adjective was kept constant (e.g., *souris bordeau* vs. *MAISON bordeau*; **contrastive noun**) (3)
274 both the noun in the 2nd fragment contrasted to the noun in the 1st fragment and the adjective
275 in the 2nd fragment contrasted to the adjective in the 1st fragment (e.g., *souris violette* vs.
276 *MAISON BORDEAU*; **contrastive fragment**).

277

278 The combination of the two experimental factors (type of knowledge and informational
 279 status of target fragment) led to 6 experimental conditions. The 6 noun-adjective pairs
 280 corresponding to the 6 experimental conditions for the *maison bordeau* target fragment were
 281 given as example in Table 1. The target fragment was always the second fragment of the pair
 282 (e.g., red house in the ‘red mouse vs. red house’ pair). Each participant saw the 84 screens (2
 283 knowledge conditions x 3 informational status conditions x 14 target noun-adjective fragments)
 284 that appeared simultaneously on his/her computer’s screen and on the computer’s screen of
 285 his/her addressee. The order of presentation of screen view was fixed across participants but
 286 was semi-randomized within participants so that the same target item did not appear more than
 287 twice in succession and the same experimental condition was not repeated more than twice in
 288 a row.

289

290 Table 1. Noun-adjective pairs, corresponding to the *maison bordeau* ‘red house’ target
 291 fragment, elicited by the task and thus produced by directors for each of the 6 experimental
 292 conditions.

293

	Shared knowledge	Not-shared knowledge
	<i>Directors knew addressees shared the same object/color</i>	<i>Directors knew addressees did not necessarily share the same object/color</i>
Contrastive adjective	maison violette/maison bordeau 'purple house/red house'	maison violette/maison bordeau 'purple house/red house'
Contrastive noun	souris bordeau/maison bordeau 'red mouse/red house'	souris bordeau/maison bordeau 'red mouse/red house'

Contrastive fragment

souris violette/maison bordeau

souris violette/maison bordeau

'purple mouse/red house'

'purple mouse/red house'

294

295

296 2.3. Procedure

297

298 The director and the addressee were seated, facing each other, both with a computer

299 screen. They could not see each other's computer screen. The director was asked to give

300 instructions to the addressee about where to place the cross by indicating to the addressee the

301 type and the color of objects (e.g., *Tu mets la croix entre la souris bordeau et la maison*302 *bordeau*; 'You put the cross between the red mouse and the red house'). The addressee was

303 always the same confederate (i.e., a native French female speaker). She was introduced to the

304 participant as the experimenter but was naïve about the scientific aims of the experiment.

305 However, since the addressee performed the task 22 times in the exact same order, she received

306 instructions about the feedbacks she was allowed to produce. These instructions were intended

307 to homogenize the feedbacks she produced from one participant to another. When she

308 interacted with the participant, she was explicitly asked to give him/her continual feedback of

309 two types. First, in cases in which she identified immediately and without effort where to put

310 the cross, she was asked to signal understanding with feedbacks such as *d'accord* 'okay', *ok*311 'okay', *c'est bon* 'I'm good'. Second, in cases in which she had difficulties to identify where

312 to place the cross (typically in case of black boxes), she was instructed to tell the participant

313 why she is in difficulties (e.g., *Alors je n'ai pas de souris bordeau et de maison bordeau à côté*

314 'Well, I don't have any red house next to a red mouse') and to ask them for clarifications

315 leading to several exchanges between the two interactional partners. Clarification questions

316 could be of three types: (i) vague clarification question (e.g., *Je n'ai pas bien compris* 'I don't317 understand very well', *Tu peux répéter?* 'Can you repeat?', *Je suis un perdue, est-ce que tu*

318 *peux me répéter ce que tu vois ?* ‘I’m lost, can you tell me what you see?’), (ii) precise
319 clarification question (e.g., *Où est-ce que je dois mettre la croix par rapport à la souris bordeaux*
320 *?* ‘Where should I put the cross in relation to the red mouse?’, *Dans quelle colonne se situe la*
321 *croix?* ‘In which column is the cross located?’) (iii) inference clarification question (e.g., *Est-*
322 *ce que je dois mettre la croix en dessous du bureau bordeaux?* ‘Do I have to put the cross below
323 the red desk ?, *Est-ce que la croix est à gauche du crayon marron ?* ‘Is the cross to the left of
324 the brown pencil ?’, *Est-ce que la croix est sur la même ligne que le crayon violet ?* ‘Is the
325 cross on the same line as the purple pencil?’). After identifying the location of the cross, the
326 confederate had to show the participant her understanding using feedback such as ‘okay’. We
327 recorded both the director’s and addressee’s productions, but we only analyzed the director’s
328 productions. The types of feedback did not give rise to any coding. Moreover, regardless of the
329 type of exchange that took place between the director and the addressee, the segment of speech
330 we analyzed was always extracted from the initial production of the speaker. This segment was
331 elicited before any confederate’s reaction. For instance, we did not analyze cases in which the
332 speaker repeated target fragments after a clarification question asked by the addressee. An
333 extract from the dialogue obtained between one participant and the confederate is given as
334 example in Appendix II. This extract corresponds to the screen views shown in Figure 1.

335

336 Each screen view appeared simultaneously on the director and the addressee computer’s
337 screen. Once the addressee has moved the cross with the computer mouse on his/her screen,
338 both participants were asked to simultaneously click on the space bar to make the next screen
339 view appeared. The director and the addressee were recorded in a quiet room using a Zoom
340 H4N Handy Recorder and a Headset Cardioid Condenser Microphone (AKG C520). We
341 analyzed directors’ speech only.

342

343 The experimental phase began with 6 practice screen views followed by 84 target screen
344 views. The 6 practice screen views were built from 8 new objects and 8 new colors that did not
345 appear in the 84 target screen views.

346
347 To ensure that directors could identify and use a consistent label for each target object
348 and color, before the experimental phase, participants performed a familiarization phase during
349 which they had to report aloud what they saw in 64 pictures. These 64 pictures were composed
350 of all the possible objects participants met in the experimental phase (56 different objects) in
351 one of the 4 possible colors used for the experimental phase plus the 8 objects of the 6 practice
352 items also randomly assigned to one of the possible colors. When participants mispronounced
353 or use a different labels to name the target objects and/or the target colors, they were corrected
354 by the experimenter, they were asked to repeat the correct labels and to remember these labels,
355 as they would see the same object/color in the next phase of the game.

357 2.4. Measures and annotations

358
359 One item that gave rise to object label errors in more than 75% was removed from the
360 analyses. Analyses were performed on pairs containing neither disfluencies/hesitations nor
361 object label errors for a total of 1461 pairs, thus removing 255 pairs (14.9%) out of the 1716
362 pairs. In these pairs, we analyzed both global prosodic variations and prosodic phrasing as a
363 marker of informational status of referents.

364
365 We first analyzed whether speakers modulate their global prosody depending on
366 whether they shared knowledge or not with their interlocutor. To do so, we measured speech
367 rate and pitch range. These are the two main global prosodic acoustic-phonetic features that

368 characterize clear speech production (see Smiljanic & Bradlow, 2009). We calculated average
369 speech rate in syllables/second for the entire pairs of noun-adjective fragments (e.g., ‘the red
370 mouse and the red house’). We first calculated the total duration of the entire pair of noun-
371 adjective fragments (see ‘Pair’ in Figure 2) according to the following procedure. Each entire
372 pair of noun-adjective fragments (e.g., *la souris bordeau et la maison bordeau* ‘the red mouse
373 and the red house’) had a base syllable count of 11 syllables. The first author listened to each
374 utterance and adjusted the base count for the very rare cases in which the speaker did not
375 produce the schwa at the end of the article *le* ‘the’ (e.g. [lə. mu. tɔ̃. ma. ʁɔ̃. e. **lbe.** ʁɛ. ma. ʁɔ̃]
376 instead of [lə. mu. tɔ̃. ma. ʁɔ̃. e. **lə. be.** ʁɛ. ma. ʁɔ̃] *le mouton marron et le beret marron* ‘brown
377 sheep and brown beret’). This syllable count was divided by the utterance duration to obtain a
378 rate measurement for each entire pairs of noun-adjective fragments. We then measured pitch
379 range variations. Two main types of pitch range variations have been reported in the literature:
380 the first is called ‘pitch level’ (Ladd, 1996) or ‘register’ (Cruttenden, 1997) and corresponds to
381 the raising or lowering of the melodic movements in the fundamental frequency (F0) space.
382 For instance, when we speak with higher pitch level than usual to imitate the voice of a child,
383 both F0 minima and maxima are produced higher so that the entire F0 contour is produced
384 higher in the F0 space. The other way in which pitch range is varied concerns the ‘pitch span’
385 (Ladd, 1996) also referred to as the ‘excursion size’ (‘t Hart et al., 2006). This measure refers
386 to the distance between the highest and the lowest pitches point in the contour. When the pitch
387 span is increased, F0 maxima are produced higher than usual while F0 minima stay the same.
388 Since clear speech research has reported ‘a wider dynamic pitch range’ (e.g., a larger pitch
389 span) when speakers are aware of perception difficulties from their listener, we measured pitch
390 range variations as pitch span variations in our study. We defined pitch span as the difference
391 between maximum and minimum F0 values for the entire pair of noun-adjective fragments.

392 Maximum and minimum F0 values were automatically extracted using Praat scripts (Boersma
393 & Weenink, 2015).

394

395 Second, in order to analyze whether the presence of shared knowledge between
396 participants affected the prosodic encoding of informational status of referents, we examined
397 whether speakers produced the noun in the 2nd fragment in the same AP as the following
398 adjective (e.g., [*maison bordeau*]_{AP} ‘red house’) or in a separate AP (e.g., [*maison*]_{AP}
399 [*bordeau*]_{AP}). Pre-boundary lengthening and the presence of a typical F0 rise aligned with the
400 last syllable of the AP are well known as the two main correlates of AP right boundaries in
401 French (Jun & Fougeron, 2002; Welby, 2006; Michelas & D’Imperio, 2012). Specifically, the
402 last full syllable of the AP has been shown to exhibit longer duration than unaccented syllables
403 within an AP. In addition, APs in non-final position within the utterance are typically
404 characterized by a rising F0 contour aligned with the last full syllable of the AP. For these
405 reasons, we measured duration of the first and second syllables of target nouns (respectively
406 S1 and S2 in Figure 2). We also extracted F0 values corresponding to the F0 minimum hertz
407 value of the 1st syllable of target nouns and to the F0 maximum hertz value of the last syllable
408 of target nouns. Based on these tagged features, we used two specific criteria to define the
409 presence of an AP right boundary (see Michelas et al., 2014 for the same procedure). First,
410 following the pitch accent detection criterion used by Astésano, Bard & Turk (2007) on French
411 data, we considered that an AP right boundary was actually produced by the speaker after the
412 target noun if the F0 maximum hertz value of the last syllable of the target noun was at least
413 10% higher than that of the F0 minimum hertz value of the 1st syllable of target nouns. Second,
414 we also verified that the syllable was lengthened. To do so, we checked whether the duration
415 of the last syllable of the target noun was at least 10% longer than that of the preceding syllable.

416 This is in line with Michelas et al. (2014) who found that healthy control speakers parsed
417 contrastive referents in a separate AP using two specific criteria to define the presence of an
418 AP boundary (a 10% threshold for pitch accent detection and a 10% threshold for syllable
419 lengthening). Michelas and D'Imperio (2012) reports results from a corpus study that compared
420 vowel duration in different prosodic positions (within words, in AP-final position, in ip-final
421 position and in IP-final position) at different speech rates (i.e., normal and fast speech rates)
422 and for different speakers (i.e., males and females). This study also demonstrated that AP-final
423 vowels are more or less 10% longer than a vowel contained within an AP and is thus in
424 accordance with the criterion used by Michelas et al. (2014).

425

426 All the performed annotations are summed up below and illustrated in Figure 2:

427

428 Duration measures:

429 **Pair:** entire noun-adjective pair

430 **det':** determinant of the 1st fragment

431 **S1':** 1st syllable of the 1st fragment

432 **S2':** 2nd syllable of the 1st fragment

433 **S3':** 3rd syllable of the 1st fragment

434 **S4':** 4th syllable of the 1st fragment

435 **det:** determinant of the 2nd fragment

436 **S1:** 1st syllable of the 2nd fragment

437 **S2:** 2nd syllable of the 2nd fragment

438 **S3:** 3rd syllable of the 2nd fragment

439 **S4:** 4th syllable of the 2nd fragment

440

441 F0 measures:

442 **F0min_Pair:** F0 minima for the entire pair of fragments

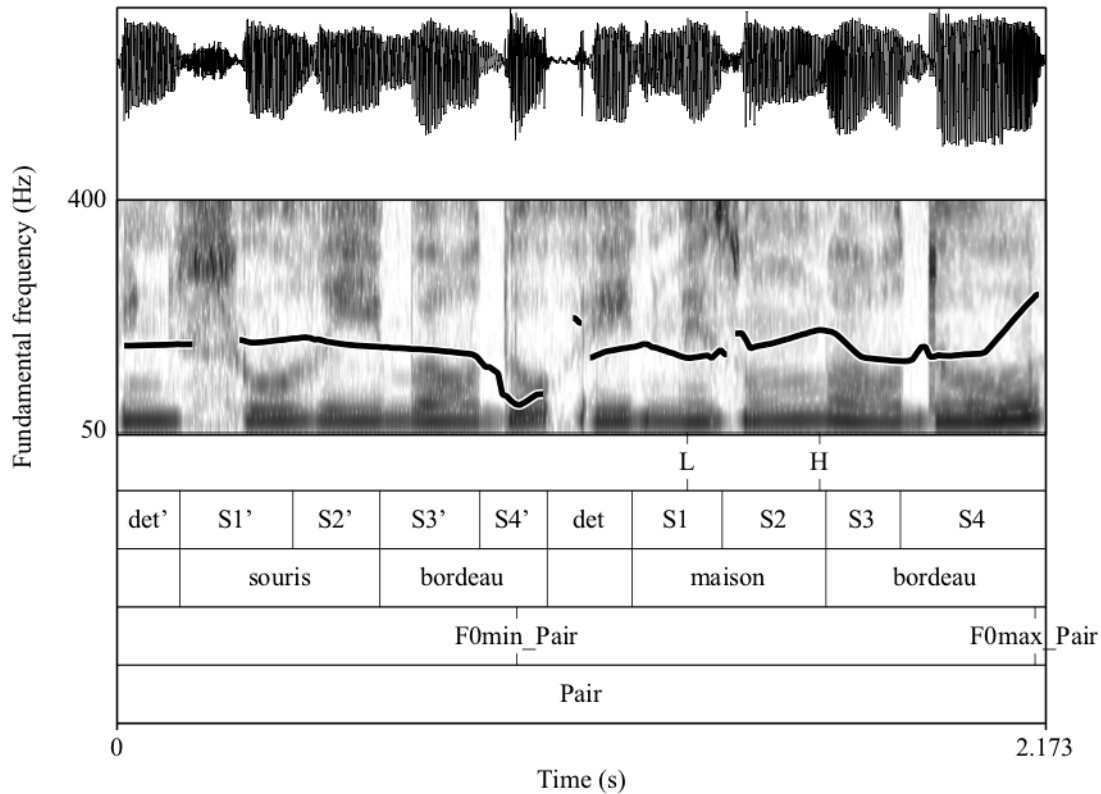
443 **F0max_Pair:** F0 maxima for the entire pair of fragments

444 **L:** low inflection point in the F0 curve near the beginning of the first syllable of the noun in
445 the 2nd fragment

446 **H:** F0 maximum hertz value of the last syllable of the same target noun

447

448 All the values corresponding to these labels were extracted thanks to Praat (Boersma &
 449 Weenink, 2015) scripts written by our own.



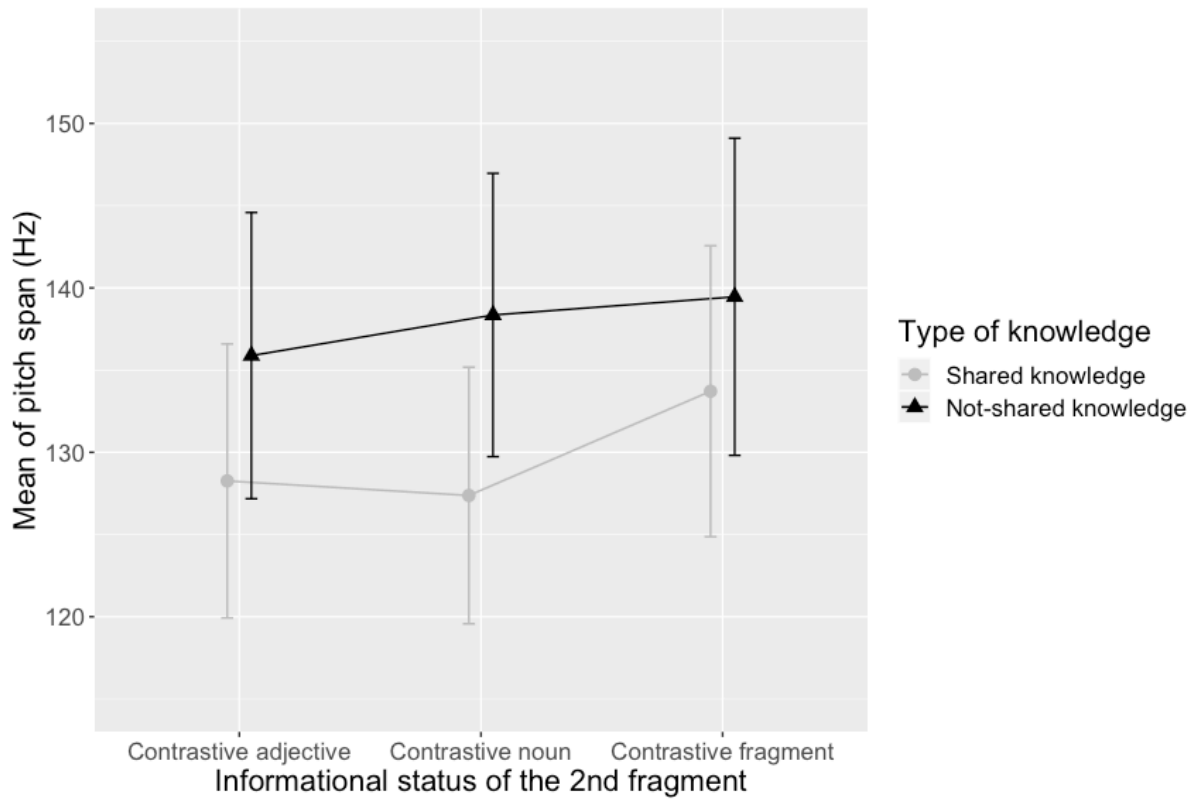
450
 451 Figure 2. Schema annotation for the pair *une souris bordeau et une maison bordeau* ‘the red
 452 house and the red house’.

3. Results

3.1. Global prosodic variations

459 Mean of pitch span (in Hz) and mean of speech rate (in syl/s) are shown in Figure 3
 460 and in Figure 4 respectively.

461



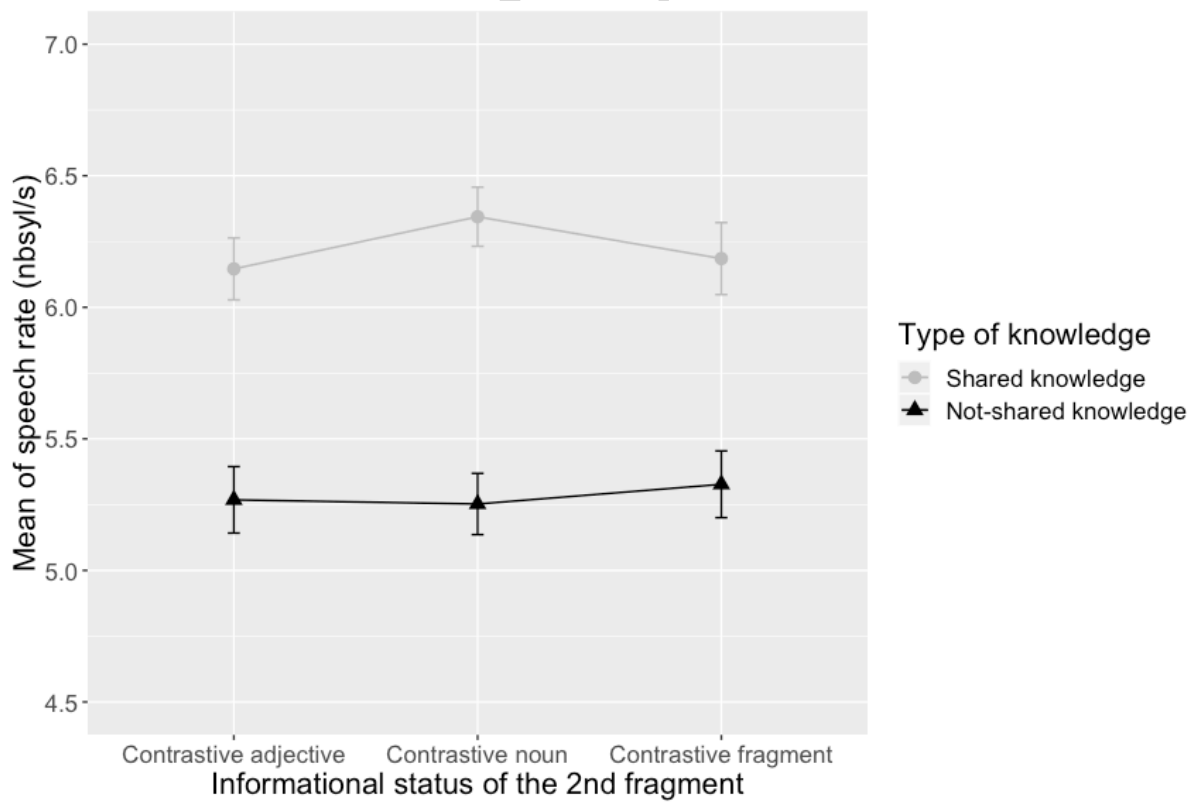
462

463

Figure 3. Mean of pitch span depending on the type of knowledge and the informational

464

status of the target fragment. Error bars show 95% confidence intervals.



465

466 Figure 4. Mean of speech rate depending on the type of knowledge and the informational
467 status of the target fragment. Error bars show a 95% confidence interval.

468

469 Two linear mixed-effects regression models were used on the log-transformed pitch span
470 and speech rate values. Taking the log transformations ensures a normal distribution of the
471 residuals (Baayen et al., 2008). Both models included the informational status, the type of
472 knowledge and the interaction between the two factors as fixed effects. Following Barr et al.
473 (2013), we used the maximal random effects structure that allowed the model to converge.
474 Participants and items were thus included as random intercepts (random slopes by participant
475 for the effect of informational status and type of knowledge were included only for the speech
476 rate model since they did not allow the pitch span model to converge). Global effects were
477 obtained using the afex::mixed function.

478

479 The pitch span model revealed a significant main effect of type of knowledge ($X^2 = 8.65$
480 $p < .01$) due to a larger pitch span in the not-shared knowledge condition compared the shared-
481 knowledge condition. The effect of informational status ($X^2 = 3.07$, $p > .20$) and the
482 informational status x type of knowledge interaction ($X^2 = 3.90$, $p = .14$) were not significant.
483 Since referents were repeated 6 times across our 6 experimental conditions and since repetition
484 is a well-established contributor to acoustic attenuation, we reran the same model with
485 repetition as a fixed factor. Repetition was coded R1-R6 and was not centered because it was
486 not a numerical variable. This additional model showed that there was no significant effect of
487 the repetition factor ($X^2 = 2.69$, $p > .20$).

488

489 The speech rate model revealed a significant main effect of type of knowledge ($X^2 =$
490 37.37 , $p < .0001$) due to faster speech rate in the shared knowledge condition compared to the

491 not-shared knowledge condition. The effect of informational status was not significant ($X^2 =$
 492 4.25, $p = .12$). The informational status x type of knowledge interaction was also significant
 493 ($X^2 = 7.60$, $p < .05$). Multiple comparisons (Tukey test, $p < .05$) obtained with the glht function
 494 from the multcomp package (Bretz et al., 2016) with p-values adjusted by the single step
 495 method revealed that this significant interaction was due to faster speech rate in the contrastive
 496 noun condition compared to the contrastive fragment condition ($z = 3.00$, $p < .05$) and a
 497 tendency for faster speech rate in the contrastive noun condition compared to the contrastive
 498 adjective condition ($z = 2.81$, $p = 0.056$) in the shared knowledge condition. As for the pitch
 499 span model, we reran our statistical model with repetition as a fixed factor. In a similar way to
 500 what happened for the pitch span model, there was no significant effect of repetition on the
 501 log-transformed speech rate values ($X^2 = 4.66$, $p > .20$).

502

503 3.2. Prosodic encoding of informational status

504

505 Proportion of items in the coded sample that included 2 APs depending on the type
 506 knowledge shared by participants and the informational status of target fragments shared is
 507 shown in Table 2.

508

509 Table 2. Proportion of items in the coded sample that included 2 APs phrasing (in %)

	Type of knowledge	
	Shared knowledge	Not-shared knowledge
Contrastive adjective	20	25
Contrastive noun	24	21

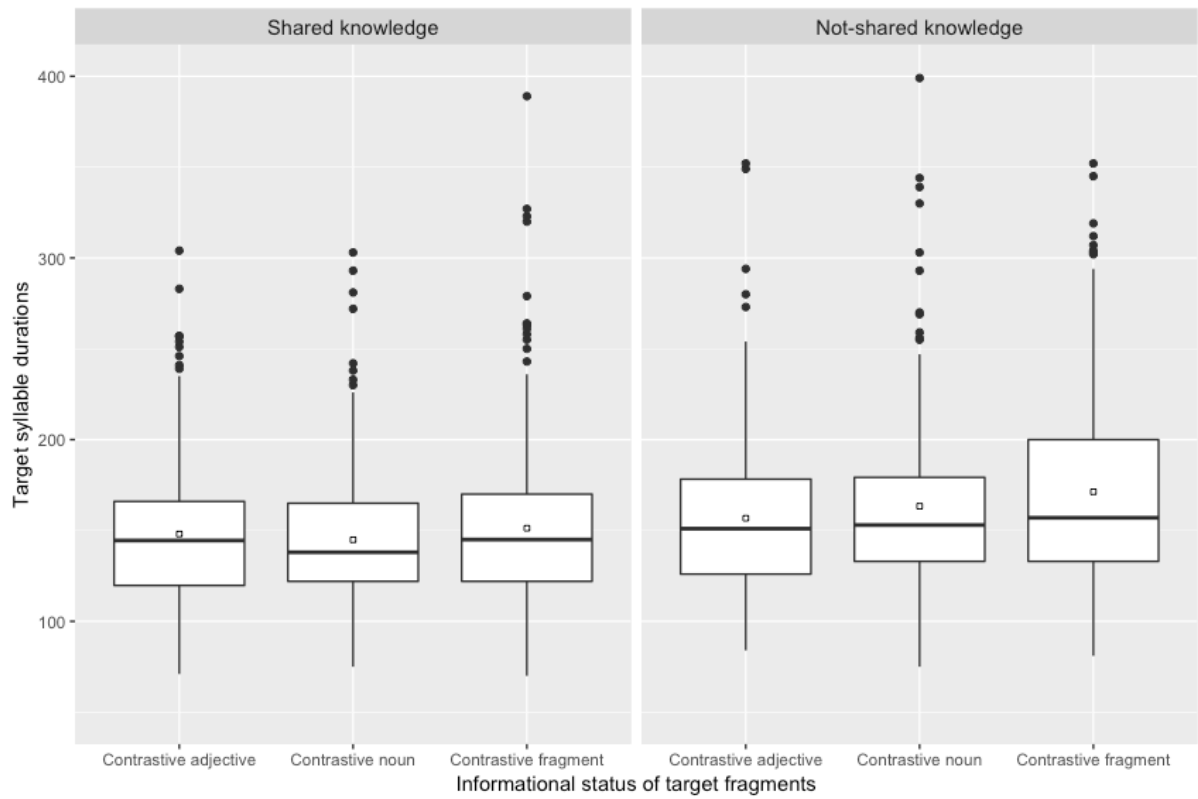
510

511 As shown in Table 2, there was a large bias in favor of the 1 AP phrasing compared to
512 the 2 AP phrasing regardless of the type of knowledge and of the informational status of target
513 fragments. We used a mixed-effects regression model (lme4 package in R-studio statistics
514 Version 0.99.903) with a logistic linking function to confirm the statistical relevance of this
515 bias. The model included the AP phrasing as dependent variable (1 = 2 APs, 0 = 1AP), the
516 informational status (contrastive adjective, contrastive noun, contrastive fragment), the type of
517 knowledge (shared-knowledge, not-shared knowledge) and the interaction between the two
518 factors as fixed effects. Following Barr et al. (2013), we used the maximal random effects
519 structure that allowed the model to converge. Participants and items were thus included as
520 random intercepts only (random slopes by participants and by items were not included because
521 they did not allow the model to converge). Neither the effect of the two factors (effect of
522 informational status: $X^2 = 0.24$, $p > .20$; effect of type of knowledge: $X^2 = 0.0004$, $p > .20$), nor
523 the informational status x type of knowledge interaction was significant ($X^2 = 2.00$, $p > .20$).
524 As with the global prosodic variations, we reran our statistical model with repetition as fixed
525 factor. The additional model showed that there was no significant effect of repetition on the
526 AP-phrasing produced by participant ($X^2 = 3.19$, $p > .20$).

527

528 We also fitted two linear mixed effects models on the acoustic parameters we used
529 (duration and F0 values associated with target syllables) to define the presence or the absence
530 of an AP at the right edge of target noun. Durations of target syllables in milliseconds (S2' in
531 Figure 2 above) are illustrated in Figure 5 while logarithms of F0 maxima associated with target
532 syllables (H in Figure 2 above) are illustrated in Figure 6.

533



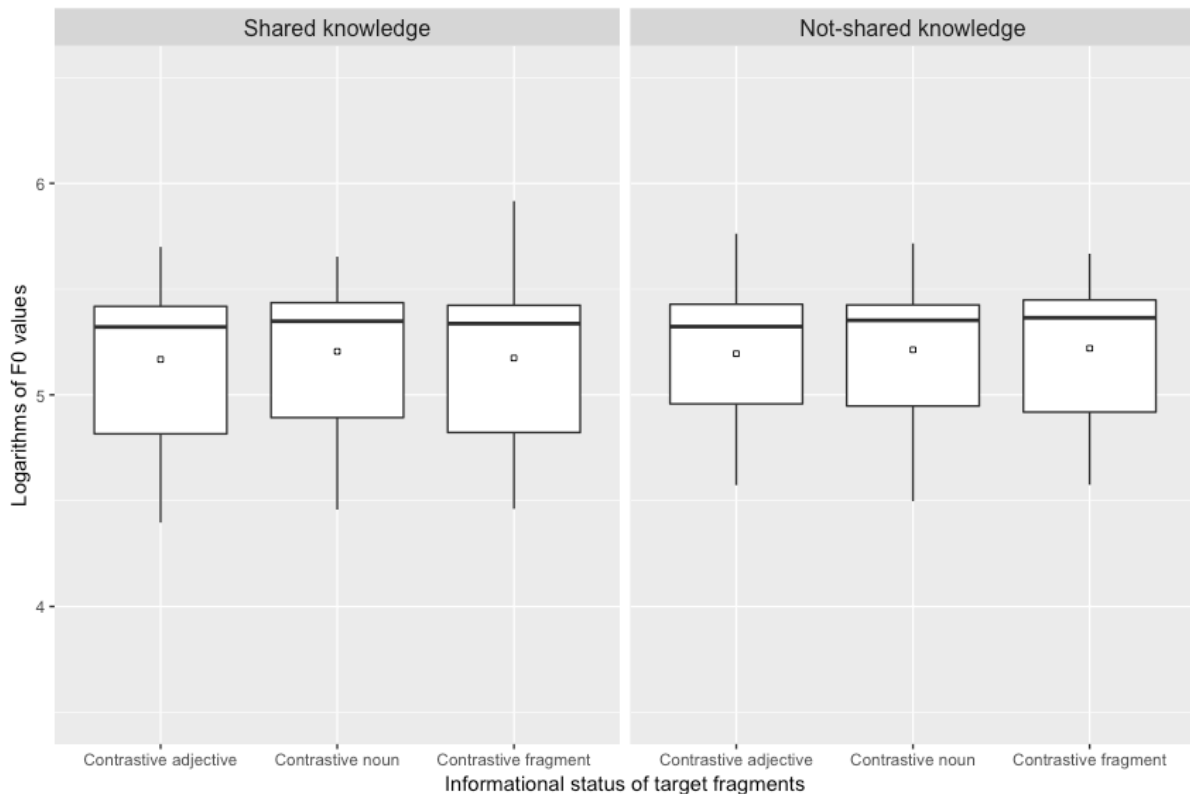
534

535 Figure 5. Boxplots of target syllable duration (in ms) depending on the informational status of
 536 target fragments and knowledge shared by participants.

537

538

Author's



539

540 Figure 6. Boxplots of logarithms of maximum F0 values associated with target syllables

541 depending on the informational status of target fragments and knowledge shared by

542 participants.

543

544 The two models included log-transformed values as dependent variables to ensure a

545 normal distribution of the residuals (Baayen et al., 2008). The informational status of referents

546 (contrastive adjective, contrastive noun, contrastive fragment), the type of knowledge (shared-

547 knowledge, not-shared knowledge) and the interaction between the two factors were included

548 as fixed effects. Following Barr et al. (2013), we used the maximal random effects structure

549 that allowed the model to converge. The duration values model also included random intercepts

550 by participant and by item and random slopes by participant for the two fixed effects. For the

551 F0 values model, participants and items were included as random intercepts only (random

552 slopes by participants and by items were not included because they did not allow the model to

553 converge).

554

555 For duration values, the effect of type of knowledge ($X^2 = 87.03$, $p < .0001$) was
556 significant due to longer values in the not-shared knowledge condition than in the shared
557 knowledge condition. The effect of informational status was also significant ($X^2 = 20.49$, $p <$
558 $.0001$) due to longer values in the contrastive fragment condition compared to the two other
559 conditions. The type of knowledge x informational status interaction was not significant ($X^2 =$
560 2.62 , $p > .20$). For F0 values, neither the effect of type of knowledge ($X^2 = 2.33$, $p = 0.13$) nor
561 the effect of information status ($X^2 = 4.68$, $p = .10$) reached significance. The type of knowledge
562 x informational status interaction was not significant ($X^2 = 0.93$, $p > .20$). For both models there
563 was no significant effect of the repetition when it was added to the models as a fixed factor (for
564 duration values: $X^2 = 2.62$, $p > .20$; for F0 values: $X^2 = 2.03$).

565

566

4. Discussion

567

568 The aim of the present study was to investigate whether the manipulation of shared knowledge
569 between interlocutors impacts the prosody produced by the speaker during semi-spontaneous
570 interactions. We were interested in two kinds of prosodic variations: global prosodic variations
571 that affect entire utterances (i.e., pitch range and speech rate variations) and more local prosodic
572 variations that encode informational status of words (i.e., prosodic phrasing to encode the
573 informational status of referents). To answer this question, we used a collaborative game in
574 which we manipulated both the presence/absence of shared knowledge between a director and
575 an addressee and the informational status of target words.

576

577 Based on previous studies, our first hypothesis was that, in the absence of shared
578 knowledge, participants would modify their global prosodic features to facilitate addressees'

579 comprehension. In line with our hypothesis, we found that directors spoke more slowly and
580 with larger pitch excursions when they did not necessarily share knowledge on target objects
581 with their addressee than when they did it. This result corroborates studies from clear speech
582 literature (see Smiljanic & Bradlow, 2009, for a review) by showing that directors adapt their
583 prosodic production to their addressee when they are aware of potential speech perception
584 difficulties on their part. It is important to note that our study did not explicitly test the
585 mechanism underlying such accommodation process. However, we can speculatively offer
586 some possibilities. An audience design explanation of the decrease of speech rate and increase
587 of pitch span we observed is that speakers ‘clarified’ their speech to increase addressee
588 understanding (Wright, 2004; Smiljanic & Bradlow, 2009). Within this framework, speakers
589 in the not-shared knowledge condition recognized that their addressees needed extra help. This
590 realization may have triggered a speaking mode similar to ‘clear speech’ that provided
591 additional information, which would help the addressee to complete the task. According to
592 Smiljanic & Bradlow (2009), clear speech modification refers to a distinct speaking style that
593 speakers adopt when they are aware of a speech perception difficulty on the part of the listener
594 due to background noise, a hearing impairment, or a different native language. All clear speech
595 research converges on the idea that clear speech typically involves a wide range of
596 acoustic/articulatory adjustments, including a decrease in speech rate (longer segments as well
597 as longer and more frequent pauses), wider dynamic pitch range, greater sound-pressure levels,
598 more salient stop releases, greater rms intensity of the non-silent portions of obstruent
599 consonants, higher-voice intensity, vowel space expansion and vowel articulation. In our
600 study, the not-shared knowledge condition induced a decrease in speech rate and a wider
601 dynamic pitch range which are the two main acoustic adjustments observed in case of clear
602 speech contexts. We can thus reasonably interpret these two modifications as resulting from an

603 adjustment the speaker made because (s)he anticipated a perception difficulty from his/her
604 conversational partner, which is exactly the definition of “clear speech”.

605 However, as for the majority of clear speech studies conducted in the past, we cannot
606 exclude that the effect of not-shared knowledge in our study is not the result of audience design
607 per se, but rather results from production constraints on the speaker’s part (see also Arnold et
608 al., 2012; Bard et al., 2000; Bard & Aylett, 2004; Fraundorf et al., 2015). Specifically, not-
609 shared knowledge between interlocutors could have slowed down the planning process of
610 speech which results in a decrease of the speaker’s speech rate and an increase of his/her pitch
611 excursions. Our results are thus also consistent with a mechanism in which not-shared
612 knowledge between interlocutors complicates production processes on the speaker’s part,
613 rather than induce adjustment of prosodic forms on the basis of a specific representation of the
614 addressee’s needs. Thus, research must be further developed to determine whether shared
615 knowledge manipulation affects prosody of the speaker via planning effects or not.

616

617 Regardless of the mechanism underlying the effect of shared knowledge we observed
618 on speech rate and pitch range, our result can also be interpreted in parallel with those of Rosa
619 et al. (2015). In their study, speakers participated in two instruction-giving experiments in
620 which they gave instructions to listeners to move objects to locations on a board. The attention
621 of the listener was manipulated through the number of tasks in which they were involved: in
622 the distraction condition, addressees might experience difficulties since they were also
623 completing a demanding secondary computer task; in the attentive condition they paid full
624 attention to the main task. The authors found that speakers use longer words for distracted
625 listeners compared to attentive listeners. Since in our study we found more slowly speech rate
626 (measured as number of syllables/s) and longer durations of the last syllable of target words

627 when knowledge was not-shared between interlocutors compared to when it was shared, our
628 result is in line with those of Rosa et al., by showing that speakers modify duration of speech
629 segments when they were aware of an interlocutor's difficulty. For Rosa et al., this perception
630 difficulty consisted in distraction of listeners' attention, while in our study, this perception
631 difficulty was due to an absence of shared knowledge between interlocutors. However, it is
632 important to note that in our study the modification of durations of the last syllable of target
633 words resulted from a global speech rate modification. By contrast, in Rosa et al.'s study, this
634 duration effect seems to be localized to the target word and does not appear to affect global
635 speech rate. However, given that the authors have not explained how speech rate was
636 calculated, one could argue that the different results obtained could be linked to the different
637 portion of speech segments used to define speech rate. The entire fragment was used in our
638 study whereas smaller portions of speech segments appear to have been used in Rosa et al.'s
639 study. Moreover, it is interesting to notice that in both studies, duration of target words did not
640 vary depending on whether target nouns were contrastive/unpredictable vs. non-
641 contrastive/predictable, suggesting that in these two kinds of tasks, participants did not
642 prosodically encode informational status of words.

643

644 Indeed, in our study, the observation that directors more often produced 1 AP phrasing
645 (between 75% and 80%) rather than 2 AP phrasing whatever the knowledge shared by
646 participants and the informational status of words was unexpected. Specifically, when
647 knowledge was shared, we expected participants to prosodically encode the contrastive status
648 of target nouns by producing target nouns in a separate AP from the following adjective. We
649 thus expected more 2 AP phrasing in the contrastive noun condition compared to the contrastive
650 adjective condition in the shared knowledge condition. By contrast, speakers did not increase
651 their 2 AP phrasing productions regardless of the informational status of target noun and of the

652 knowledge shared with their addressee. This result contrasts with previous studies in the field
653 reporting that speakers phrased the contrastive information in a separate AP when knowledge
654 is shared between interlocutors (Féry, 2001; Dohen & Loevenbruck, 2004; Michelas et al.,
655 2014).

656

657 A way to explain this result is to consider the cognitive load of the task due to visual
658 salience effects. Interestingly, Kantola & Van Gompel (2011, 2016) showed an effect of the
659 presence of visual competitors on linguistic choices of the speakers when an addressee was
660 present while this visual effect of the competitor disappeared in the absence of an addressee.
661 In their experimental paradigm, they manipulated the visual saliency of the referent by
662 presenting a picture in which a visual competitor was present or not. The authors found an
663 effect of the visual competitor on the choice of the referring expression: speakers used fewer
664 pronouns and more repeated noun phrases when a visual competitor was present in the scene
665 than when there was no visual competitor. The authors concluded that visual salience effects
666 were due to adjustments that speakers make when they speak to an addressee. Damen et al.
667 (2019) did not find such adjustment to the addressee. Using a referential communication task,
668 they showed that giving an explicit focus on addressee's perspective did not influence reference
669 production of the speaker. While speakers were able to engage in perspective taking, they did
670 not use this knowledge during the task. In other words, they referred to their privileged
671 knowledge, disregarding the addressee's perspective. In the present study, visual context
672 always contained two target objects and a competitor regardless of the type of knowledge
673 condition. In line with Kantola & Van Gompel (2011, 2016), we hypothesized that considering
674 the visual context of the target objects (meaning the two target objects plus the competitor)
675 would require a lot of attention making the task cognitively demanding for the director.
676 Interestingly, in Michelas et al. (2014)'s study in which control speakers prosodically encoded

677 the informational status of referents, no competitor (i.e., no object of the same type but with a
678 different color for the given noun condition and no object of the same color but a different type
679 for the contrastive noun condition) was present in the map for each given location. The absence
680 of such competitors in the visual context could have made the task less cognitively demanding
681 for speakers than in the current study. Thus, assuming a high cognitive load of our task, it could
682 be the case that directors were not able to keep track of the salience of objects (in terms of non-
683 contrastive/contrastive status) and thus did not prosodically encode this salience. It is
684 interesting to note that this difficulty of speakers under cognitive load to take discourse salience
685 into account was also found by Vogels et al. (2015). In this study, speakers under cognitive
686 load (they had to spread attentional resources between two tasks) had more difficulties to use
687 pronouns to refer to less salient referents than speakers who performed only one task. Similarly,
688 in our task, it can be the case the speakers under cognitive load had chosen the most economical
689 linguistic forms for themselves which consisted in not using prosodic phrasing to encode
690 informational status of target nouns. But regardless of the cognitive load of our task, it is worth
691 noting that in 100% of cases, the participants successfully completed the interactive task. There
692 was no case in which the task was interrupted because of great difficulties to complete it.
693 Moreover, it is interesting to note that in only 2% of cases, the confederate put the cross in the
694 wrong place showing that in 98% of cases, speakers succeeded in guiding the confederate to
695 put the cross in the right place. These two points reinforce the idea that the task we have created
696 is not a burdensome task for the participants but that it requires more attentional resources than
697 a task such as the one used by Michelas et al. (2014) in which the participants can focus on the
698 prosodic encoding of the informational status of referents.

699 Interestingly, when taking a closer look at the acoustic cues of prosodic phrasing
700 produced by our participants, we found an effect of shared-knowledge manipulations on
701 duration of target syllables but not on height of these syllables. This effect confirms the fact

702 that the increase in duration we observed in the not-shared condition resulted from global
703 speech rate manipulation and did not occur specifically on target words. Indeed, if this
704 modification of duration has translated variations of prosodic phrasing, we would have
705 expected that the height of the syllable increased at the same time as the duration of this one.
706 However, this was not the case and this confirms the relevance of the criteria we used to
707 differentiate cases in which target nouns were produced in the same AP as the following
708 adjectives (1 AP phrasing) from cases in which they were produced in a separate AP (2 AP
709 phrasing; see also Michelas et al. 2014 for a similar procedure). Moreover, the fact that syllable
710 duration did not vary as a function of the informational status of target words also indicates
711 that the increase of target duration we observed reflects global speech rate modification and
712 not acoustic prominence to encode informational status of words.

713

714 Concerning the speaker's global prosodic modifications we observed in the not-shared
715 condition, one could argue that these modifications could be impacted by the fact that, in our
716 task, the addressee was a confederate. Indeed, we know the attributions that language users
717 have about their task partner can impact how they adapt their language (Duran, Dale & Kreuz,
718 2011) and that using a confederate in the addressee role might be risky for studies of language
719 in dialogue contexts (Kuhlen & Brennan, 2003). For these reasons, we run an additional pilot
720 study using exactly the same procedure as previously described except that the additional
721 participants interacted with other naïve participants and not a confederate (see Appendix III).
722 These preliminary findings are in line with the current study and suggest that participants
723 tended to modify their global prosodic variations (i.e., pitch range and speech rate variations)
724 when speaking to naïve addressees in the same way as they did when speaking to a confederate
725 while they tended to not prosodically encode the informational status of referents regardless of
726 the knowledge condition. Thus, more research is now required to test our hypothesis that if the

727 speaker has to take into account the possible visual competitors to give relevant information to
728 his/her addressee each time s/he prepares the prosodic encoding of referents, s/he would
729 probably not encode the informational status of these referents. In a more general manner, a
730 large number of previous studies considers that the difficulty of the communicative task may
731 influence the degree to which speakers appear to be modeling their listeners (see for instance
732 Horton & Keysar, 1996; Bard et al., 2000; Bard & Aylett, 2004; Kantola & Van Gompel, 2011,
733 2016; Rosa et al., 2015). Thus, the next step of this research will be to determine the constraints
734 under which speakers could be more/less sensitive to addressee's knowledge.

735

736 To conclude, the present study reveals that in French the speaker's knowledge of what
737 the addressee knows or does not know affects his/her prosodic choices. When knowledge was
738 not shared, we found that speakers modified their global prosody in terms of speech rate and
739 pitch span modulations. However, and unexpectedly, we also found that speaker-internal
740 constraints seemed to affect local prosodic encoding of informational status of words
741 production since our participants who were involved in a more cognitive demanding task than
742 in previous studies (e.g., Michelas et al., 2014) did not modify their AP phrasing to encode
743 pragmatic status of referents when knowledge was shared. These results are in line with the
744 idea that complex reasoning usually implicated in constructing a model of the addressee co-
745 exists with speaker-internal constraints such as cognitive load of the task to affect the
746 production of speech (Bard et al., 2000; Bard & Aylett, 2004).

747

Acknowledgements

We are grateful to Alexia Galati, Emiel Kraemer and one anonymous reviewer for their helpful comments on earlier versions of the manuscript.

Author's version

References

- Astésano, C., Bard, E. G., & Turk, A. (2007). Structural influences on initial accent placement in French. *Language and Speech*. 2007;50(3):423-446.
- Arnold, J. E., Kahn, J. M., Pancani, G. C. Audience design affects acoustic reduction via production facilitation. *Psychonomic bulletin & review*. 2012;19(3):2012: 505-512.
- Baayen, R. H., Davidson, D. J., Bates, D. M. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., Newlands, A. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*. 2000; 42: 1–22.
- Bard, E. G., Aylett, M. P. Referential form, word duration, and modeling the listener in spoken dialogue. In: Trueswell JC, Tanenhaus MK, editors. *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press; 2004. pp. 173–191
- Beysade, C., Hemforth, B., Marandin, J.-M., Portes, C. “Prosodic Markings of Information Focus in French”. In : Yoo HY, Delais-Roussarie E, editors. *Actes d’Interfaces, Discours et Prosodie*. Paris, ISSN 2114-7612; 2009. pp. 109-122.
http://makino.linguist.jussieu.fr/idp09/actes_fr.html.
- Boersma, P., Weenink, D. Praat. Doing phonetics by computer (Version 5.4.01, 2015),
Computer program: www.praat.org.
- Bradlow, A. R., Kraus, N., Hayes, E. Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*. 2003; 46(1): 80-97.

- Bretz, F., Hothorn, T., Westfall, P. Multiple comparisons using R. CRC Press. 2016.
- Chafe, W. (1976). Givenness, contrastiveness, subject, topic, and point of view. In: Li C, editors. Subject and Topic. New York: Academic Press; 2016. pp. 25-55.
- Chen, A., Destruel, E. Intonational encoding of focus in Toulousian French. In Proceedings of Speech Prosody; 2010.
- Clark, H. H., Haviland, S. E. Comprehension and the given-new contract. In Freedle RO: Discourse production and comprehension. Norwood, NJ: Ablex Publishing; 1977. pp. 1-40.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language. 2013; 68(3):255-278.
- Cruttenden, A. Intonation. Cambridge: Cambridge University Press; 1997.
- Damen, D., van der Wijst, P., van Amelsvoort, M., & Kraemer, E. Perspective-Taking in Referential Communication: Does Stimulated Attention to Addressees' Perspective Influence Speakers' Reference Production? Journal of Psycholinguistic Research. 2019, 48 (2), 257-288.
- Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. Cognition; 121(1), 22-40.
- Dohen, M., Løevenbruck, H. Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French. In: Proceedings of Interspeech; 2004.
- Ferguson, S. H., Kewley-Port, D. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. The Journal of the Acoustical Society of America. 2002; 112(1): 259-271.

- Féry, C. Intonation of focus in French. In: Féry C., Sternefeld W, editors. *Audiatur Vox Sapientes: A Festschrift for Arnim von Stechow*. Berlin: Akademi Verlag; 2001. pp. 153-181.
- Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. Reduction in prosodic prominence predicts speakers' recall: implications for theories of prosody. *Language, cognition and neuroscience*. 2015; 30(5): 606-619.
- Halliday, M. A. K. Notes on transitivity and theme in English, part2. *Journal of Linguistics*. 1967; 3:199–244.
- Hart, J. t', Collier, R., Cohen, A. *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge University Press; 2006.
- Horton, W. S., & Keysar, B. When do speakers take into account common ground?. *Cognition*. 1996; 59(1): 91-117.
- Jun, S.A., Fougeron, C. Realizations of accentual phrase in French. *Probus*. 2002; 14:147-172.
- Kaland, C., Krahmer, E., & Swerts, M. White bear effects in language production: Evidence from the prosodic realization of adjectives. *Language and speech*. 2014; 57(4): 470-486.
- Kantola, L., van Gompel, R. P. Does the addressee matter when choosing referring expressions?. *Proceedings of PRE-Cogsci: Bridging the gap between computational, empirical and theoretical approaches to reference*. Boston; 2011.
- Kantola, L., van Gompel, R. P. Is anaphoric reference cooperative? *The Quaterly Journal of Experimental Psychology*. 2016; 69:6:1109-1128.
- Krause, J. C., Braid, L. D. Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*. 2004; 115(1); 362-378.

- Kuhlen, A. K., & Brennan, S. E. Language in dialogue: when confederates might be hazardous to your data. *Psychonomic bulletin & review*. 2013; 20(1): 54-72.
- Ladd, Robert D. *Intonational Phonology*. Cambridge University Press, Cambridge; 1996.
- Liu, S., Del Rio, E., Bradlow, A. R., Zeng, F. G. Clear speech perception in acoustic and electric hearing. *The Journal of the Acoustical Society of America*. 2004; 116(4): 2374-2383.
- Matthies, M., Perrier, P., Perkell, J. S., Zandipour, M. Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research*. 2001; 44(2): 340-353.
- Michelas, A., D'Imperio, M. When syntax meets prosody: Tonal and duration variability in French Accentual Phrases. *Journal of Phonetics*. 2012; 40(6) : 816-829.
- Michelas, A., Faget, C., Portes, C., Lienhart, A.-C., Boyer, L., Lançon, C., Champagne-Lavau, M. Do patients with schizophrenia use prosody to encode contrastive discourse status?. *Frontiers in Psychology*. 2014; 5:755.
- Michelas, A., Cau, C., & Champagne-Lavau, M. (2017). How does the absence of shared knowledge between interlocutors affect the production of French prosodic forms?. *Proc. Interspeech 2017*, 3191-3195.
- Perkell, J. S., Zandipour, M. Economy of effort in different speaking conditions. II. Kinematic performance spaces for cyclical and speech movements. *The Journal of the Acoustical Society of America*. 2002; 112(4):1642-1651.
- Picheny, M. A., Durlach, N. I., Braida, L. D. Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*. 1986: 29(4), 434-446.
- Picheny, M. A., Durlach, N. I., Braida, L. D. Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility

- between clear and conversational speech. *Journal of Speech, Language, and Hearing Research*. 1989; 32(3), 600-603.
- Prince, E. F. On the syntactic marking of presupposed open propositions. In: Farley A, Farley P, McCullough KE, editors. *Papers from the Parasession on Pragmatics and Grammatical Theory. 2nd Regional Meeting in Chicago, IL: Chicago Linguistic Society*; 1986. pp. 208-222.
- Rosa, E. C., Finch, K. H., Bergeson, M., Arnold, J. E. The effects of addressee attention on prosodic prominence. *Language, Cognition and Neuroscience*. 2015; 30(1-2): 48-56.
- Smiljanić, R., Bradlow, A. R. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass*. 2009; 3(1): 236-264.
- Smiljanić, R., Bradlow, A. R. Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, 2005; 118(3):1677-1688.
- Swerts, M., Krahmer, E., & Avesani, C. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*. 2002; 30(4) : 629-654.
- Wright, R. A. Factors of lexical competition in vowel articulation. In: Local JJ, Ogden R, & Temple R, editors. *Laboratory phonology*. Cambridge, U.K.: Cambridge University Press; 2004. Vol. 6, pp. 26–50.
- Vogels, J., Krahmer, E., Maes, A. How cognitive load influences speakers' choice of referring expressions. *Cognitive science*. 2015; 39(6): 1396-1418.
- Welby, P. French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*. 2006; 34(3): 343–371.

Supporting information captions

Appendix I. Complete instructions given to participants.

Appendix II. Extract from the dialogue between one participant (the director) and the confederate (the addressee) corresponding to the screen views shown in Figure 1.

Appendix III. Pilot study including 5 additional participants that interacted with other naïve participants.

Author's version

Appendix I. Instructions given to participants.

The confederate - who was introduced to the participants as the experimenter but was naïve about the scientific aims of the experiment - gave instructions to them.

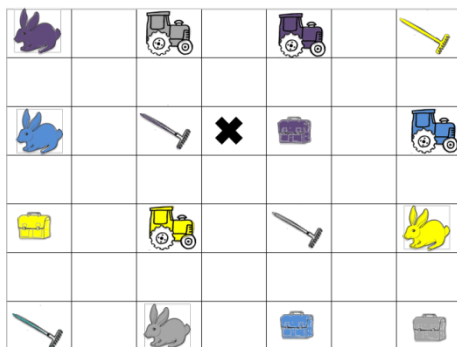
During this experience, we will play a game together.

We will both see different objects and different colors on our respective screens. These objects will appear on grids and we will have the same grids at the same time.

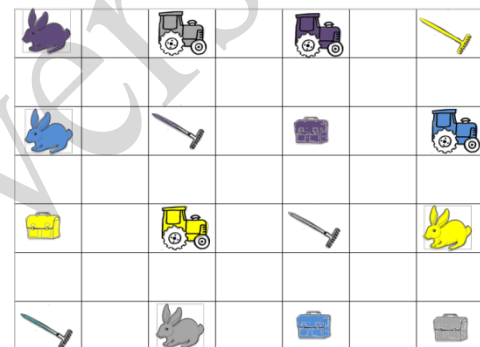
For each grid, you will see a cross on your own screen, but I won't have this cross on my own screen. Your task is to tell me where to place the cross between different objects by telling me the type of objects and the color of objects of the two critical objects between which the cross is located.

For instance, in the following grid, **the cross is located between the purple rake and the purple schoolbag.**

YOUR GRID

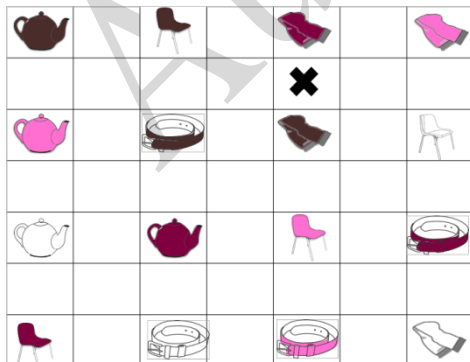


MY GRID

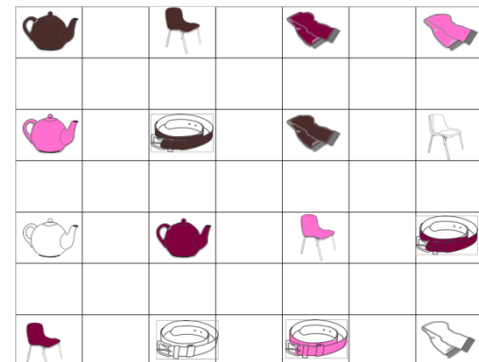


The cross can be located vertically (such as in the picture above and in this case you have to tell me the objects from left to right) or the cross can be located vertically (such as in the picture below and in the case you have to tell me the objects from top to bottom).

YOUR GRID



MY GRID



The cross is located between the red scarf and the brown scarf.

Once I have placed the cross on my grid, we will click together on the OK button to bring up

the next grid.

To make the game more difficult, in some grids there will be white boxes while in some others, there will be black boxes. In case of white boxes, you are sure that I share exactly the same objects and colors as you. But in case of black boxes, it is possible that you have either a different type of object or a different color of object as me in those boxes.

At the end of the game, we win if I have correctly placed all the crosses in the right places.

We are going to start with 6 practice grids to make sure everything is going well.

But before that, we are going to make sure that we name all the objects and the colors in the same manner. To do so, you are going to tell me what you see in the following pictures by naming the objects by their type and their color.

Do you understand? Do you have any questions?

Author's version

Appendix II. Extract from the dialogue between one participant (the director) and the confederate (the addressee) corresponding to the screen views shown in Figure 1.

Director : *ok donc là tu mets la croix entre la souris bordeau et la maison bordeau et les deux sont dans des cases noires.*

Addressee: *alors je vois bien la souris bordeau mais je n'ai pas de maison bordeau à proximité. Où est-ce que je dois mettre la croix par rapport à la souris bordeau ?*

Director : *En fait il faut que tu mettes la croix à droite de la souris bordeau ce qui fait que la croix se situe dans la deuxième colonne de la grille.*

Adresse : *ok parfait je vois. C'est bon j'ai mis la croix.*

Director: Ok then you put the cross between the red mouse and the red house, and both are in black boxes.

Addressee: Well, I can see the red mouse, but I don't have a red house nearby. Where do I have to put the cross relative to the red mouse?

Director: You have to put the cross on the right of the red mouse so that the cross is in the second column of the grid.

Addressee: ok perfect I see. I put the cross, It's fine for me.

Appendix III. Pilot study including 5 additional participants that interacted with other naïve participants.

Method:

Participants: Five participants participated in the pilot experiment. None has participated in the main experiment.

Material and procedure: The material and procedure were the same as in the main experiment except that the additional participants interacted with other naïve participants and not a confederate.

Results: The same analyses as previously described were conducted. These analyses showed that participants increased their pitch range (154 Hz vs. 165 Hz; $X^2 = 5.19$, $p < .05$) and decreased their speech rate (5.56 syl/s vs. 5.07 syl/s; $X^2 = 8.00$, $p < .01$) in the not-shared knowledge condition compared to the shared condition when speaking to naïve addressees in the same way as they did when speaking to a confederate. Moreover, in these additional analyses, the informational status of referents x type of knowledge interaction was not significant ($X^2 = 3.90$, $p = .142$) showing that participants did not use more 2 APs than 1 APs phrasing when the noun was contrastive (38% of items included 2 APs) compared to when the adjective was contrastive (32% of items included 2 APs) in the shared knowledge condition.