Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Intonation in the processing of contrast meaning in French: an eye-tracking study

Núria Esteve-Gibert[1], *Cristel Portes*[1,2], *Amy Schafer*[3], *Barbara Hemforth*[4], *Mariapaola D'Imperio*[1,5]

[1] Aix Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France
[2] IMS, University of Stuttgart, Germany
[3] University of Hawai'i at Mānoa, Hawai'i
[4] Laboratoire de Linguistique Formelle, CNRS, Université Paris Diderot
[5] Institut Universitaire de France (IUF)

nuria.esteve-gibert@blri.fr, cristel.portes@lpl-aix.fr, aschafer@hawaii.edu, barbara.hemforth@linguist.jussieu.fr, mariapaola.dimperio@lpl-aix.fr

Abstract

Listeners rapidly process tonal composition and pitch accent placement within an utterance to create expectations about its pragmatic meaning and information structure. It is still unknown whether the nuclear pitch accent alone or a combination of pitch accent and the following edge tone are needed in order to process intonational meaning in French. This study investigates the online comprehension of the French (L)H*L% rise-fall "implication" contour, which evokes a contrast meaning. Twenty-nine speakers participated in an eye-tracking experiment. The critical stimuli were sentences whose interpretation could be anticipated by successfully processing the implied meaning evoked by the (L)H*L% rise-fall contour on the critical word (hereafter CW). The results showed that participants are able to associate the implication contour with a contrast meaning, and that they start doing this only after the H* peak of the rise-fall intonation movement has been processed, hence when part of the L% falling movement has been perceived.

Index Terms: intonation processing, eye-tracking, implication contour, French

1. Introduction

Intonation is a tool to express pragmatic meaning and information structure. In English, for instance, the L+H* pitch accent can signal a contrastively focused element, for which a speaker selects an element within a set of possible alternatives [1]. The intonational structure of French differs widely from that of English: the domain of stress is larger than the word, being the Accentual Phrase (AP) [2], with prosodic prominence marking the right edge of the AP. Hence, the presence of prosodic prominence on a specific element does not necessarily indicate that that element is contrastively focused. Instead, other intonation cues like an initial rise (LHi) on the left edge of the focused constituent or post-focal pitch range compression is probabilistically employed for contrastive focus use (e.g. [2]–[4]). Moreover, in French there is no specific pitch accent indicating that an element is contrastively focused. Instead, several intonation contours with various degrees of speaker commitment and attitude attribution are found to occur in contrastively focused contexts: a rising H*H%, found in confirmation questions, a rise-fall-rise H+!H*H%, indicating disbelief on the part of the speaker relative to her interlocutor's proposition, or a rise-fall (L)H*L%[1], called the "implication contour", found when there is a contrast between the interlocutors' beliefs [5], [6].

Online processing of intonation in contrastively focused elements has been studied mainly in stress-accent languages such as English or German, revealing that speakers use pitch accent location and pitch accent type to create online expectations about an upcoming referent [7]–[12]. In [9], participants first heard a sentence like "Put the candy/candle below the triangle" and then a sentence like "Now put the candle above the square", where 'candle' could either be accented or deaccented. Participants' looks to the competitor 'candy' when hearing the second sentence showed that listeners interpreted the noun as anaphoric when it was deaccented and as non-anaphoric when accented.

[13] found that speakers rely on the placement and type of pitch accent to infer contrast in American English. Participants heard a sentence like "Hang the green drum" that was followed by a sentence like "Now hang the BLUE drum" or "Now hang the blue BALL". Their results showed that in the first case, fixations were speeded to the target image if a L+H* accent was placed on the adjective, while this was not the case when the same pitch accent was placed on the noun. Interestingly, in sentences like "Now hang the BLUE ball", participants expected the target noun to be "drum" and not "ball" when they heard a contrastive accent on the adjective, and produced eye fixations on a set of drums even when segmental cues had already disambiguated the target.

As for French, intonational meaning has not yet been explored with online techniques. Moreover, since the contour under investigation is composed of a rise, which is quite frequent in AP-final position in French, and a subsequent fall, we also ask whether listeners can infer the contrast meaning once the pitch accent rise has been perceived or if they have need to process the following L% [see 11 for English]. The present study employed the Visual World eye-tracking

[1] Note that we employ a transcription for the implication contour that is more transparent, since it includes the preceding L(ow) target, though we include it in parentheses to indicate that its contrastive status is still not known

 doi: 10.21437/SpeechProsody.2016-252

paradigm [14] to investigate if French speakers associate the (L)H*L% rise-fall "implication" contour to a specific meaning, i.e. a contrast between the interlocutors' beliefs, and if so, at which point during the contour listeners would be able to extract the contrastive meaning.

2. Methods

2.1. Participants

Twenty-nine French-speakers were recruited in the Paris area (7 males). Four additional participants took part in the study but were excluded due to experimental errors (N=2) or to exclusive fixations to the center of the screen (N=2).

2.2. Materials

Eighteen test suggestion-response sentence pairs were created to evoke a dialogue in a card game in which players guess which cards the other player holds. The critical stimuli pair had the form of *Je pense que tu as un/e X* ('I think you have a/an X') – *J'ai un/e X* ('I have a/an X'). All suggestions were produced with a falling LHiL*L% intonation and included a homophone in phrase-final position, presented with a visual display that depicted the subordinate alternative (Fig. 1, left panel).

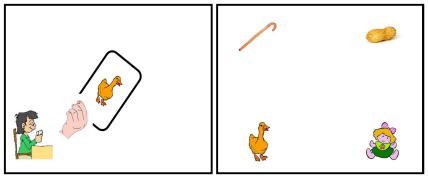

Figure 1: *Visual displays in suggestions (left) and responses (right).*

Condition	Target	Competitor (1)	Competitor (2)	Unrel. Dis.
Confirmation-homophone	*Cane* 'Duck'	*Canne* 'Stick'	-	*Poupée* 'Doll' *Cacahouète* 'Peanut'
Contrast-homophone	*Canne* 'Stick'	*Cane* 'Duck'	-	*Poupée* 'Doll' *Cacahouète* 'Peanut'
Contrast-control	*Poupée* 'Doll'	*Cane* 'Duck'	*Canne* 'Stick'	*Cacahouète* 'Peanut'

Table 1: *Summary with an example of the images corresponding to each position across conditions.*

Critical responses were of three types (Table 1): (a) confirmation-homophone sentences, produced with a LHiL*L%, including the same homophone as in the suggestion (e.g., *cane* 'duck') and followed by the segmental disambiguation *bien sûr* plus a clarifying phrase, e.g., *l'animal* ('indeed, the animal'); (b) contrast-homophone sentences, produced with a (L)H*L%, followed by the segmental disambiguation *plutôt*, and clarification to the dominant alternative of the homophone pair ('… a cane, instead, for walking'); (c) contrast-control sentences, produced with a (L)H*L% and followed by the segmental disambiguation *plutôt, […]* ('instead, […]'), but including a non-homophone CW. Fig. 2 shows the spectrograms of the three possible responses.

In scenes accompanying test responses (Fig. 1, right panel), the image at the bottom left always coincided with the suggested word, while counterbalanced in the other positions were the alternative interpretation of the homophone, the target non-homophone of the contrast-control sentences, and a non-homophone unrelated distractor.

Thirty-six filler sentence pairs were also created, with the same form as test sentence pairs but including non-homophones in the critical position. There were two types of filler pairs: (a) filler-confirmation pairs (N=21), with a non-homophone in the suggestion and the same non-homophone in the response, produced with a LHiL*L% confirmation intonation; (b) filler-contrast pairs (N=15), with a non-homophone in the suggestion and a different non-homophone in the response, produced with a (L)H*L% contrast intonation. All intonation contours were felicitous: 27 trials in the experiment used confirmation intonation for correct guesses, and 27 trials used contrast intonation for incorrect guesses.

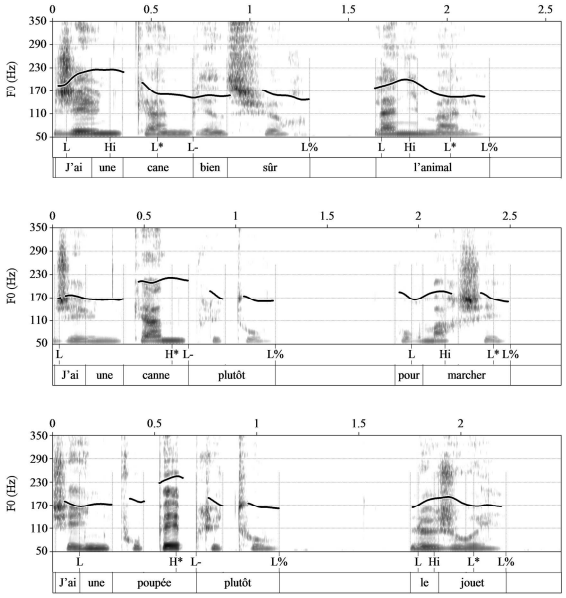

Figure 2: *Spectrograms of the responses in the confirmation-homophone condition (top), the contrast-homophone condition (middle), and the contrast-control condition (bottom).*

2.3. Procedure

Participants' eye fixations were tracked by an Eyelink II Eye-tracker. First, participants were told a story about a girl who had to make a guess about another girl's cards, and then the other girl either confirmed or contradicted that guess. Three practice trials preceded the test phase. The test phase consisted of 18 test trials (6 per test condition) and 36 filler trials (21 filler-confirmation and 15 filler-contrast) presented in randomized order. Test conditions were counterbalanced across three presentation lists using a Latin-square design. Participants were randomly assigned to one of the three lists.

2.4. Predictions

We predicted that before the CW, participants would look equally to the four images because no prosodic or segmental cues would reveal the intended target (although we anticipated that there could be some bias for or against the repeated image, i.e. the duck in Fig. 1). As the beginning of the CW unfolded, we expected that in the confirmation- and contrast-

homophone conditions participants would look equally at the two homophone images since no segmental cues to disambiguate the target nor nuclear tonal ones would have been perceived yet. However, in the contrast-control condition participants were predicted to look less at the suggested image because they would have already perceived segmental disambiguation. We further predicted that once the entire tonal configuration had been perceived by the end of the CW, the two homophone conditions would diverge: participants would look more at the suggested image if a confirmation intonation was presented, but less at it if they heard a contrast intonation. Finally, we expected that during the segmental clarifying phrase, the confirmation condition (but not the other ones) would elicit more looks to the suggested image.

3. Results

Fig. 3 plots fixations to the image depicting the suggested word across the 3 conditions, averaged across participants. The x-axis shows time (ms) from the onset of the CW, and the grey region indicates a 200ms offset to account for saccade planning. Note that looks at the suggested image were similar across conditions before the CW (before the grey region). At the beginning of the CW, and so as soon as participants perceived a small portion of the segmental material, their looks to the suggested image decreased sharply in the contrast-control condition. If we compare the two homophone conditions, we observe that during the CW participants increased their looks at the suggested image even if the intonation signaled contrast. However, shortly before the end of the CW and coinciding with the presence of the H* peak in the conditions with contrast intonation (dotted vertical line in Fig. 3, offset for saccade planning), looks at the suggested image begin to decrease (and do so even more after the CW), coinciding with the region in which the L% fall is realized.

For the statistical analyses we calculated the proportion of fixations (out of trials with valid fixations) on each of the four picture elements, by time steps of 20 ms. We then aggregated the time segments into larger time regions of interest by counting the number of 20ms time slots in which a fixation to the suggested image, the image depicting the homophone competitor, or to the two non-homophones (one of which was the target in the contrast-control condition) occurred. Five regions of interest were analyzed: (1) the region prior to the CW, (2) the region within the CW preceding the H* peak, (3) the 100 ms window where the H* peak was perceived (or not perceived in the LHiL*L% condition), (4) the region following the presence or absence of the H* peak within the CW, and (5) the first 100 ms of the segmental disambiguating region. Time was aligned with the beginning of each CW, and then offset by 200ms for saccade planning. Because CWs had different durations each region of interest was calculated as a function of each word's properties.

We then calculated logodds of looks to the image depicting the suggested word vs. looks to the other three areas of interest in the visual scene. Using this dependent measure we fit linear mixed-effects models using the *lmer* function of the R package *lme4*. Participants and items were treated as random effects to accommodate by-subject and by-item variation in one model [15]–[17]. Condition (confirmation-homophone vs. contrast-homophone vs. contrast-control was included as a Helmert coded predictor to allow comparison of the contrast-control level with the other two levels, and the two homophone conditions to each other. First we present the results for the comparison between the two homophone conditions (confirmation-homophone and contrast-homophone) vs. the contrast-control condition in each region of interest. Second, we report the results of the comparison between the two homophone conditions in each region of interest. Table 2 shows *ß estimates, SEs, t,* and *p* values for all comparisons. We determined *p* values by chi-square tests from nested model comparisons.

For the first comparison, we observed that there was no effect of condition in the region preceding the CW (t = .034, p = .973), but that once the CW started and in all of the following regions of interest there was an effect of condition (beginning of the CW: t = -2.006, p < .05; F0 peak: t = -6.148, p < .001; post-peak within the CW: t = -7.888, p < .001; first 100 ms after the CW: t = -7.786, p < .001). This shows that participants looked less at the suggested image in the control condition than in the other conditions, due to the presence of disambiguating segmental material, and that this effect began when the CW started.

For the comparison between conditions with a homophone in the CW, we observed no effect of condition before the CW (t = .172, p = 0.863), with equal looks to all images at this point. Once the CW started and during the region that preceded the H* peak, we observed a marginal effect of condition (t = 1.852, p = .06), with an unexpected bias to look at the suggested image in the contrast (L)H*L% homophone condition. This early bias became fully significant in the following region (when the H* peak was presented) (t = 1.945 p < .05). Notably, after the H* peak had been perceived and during the following L% fall (i.e. during the rest of the CW),

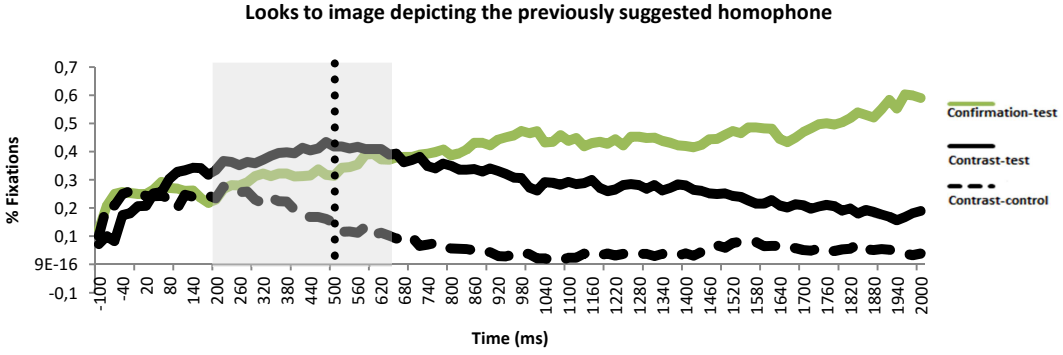

Figure 3: *Proportion of fixations to the suggested image across conditions. The grey region depicts looks when hearing the critical word, and the dotted vertical line indicates looks when perceiving the H target location.*

this bias disappeared (t = 0.613, p = .506), revealing that in the contrast (L)H*L% homophone condition participants started looking less at the suggested image only in response to the falling contour composed of the H* peak and the following fall.

	Estimate	SE	t-value	p-value
Pre-CW				
Intercept	-0.4373	0.0543	-8.046	
Homoph. vs Control.	0.0024	0.0721	0.034	.973
Confirmation vs Contrast	0.0142	0.0829	0.172	.863
Beginning of CW				
Intercept	-0.5973	0.0733	-8.145	
Homoph. vs Control.	-0.2282	0.1138	-2.006	.045*
Confirmation vs Contrast	0.2432	0.131	1.856	.063•
F0 peak				
Intercept	-0.5659	0.1113	-5.086	
Homoph. vs Control.	-0.7322	0.1191	-6.148	.000***
Confirmation vs Contrast	0.2758	0.1418	1.945	.043*
Post-peak within CW				
Intercept	-0.6377	0.1025	-6.217	
Homoph. vs Control.	-0.6378	0.1024	-6.228	.000***
Confirmation vs Contrast	0.0878	0.1432	0.613	.506
First 100 ms after CW				
Intercept	-0.6613	0.0970	-6.814	
Homoph. vs Control.	-0.9078	0.1166	-7.786	.000***
Confirmation vs Contrast	0.0661	0.1341	0.493	.622

Table 2: *Estimates, Standard Errors, t values and p values for all comparisons in all regions of interest.*

4. Discussion

The purpose of this study was to investigate if French speakers associate the "implication contour" with the specific meaning of a contrast between the interlocutors' beliefs, and at which point within the nuclear contour they have fully processed this meaning. In other words, we ask whether listeners can anticipate the contrast by perceiving the LH* rising portion of the contour or if they have to wait until the L% falling tone is processed. The results of an eye-tracking task revealed three main findings. First, and as expected, as soon as participants perceived segmental cues identifying the CW they stopped looking at the image depicting the interlocutor's suggested word if this was not the target. Second, there was an early bias for looks to the suggested image in the contrast-homophone condition. All three critical conditions presented a homophone as the suggested image, and the contrast-homophone condition employed the same contour as the contrast-control condition, yet the difference in looks begins prior to the segmental divergence between the two contrast conditions (see Fig. 3). This finding was unexpected, especially for the region prior to the onset of the CW. Third, and importantly for our study, participants' bias to look at the suggested image decreased in the contrast homophone condition but only after the H* peak had been realized, revealing a rapid effect of intonation (from the L% fall) in reducing the bias to the suggested image and in supporting a contrastive interpretation.

In the contrast-homophone condition we saw that participants increased looks to the suggested image when only pre-nuclear material was available, and that this effect continued as the CW unfolded (reaching full significance in the comparison of the two homophone conditions at the F0 peak). Note that the (L)H*L% implication contour might be more acoustically (hence perceptually) salient in terms of its stronger F0 rise before the L% fall. The increase in perceptual salience might result in increased participants' attention (hence increased looks). Another possible explanation might be related to the pre-nuclear segment in the rise-fall (L)H*L% contour. While the fall LHiL*L% contour begins in an H tone, the rise-fall (L)H*L% contour begins with an L tone. Until the beginning of the rise, the (L)H*L% contour is similar to another intonation contour of the French inventory, the rising LH*H% contour (although the rise starts earlier in the (L)H*L% case) [18]. This rising LH*H% contour is used for continuation declaratives, a pragmatic meaning that is very frequent in French speech. It could be that participants processed the first part of the rise-fall (L)H*L% 'implication' contour as indicating a continuation declarative and not as a contrastive meaning, since the first meaning is more frequent than the second one in French. Future analyses will try to correlate our findings with participants' individual differences. Finally, participants used the prosodic information following the H* peak in the implication contour to extract the contrastive meaning. The results show that the change in looks began only after the H* peak had been processed and in the tonal region where the L% falling tonal movement begins, suggesting that contrast meaning processing in French may require both the H* peak and its subsequent L% fall, or at least part of it, to have unfolded.

Similarly, [11] found that participants needed to perceive a L-H% boundary tone to derive the meaning of implied contrast in English. On the other hand, [7], using a design that made the implied meaning more easily predictable from the presence of an L+H* accent, found effects prior to the occurrence of the boundary tone. Since in French both the H* peak and the fall occur during the last prominent syllable for the implication contour, it can be difficult to tease apart whether participants' decisions were driven by the pitch accent alone or by a combination of the pitch accent and the boundary tone. This is especially true since there are open questions about how much processing time is required to build an implied meaning subsequent to perceiving the acoustic evidence that evokes it. An additional complication relates to how quickly an intonational category can be identified. In a corpus-based analysis of French rise-fall and rise intonation movements, [15] found that some rise-falls can be confused with rises due to delayed H* peak alignment. Our implication contour rise-fall might contain some instances of delayed H* peak, causing then some confusion in meaning processing. Nevertheless, the results are most consistent with a need to process the L% to construct the contrast interpretation.

In sum, results show that listeners associated the (L)H*L% French implication contour with a contrast between speakers' beliefs, and that they started doing this form-meaning mapping immediately following the H* peak of the rise-fall movement. Many questions still remain to be answered, especially since most of the research on the online processing of intonation has been carried out with stress-accent languages like English, German or Dutch which are typologically very different from French. However, we believe the present findings will contribute to a better understanding of some of the cognitive processes and individual differences involved in intonation processing.

5. Acknowledgements

We thank Céline Pozniak for her help running the experiment. This work, carried out within the Labex BLRI (ANR-11-LABX-0036), has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the project title Investments of the Future A*MIDEX (ANR-11-IDEX-0001-02).

6. References

[1] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in interpretation of discourse," in *Intentions in Communication*, P. Cohen, J. Morgan, and M. Pollack, Eds. Cambridge, USA: MIT Press, 1990.

[2] S. Jun and C. Fougeron, "A Phonological Model of French Intonation," .

[3] J. S. German and M. D'Imperio, "The Status of the Initial Rise as a Marker of Focus in French," *Lang. Speech*, 2015.

[4] A. Di Cristo, "Vers une modélisation de l'accentuation du français (seconde partie)," *J. French Lang. Stud.*, vol. 10, pp. 27–44, 2000.

[5] C. Portes, C. Beyssade, A. Michelas, J.-M. Marandin, and M. Champagne-Lavau, "The dialogical dimension of intonational meaning : Evidence from French," *J. Pragmat.*, vol. 74, pp. 15–29, 2014.

[6] C. Portes and U. Reyle, "The meaning of French 'implication' contour in conversation," in *Speech Prosody*, 2014.

[7] C. Kurumada, M. Brown, S. Bibyk, D. F. Pontillo, and M. K. Tanenhaus, "Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings.," *Cognition*, vol. 133, no. 2, pp. 335–342, 2014.

[8] A. Weber, B. Braun, and M. W. Crocker, "Finding referents in time: eye-tracking evidence for the role of contrastive accents.," *Lang. Speech*, vol. 49, no. 681, pp. 367–392, 2006.

[9] D. Dahan, M. K. Tanenhaus, and C. G. Chambers, "Accent and reference resolution in spoken-language comprehension," *J. Mem. Lang.*, vol. 47, pp. 292–314, 2002.

[10] D. G. Watson, M. K. Tanenhaus, and C. a Gunlogson, "Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*.," *Cogn. Sci.*, vol. 32, no. 7, pp. 1232–44, Oct. 2008.

[11] H. Y. Dennison and A. J. Schafer, "Online construction of implicature through contrastive prosody," in *Speech Prosody*, 2010, pp. 1–4.

[12] W. F. Heeren, S. A. Bibyk, C. Gunlogson, and M. K. Tanenhaus, "Asking or Telling – Real-time Processing of Prosodically Distinguished Questions and Statements," *Lang. Speech*, pp. 1–28, 2015.

[13] K. Ito and S. R. Speer, "Anticipatory effects of intonation: Eye movements during instructed visual search," *J. Mem. Lang.*, vol. 58, no. 2, pp. 541–573, 2008.

[14] M. K. Tanenhaus and M. J. Spivey-Knowlton, "Integration of visual and linguistic information in spoken language comprehension," *Science (80-.).*, vol. 268, no. 5217, p. 1632, 1995.

[15] R. H. Baayen, *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press, 2008.

[16] D. J. Barr, "Analyzing 'visual world' eyetracking data using multilevel logistic regression," *J. Mem. Lang.*, vol. 59, no. 4, pp. 457–474, 2008.

[17] T. F. Jaeger, "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," *J. Mem. Lang.*, vol. 59, pp. 434–446, 2008.

[18] C. Portes and L. Lancia, "A corpus based investigation of the contrast between French rise-fall and rise via wavelet based functional mixed models," in *Laboratory Phonology*, 2012.