Perceptual relevance of the French Initial Rise in identifying the left edge of a contrastive focus

Axel Barrault, Pauline Welby, James S. German

Aix Marseille Université, CNRS, LPL, Aix-en-Provence, France
axel.barrault@univ-amu.fr, pauline.welby@univ-amu.fr, james.german@univ-amu.fr

Abstract

This study addresses the interplay between information structure and intonation in French. In a forced-choice perception task, we tested whether the French Initial Rise (Initial Rise) is of relevance in identifying the left edge of a non-corrective contrastive focus. Participants listened to two-clause sentences with parallel structures in the two clauses in which the segmental information of the direct object head noun in the second clause was missing (see Figure 1). Crucially, there was always a contrast on the final color adjective modifying the noun. On the intonational residual, we manipulated the presence (LHILH*) or absence (LLH*) of an Initial Rise. Participants were asked to choose between the Given or a New referent. The latter would reflect a contrastive interpretation of the entire post-verbal argument, while the former would reflect a contrastive interpretation restricted to the final adjectival modifier of the direct object noun phrase. Results showed no bias toward either of the two interpretations as a function of the presence or absence of the Initial Rise as marking the left edge of a contrastive focus. As anticipated, however, the sentences’ parallelism induced an overall bias toward a contrastive interpretation. Several factors that may explain the present null result are discussed under the hypothesis that the Initial Rise is a weak cue to focus.

Index Terms: prosody, intonation, vocoding, non-corrective contrastive focus, perception, parallelism, probabilistic processing.

1. Introduction

In everyday communication, intonational modulations across utterances are pervasive. One of the primary and most extensively studied acoustic correlates is fundamental frequency (F0) which encodes several – potentially interacting – aspects of meaning at different linguistic levels, depending on the language. Typically, F0 modulations express pragmatic meaning (i.e., information structure, speech act marking, belief status, inter alia) in intonational languages such as French or English, but also lexical meaning in tonal languages such as Mandarin Chinese or Thai. Furthermore, F0 also helps to convey paralinguistic information such as the talker’s emotional state as well as indexical information. This multitasking role makes intonation a fundamental element of human communication – and therefore of linguistic research – but also contributes to the difficulty in understanding the mappings between acoustic form and discourse function. Several theoretical proposals have tried to model the mapping between F0 modulations and discourse functions in intonational languages. Whether this mapping is direct or mediated by an abstract phonological level of representation and whether the proposals have argued for holistic or strongly compositional accounts, these models have provided a rich and elaborated conceptual framework and useful methodologies. Yet, a growing body of evidence today demonstrates that different discourse functions rely on the same tonal marking, and conversely, that different marking can be associated with the same discourse function. In other words, there is no one-to-one mapping between tonal form and discourse function. The pervasive variability in the way speech is produced and processed lies in part in the multitasking of prosody. Therefore, investigating the precise role of intonation in listeners’ ability to accurately derive intended meaning has been one of the major challenges of the past decades. In view of these developments, a promising line of research have recently put forth that the non-deterministic regularity of tonal marking may be captured through a probabilistic account, integrating multiple cues to account for the interplay between acoustic forms and discourse functions.

Among the discourse functions cued by tonal marking in intonational languages, information structure [1] has been the topic of many studies. Th information structure of a sentence formally expresses the pragmatic delineation of a proposition in a discourse [2]. It refers to the “packaging” of information in Chafe’s [3] terms, which is independent of logico-semantic meaning, and which reflects the speaker’s effort to optimally direct listeners to retrieve the information encoded in a sentence and enter it into their knowledge-store [4]. Intonational languages nonetheless differ in the extent to which tonal prominence are required to convey information structure and the way they do so. For instance, contrastive focus marking in French is said to rely less on tonal prominence than a language like English, although both languages have the phonetic and phonological means to encode contrastive focus and do so quite similarly [5]. In French intonation, there is typically a final rise in F0 associated with the last full syllable of a non-utterance-final accentual phrase (AP), the basic phrasing unit. Another F0 rise can occur at the left edge of the AP and is often realized at the beginning of the first content word of the AP (the Initial Rise (Initial Rise) or LH in a two-rise LHILH* pattern) [6][7][8]. However, precise placement of the Initial Rise and its conditioning factors vary, as does its informational function. Several factors have been shown to influence the occurrence of this Initial Rise, including syntax [9], speech rate and phonological length [6][8][10], style [11][12], segmental material [5], and crucially, focus structure [9]. Descriptively, intonational encoding of focus has traditionally been claimed to be restricted to corrective cases in French, while English encodes other types of focus, including question-answer congruence [13][14] and non-corrective contrastive focus [15]. In addition, previous work on the prosodic marking of focus has often investigated corrective focus using the term contrastive. Yet, results from an interactive speech production study demonstrated a probabilistic association between the Initial...
Rise and the left edge of non-corrective contrastive focus using wh-interrogatives [16]. These authors argue, in line with the recent probabilistic development of intonation, that the relatively weak association between the non-corrective contrastive focus found in their quantitative experimental assessment can be informative in an interpretative model integrating multiple parameters to Initial Rise occurrence. To the best of our knowledge however, the use of the Initial Rise as a two-part cue [17] to identify the left edge of a focus has not been investigated experimentally.

The current study aimed at experimentally assessing the perceptual relevance of the Initial Rise in signaling the left edge of a contrastive focus. To this end, we designed a two-alternative forced choice (2AFC) perception task using sentences as in Figure 1 involving a particular form of a syntactic structure typical of the oral modality and that we henceforth refer to as parallelism. Consider the example in (1).

(1) He ate a GREEN apple, and then, he ate a RED apple.

By the parallelism of its two constitutive independent clauses, in which a significant part of the first clause linguistic material is repeated in the second clause, this structure naturally involves contrastive focus (e.g., green vs. red). Other types of parallelism have been investigated in English as well as in Spanish [18] (inter alia). This structure has the advantage of allowing a more direct comparison between the English and French intonational systems – and more generally between Germanic and Romance languages [18], although French differs greatly from other Romance languages. Moreover, French has been typically described as relying less on prosody to mark non-corrective contrastive focus. Specific syntactic processes such as clitic and its derivatives, which extract the constituent under focus have been claimed to prototypically express focus in French [2][20]. Although always accompanied by an intonational pattern possibly comprised of a signalling prominence, the study of how intonational structure contributes to contrastive focus perception in these syntactic structures is difficult to untangle from the contribution of syntax to prominence in French. Using parallelism thus allows us to test the hypothesis that the Initial Rise is a perceptual cue to the non-corrective contrastive focus at the left edge. Descriptively, the prediction drawn from [16] is that if the Initial Rise is a cue facilitating the identification of the left edge of a contrastively focused constituent, we should observe that participants choose the New referent – reflecting a contrastive interpretation – over the Given one when they hear the Initial Rise than when they did not.

2. Method

2.1. Participants

A total of 58 native speakers (16 men) of Hexagonal French who did not declare any language impairment or hearing disorder participated in our study. Their age ranged from 18 to 68 years old (mean= 21.6; SD=11.2). All resided in France at the time of the study. They were compensated with gift certificates (valued at €10). The data from four participants were excluded from the analyses (three because of a technical issue that prevented them from completing the task, one who misinterpreted the anaphoric pronoun as a new referent). Of the remaining 54 participants, only the data of participants who chose each of the two types of choices (New, Given) at least once were included in the final statistical analysis. Under this criterion, another five participants’ data were excluded, leaving a total of 49 participants.

2.2. Materials

2.2.1. Stimuli

Participants listened to 56 sentences comprised of two parallel independent clauses like the example in Figure 1.

Figure 1: A typical test sentence

Through its parallel structure, the sentence’s first clause sets up the upcoming contrast within the post verbal argument (i.e., direct object) comprised of a noun phrase (e.g., le monument sur fond vert ‘the monument on a green background’) with a color adjective modifier. In French, adjectives are post-nominal. Yet, this difference does not seem to affect the realization of focus [5]. Therefore, to investigate a post-verbal but non-final intonational contrast in French, the contrast must be made on the head noun, contrary to previous work in English, a language with pre-nominal adjectives (see for instance [19]). In the second clause, the segmental information of the non-final head noun in the post-verbal direct object of the penultimate accentual phrase was rendered inaccessible by a low-pass filtering resynthesis procedure. This was realized through Praat [21] via an adaptation of a script [22]. On the intonational residual, we manipulated the presence (LHiLH*) or absence (LLH*) of an Initial Rise. Minimaliy, there was always a contrast on the final color adjective (e.g., vert ‘green’ vs. rouge ‘red’), regardless of the experimental condition (i.e., the tonal marking, LHiLH* or LLH*). The rising-falling marking (i.e., LH*+L, see Figure 1) on the color in the last AP of the utterance always supported the contrastive new information. The phrasing remained fixed regardless of the condition as did the rest of the intonational patterns of the carrier sentences.

In line with the prediction drawn from the material used in [16], we had to allow the full realization of the tonal sequences (i.e., LHiLH* vs LLH*) on the target word where the final rise is a continuation. French declaratives typically end with a fall, and we aimed at avoiding the interaction with the required Intonational Phrase boundary (i.e., L-L%E) – even under focus. Therefore, the hypothesis tested here would have been less straightforward if the contrastively focused color adjective had immediately followed the target noun since the adjective would have been in utterance-final position. Thus, to allow this full realization, we embedded the contrastively focused color adjective modifying the direct object noun phrase in a prepositional phrase (e.g., sur fond vert, ‘on a green background’) so that the target word was more naturally expected to be realized in its own AP. Furthermore, the sequence sur fond (‘on (a) background’), allows the possibility of observing deaccenting. Fifty-six pairs of trisyllabic nouns were used as target words, avoiding voiceless consonants. Since the probability of observing an Initial Rise increases as a
function of number of syllables of a content word in the AP, especially between 3 to 4 syllables [6][8], and since the absence of the Initial Rise (yielding LLH+) is also quite as common for content word constituents of this size [6], we chose to use three-syllable target nouns on which we manipulated the presence or absence of the Initial Rise. The critical and target AP (see Figure 1) are therefore four-syllables APs. A total of 112 trisyllabic nouns, four bisyllabic verbs, and five monosyllabic color adjectives were selected using the Lexique online database [23]. The nouns were paired such that the two nouns had a similar onset but did not form a minimal pair (e.g., médaillon, ‘medallion’ vs. mémorial, ‘memorial’), to avoid metalinguistic corrective focus, which was not the type of contrastive focus under investigation. Moreover, metalinguistic corrective focus may have its own intonational features and merits its own study. The remaining linguistic material was kept constant in terms of number of syllables in each syntactic constituent, syntactic structure, intonation, and breaks, but varied at the level of the lexicon to consistently induce new referents. All the sentences had the same syntactic structure. We used four bisyllabic transitive verbs (i.e., il/elle: dépasse, franchit, contourne, traverse. ‘s/he passes’, ‘crosses’, ‘goes around’, ‘goes through’). The 56 subject NPs used bisyllabic animal names (e.g., le canard ‘the duck’). The two clauses were conjoined by the same conjunction-adverb sequence (et ensuite ‘and then’). Each set allowed us to generate four test sentences (2 words X 2 intonational tonal patterns) with the same carrier sentence. Each participant heard only one out the four possible sentences for each of the 56 pairs) to always induce new items. The four sentences were therefore distributed across four lists. Participants thus completed 28 test trials in each of the two experimental conditions (Initial Rise vs. No-Initial Rise).

2.2.2. Resynthesis

The test sentences were all recorded in a sound attenuated chamber with the nonsense sequence /ma.ma.ma/ produced by a female native speaker of French, who is a trained phonetician. As a precaution, we used sonorant segments to minimize the disruption of the F0 signal. Additionally, vowel quality has also been found to influence the F0 realization and perception [25][26][27]. Therefore, the same vowel (i.e., /a/) and the same nasal consonant (i.e., /m/) were used for the three constituent syllables of the target noun to keep the quality of the relationship between F0 and the segmental material constant. The number of syllables mirrored that of the test items (i.e., three content word syllables) so that the critical item given in the context clause (e.g., monument) remains an equally probable alternative in the forced choice.

The recorded utterances were annotated on Praat TextGrids and divided in three parts: the /ma.ma.ma/ sequence and the two parts preceding and following it. A copied version of utterances was then entirely resynthesized. The resynthesis by vocoding procedure has been usually used to simulate the effect of cochlear implants (e.g., [28][29]). Descriptively, vocoding consists of an analyzing the main spectral components of the voice signal, then synthesizing a degraded version. In the current study, this second step allowed us to modify designated parameters of the voice signal corresponding to the segmental information. The vocoding procedure involves applying a low-pass filtering processing of the signal masking the segmental information, cleaning the signal, then rescaling of the intensity to fit the loudness of the unprocessed parts, by means of a Praat script [22] adapted for the current study. The resynthesized output sounded like humming (i.e., when a continuous voiced sound is produced without any supraglottic articulation with the mouth closed, and air escapes through the nasal cavity). In sum, vocoding the speech signal was a way to isolate the intonational variable by neutralizing the segmental information and in so doing, neutralizing the disambiguating lexical information. The processed sequence /ma.ma.ma/ was then cross-spliced and inserted in place of the clear recorded version in carrier sentences (see Figure 1). Each stimulus was created by concatenating the non-vocoded first part followed by the vocoded target, followed in turn by the non-vocoded final part. The side of appearance of the binary forced choice on screen was counterbalanced across two sublists in each of the four main lists yielding eight lists, to which participants were randomly assigned. The experiment was implemented and run using OpenSesame 3.2.8 [30].

2.3. Procedure

Participants were tested in groups in a quiet room of the lab with individual computer screens and headphones. Participants were asked to read instructions displayed on their individual computer screens. They were also presented with an animation supporting the game as in Figure 2. Specifically, the goal was to visually support the possibility for a given referent to be repeated in the second clause.

Figure 2: Animation supporting instructions

2.3.1. Instructions: English gloss

The instructions given were as follows.

First screen (English gloss): “You will hear sentences pronounced by the host of a board game. In this game, the pawns are animals. To win, they must progress as fast as possible, square after square until reaching the end of the maze. The board game squares contain objects or characters depicted on different backgrounds. Each player takes turn rolling two dice, one after the other. Each dice indicates the number of squares to be crossed by the pawn. The host will give you the result of each dice rolls by pointing out the path of the pawn as in the following examples.”

Second screen (English gloss): “Here are two examples:

1. « Le crocodile avance à la forteresse sur fond bleu, et ensuite, il avance à la pyramide sur fond rouge. » [‘The crocodile moves to the fortress on a blue background, and then, it moves to the pyramids on a red background.’]

2. « L’anacoda saute à la balançoire sur fond jaune, et ensuite, il saute à la balançoire sur fond rouge. » [‘The anaconda jumps to the swing on a yellow background, then it jumps to the swing on a red background.’]

However, in sentences you will hear, there has been a recording problem. Some information is missing. Your task is...
to retrieve this word. To do so, two answers will be proposed on screen. You will have to indicate your answer by clicking on it to move to the next sentence. You will have to pay close attention. There is only one chance to listen per sentence. Trust your first intuition. Concentrate. Shall we give it a try?"

2.3.2. Training phase

To ensure that they understood the answer selection process, all participants went through the five same training trials. Those trials were recorded with a male voice to maximize the difference with that of the test phase. The same vocoding procedure was applied to the training. However, instead of the direct object, the subject was vocoded, as in (2).

(2) “Le __ avance à l’âne sur fond bleu, puis il avance à l’ours sur fond vert.”

‘The __ goes to the donkey on the blue background, and then it goes to the bear on the green background’

Forced choice: merle, ‘blackbird’ versus loup, ‘wolf’

The vocoded parts were monosyllabic words to avoid eliciting the LHLH* pattern during the recording or triggering expectations about the task to come. None of the items used in the training phase appeared in the test phase.

2.3.3. Test phase

Participants were allowed to go through the 56 test trials without pause at a self-paced rate. For each trial, they were asked to choose between two possible referents displayed on a computer screen: a Given or a New referent (Figure 2). The latter would reflect a contrastive interpretation of the entire post-verbal argument, while the former would reflect a contrastive interpretation restricted to the final adjectival modifier of the noun phrase. To indicate their response, participants press one of two keys of a French keyboard corresponding to the word chosen (‘Q’ for the word displayed on the left side of the screen, ‘M’ for that on the right).

Figure 2: Time course of a test trial.

3. Results

The dependent measure was the proportion of contrastive responses for each intonation pattern. We combined all test trials of the 49 participants. Their age ranged from 18 to 67 years (mean=21.6; SD =11.2; 12 men). We ran a generalized linear mixed regression model with the lmer4 package of R (version 1.1-23) [31], with the presence of Initial Rise as fixed factor and participant as random intercepts. We did not find any significant difference between the two within-subject conditions (B=-0.011, z=-0.132, p=0.89). That is, we did not find any evidence in favor of one of the two tonal patterns marking the left edge of a contrastive focus. Rather, we observed a great deal of variability in the way participants responded to the task. As anticipated however, the sentences’ parallelism overall induced a bias toward a contrastive interpretation.

4. Discussion

This study was meant to experimentally assess whether the Initial Rise, that was found to be associated with the left edge of a post-verbal non-corrective contrastive focus in [16], is a relevant perceptual cue constraining the choice towards a New over a Given referent in an offline two-Alternative Forced Choice task. Specifically, we designed two-clause sentences with parallel structures across the clauses where a non-corrective contrastive focus was minimally present on the last word (i.e., the color adjectival modifier) but that could span from the preceding head noun. Results revealed first that participants were overall slightly biased towards choosing the contrastive noun, and that regardless of the presence or absence of the Initial Rise at the onset of the vocoded part. This result was anticipated considering the parallel structure. More importantly, the presence of the Initial Rise on the target sequence did not influence participants’ choice in this study. We cannot exclude that the Initial Rise may not be informative enough to be used to identify the left edge of a (non-corrective) contrastive focus. However, at least three factors may explain this null result, under the hypothesis that the Initial Rise is a weak cue to focus [16]. First, segmental information might be required to allow listeners to exploit prosodic cues, since we rarely experience this type of acoustic signal in daily communication. The influence of segmental material is known to influence the realization and the perception of F0 signal [24][25][26][27]. Listeners might not have fully engaged in intonational processing without this information. Second, the vocoding manipulation might have partially obscured the tonal information, despite the upscaled intensity. Third, the task may have involved a working memory challenge. Given the length of the sentences to display a contrast against two parallel clauses, the task may have involved a rather high cognitive load preventing an efficient integration of tonal information. Considering that the Initial Rise has been consistently shown to be used in word segmentation [17][32] and given the robust effects in production shown in previous work [16], the null finding of this study is somewhat surprising, though it provides a basis for future studies to better establish how the Initial rise is used by speakers and listeners. At this step, we assume that the hypothesis under which the Initial Rise is a weak cue to the left edge of a non-corrective contrastive focus remains valid.

5. Acknowledgements

This work, carried out within the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

6. References
