Do head gestures function as precursors for prosodic focus marking in the L2?

Ivy Mok¹, Lieke van Maastricht¹, Núria Esteve-Gibert²

¹Centre for Language Studies, Radboud University Nijmegen
²Universitat Oberta de Catalunya
ivy.mok@ru.nl, lieke.vanmaastricht@ru.nl, nesteveg@uoc.edu

Abstract

Prior research on the acquisition of discourse focus marking has traditionally focused on prosodic cues while disregarding visual cues to mark information status. Recently, it has been shown that French children used head gestures to highlight new/contrastive discourse referents before developing the necessary prosodic cues. As prosodic discourse marking is challenging for both L1 and L2 learners, we investigate whether this entrainment function of head gestures in prosodic focus marking also occurs in an L2 context. Hence, Catalan/Spanish learners of English were audio-visually recorded while producing semi-spontaneous utterances in three focus conditions (broad focus; contrastive narrow focus; corrective narrow focus). We analyzed (alignment between) prosodic (F0 and syllable duration) and gestural (gesture type and phases) correlates of focus. The results show a longer duration for words in the contrastive and corrective conditions than in the broad focus condition, but no effect of focus condition on pitch range or gesture presence. Moreover, pitch range in gesture-accompanied words was always higher than pitch range in non-gesture-accompanied words, irrespective of focus condition. These results imply that gesture and prosody may be so tightly coupled in this context that L2 learners emphasize the same, possibly inaccurate, part of the utterance in both modalities.

Index Terms: focus marking, L2 acquisition, head gestures, prosody, intonation

1. Introduction

Speakers use prosody to express linguistic meaning, for instance, to mark new and given information in discourse, e.g., [1]. Prior work has shown that languages use varying prosodic strategies for focus marking [2] and that these cross-linguistic differences complicate prosodic focus marking for L2 speakers. Hence, Spanish learners of Dutch and Dutch learners of Spanish generally produce pitch accent distributions that are typical of their L1 but not L2 when marking focus, even if they are advanced L2 speakers overall [3]. Moreover, these prosodic transfer effects are perceivable by L1 listeners, who consider L2 speech with prosodic transfer in pitch accent distributions to mark focus to be more foreign-accented and more difficult to understand than L2 speech with on-target pitch accent distributions to mark focus [4].

The prosodic marking of discourse focus is not only difficult for adult L2 learners; the development of prosody-gesture alignment for functional purposes (i.e., focus marking) is also challenging in L1 development. Focus marking through adult-like intonation patterns does not usually occur until children are 7 or 8 years old [5][6], even though infants can use other ways to mark focus from a much earlier age: [7] showed that 4 or 5-year-olds express information status non-verbally, even if they have not yet learned prosodic cues for that purpose. Their results showed a higher head gesture rate in the narrow focus items than in the broad focus items, but no effect of focus condition on word duration nor F0 range. 73% of the head gestures were aligned with the focused item. Hence, they conclude that head gestures function as an entrainment mechanism, working as an anchor for prosody to synchronize with and hence aiding prosody acquisition in the L1. This multimodal perspective on L1 development is in line with studies on multi-modality in adult speech, which have shown that body gestures are generally temporally aligned with tonal movements in adult L1 speech (e.g., [8-9]) causing an increase in the amplitude of speech, in combination with acoustic peaks in the F0 [10-11]. In addition, producing head gestures leads to a more pronounced prosodic focus marking in the form of higher F0 and duration excursions, which is also noticeable for L1 listeners [9, 11].

In L2 development, results are less straightforward: Some studies have shown that producing body movements and gestures facilitate L2 prosody production (e.g., [12]), while others found the opposite (e.g., [13]). To our knowledge, our study is the first to investigate head gestures as precursors and possible reinforcers of L2 prosodic focus marking. To determine whether head gestures can boost the production of prosodic marking of focus in an L2 context as it does in infants learning their L1, we asked Catalan/Spanish learners of English to interact with a fictional character in a game and ask her to grab objects of varying shape and color out of a bag. By manipulating the shape and color of the objects in the bag, we controlled the information status of the target items, eliciting utterances in which the target item was either in broad, narrow, or contrastive focus. Based on [3] and [14], we expect L2 learners to transfer their L1 prosody to their L2. In this context, in which we limited the option of word order changes, this means that we would expect Catalan/Spanish learners to place the nuclear pitch accent on the final element of the noun phrase, irrespective of its focal status (e.g., [15][16]). Contrastively, if no L1 to L2 transfer occurs we may expect Catalan/Spanish learners of English to produce the nuclear pitch accent on the focused element of the noun phrase, irrespective of its position in the utterance (final or pre-final) [1][17].

To our knowledge, there are no L2 studies on nonverbal focus marking using head gestures to base gesture transfer predictions on. If gestures serve as a precursor for L2 prosody acquisition in focus marking as they seem to do in the L1 [7], learners will use L2 gesture strategies even when transferring their L1 prosody to the L2. This would mean that they produce more head gestures with focused than unfocused elements and that their gestures align with the focused element in the noun phrase. Alternatively, gesture and prosody may be so tightly
coupled [10] that learners place prominence on the same, possibly inaccurate, part of the utterance in both modalities.

2. Methods

2.1 Participants

Data analysis is currently ongoing. So far, twelve (10 female, 2 male) adult Catalan/Spanish bilingual learners of English of the Universitat Oberta de Catalunya voluntarily participated in the study. Their mean age was 38 years (range: 23-52 years) and their English proficiency ranged from A1 to C1 via self-assessment. LexTALe scores ranged from 40 to 86 (out of 100).

2.2 Materials

Each trial had a visual display that showed Anna (the main character in the game) standing in the bottom left corner of the screen with her hand stretched out and in the top right corner an image visualizing the task that Anna should perform and the item that she should grab from the bag to perform that task. For instance, the instruction could be “Tell Anna which colored item she should grab to open the door in combination with a picture of the task (e.g., a closed door) with the target item (e.g., a green nine) on the top right corner. Finally, the middle of the screen displayed a big bag of Anna’s items (e.g., a red L, a blue M, a green 9, and a pink 9), see Figure 1. Using four monosyllabic items shapes (an M, an L, a 9, and a 1) and four colors (pink, red, green, and blue), the items that were available in the bag were manipulated and different focus types were elicited.

Participants produced a total of 40 sentences, which were evenly divided across 5 conditions (i.e., 8 sentences per condition). Each elicitation typically contained: an article, a monosyllabic adjective, and a monosyllabic noun. The five conditions were: a broad focus condition (e.g., Take the green nine, which occurred with only one item in the bag) and 4 types of narrow focus conditions which always occurred with several items in the bag: contrastive narrow focus on the Adjective; contrastive narrow focus on the Noun, corrective narrow focus on the Adjective, and corrective narrow focus on the Noun.

Given that there were several items in the bag, contrastive focus was elicited by having two items of the same color or shape in the bag, one of which was the target. For instance, if the “green nine” was the target item but there was also a “pink nine” present in the bag, then it was expected that the participant would produce an utterance in line with the contrastively focused adjective. If no L1-L2 transfer effects would occur in either gesture or prosody, one might expect participants to produce “green nine” with prosodic and gestural cues on the word “green” to show a contrast between the colors. If Anna had grabbed the pink nine instead of the green nine, the participants were asked to repeat their instruction to correct her mistake. This should elicit corrective focus on the previously contrastively focused element. Participants showing no L1-L2 transfer were hence expected to produce “green” with even more prominent prosodic and gestural cues to correct Anna’s mistake for grabbing the wrong item, see Figure 2.

2.3 Procedure

In the experiment, which was conducted online, participants were presented with the stimuli in a PowerPoint presentation while interacting with the experimenter via Zoom. The presenter controlled the pace of the PowerPoint presentation. Prior to the experiment, participants had filled in a questionnaire regarding their language background and performed the LexTALe (Lexical Test for Advanced Learners of English) [18] as an additional indication of their L2 English proficiency. The experimental session was video- and audio-recorded and took 20-30 minutes. All participants consented to be audio-visually recorded.

2.4 Prosodic and gestural analysis

Praat [19] was used for the acoustic annotation of the data. In each trial, the speech was orthographically transcribed and

Figure 1. Experiment instructions (left) and example items for broad focus (middle, right).

Figure 2. Example of items eliciting narrow (left, middle) and corrective (right) focus.
segmented at the word level (which coincided with the syllable level, since all nouns and adjectives used in the experiment consisted of only one syllable). All nouns and adjectives were coded according to their focus condition as either AN (broad focus), AFN (narrow focus on noun), ACN (corrective focus on noun), FAN (narrow focus on adjective), and CAN (corrective focus on noun). Optionally, items that were out of the ordinary were marked as NC (non-canonical). Such items include switching the order of the adjective and noun or other variations of the noun phrase. For each adjective and noun, we extracted duration and pitch range, measured as the difference in Hz between the F0 maximum and F0 minimum of the word.

ELAN [20] was used for the head gesture coding. Tier 1 included annotations of presence or absence of a head gesture (none / adjective / noun / primarily on adjective / primarily on noun). If both the adjective and the noun contained a head gesture, “primarily on...” was coded on the more visually salient gesture. If Tier 1 was coded anything other than “none”, Tier 2 was coded by gesture type (head nod / head tilt / head other / chin forward / body forward / eyebrow raising / eyebrows + nod / eyebrows + tilt / eyebrows + chin / eyebrows + head other / fuzzy). Unconventional gesture types were marked with a comment in Tier 3, see Figure 3.

![Figure 3: Example trail as analyzed in ELAN.](image)

3. Results

To evaluate whether prosodic cues and head gestures were used to mark focus by Catalan/Spanish learners of English, we conducted several Linear Mixed-Effect Model analyses for the prosodic measures and one Generalized Linear Mixed Effect Model analysis for the gestural measure in R and Rstudio [21] using the lme4 package [22]. In the prosodic analysis, we investigated whether the predictor variables focus condition (5 levels: AN, FAN, AFN, CAN, CAN, with item as a random factor), and word position (2 levels: adjective or noun, with participant as a random factor), affected the outcome variables duration and pitch range. In the gestural analysis, we examined whether these same predictor variables affected the outcome variable head gesture presence.

3.1. Duration

The results showed that there was a significant effect of both focus condition ($\chi^2(2) = 6.06, p < .05$), Intercept (broad focus): $\beta = 0.37, SE = 0.02, t = 15.78$; contrastive focus: $\beta = 0.02, SE = 0.01, t = 2.38$; corrective focus: $\beta = 0.02, SE = 0.01, t = 2.74$); and word position ($\chi^2(1) = 16.11, p < .001$), Intercept (Adjective): $\beta = 0.40, SE = 0.02, t = 19.05$; Noun: $\beta = 0.07, SE = 0.01, t = 5.82$) on duration, but no interaction effect between focus condition and word position ($\chi^2(2) = 1.89, p = .38$). This shows that target words in the contrastive and corrective conditions were on average longer than the target words in the broad focus condition. Similarly, target words in the phrase final position (Nouns) tended to be longer than words in the phrase non-final position (Adjectives), see Figure 1.

![Figure 4. Duration of target word in seconds by focus condition and word position (A- adjective, N-noun)](image)

3.2. Pitch range

A second model revealed a main effect of word position on pitch range ($\chi^2(1) = 7.01, p < 0.01$), Intercept (Adjective): $\beta = 97.56, SE = 14.72, t = 6.63$; Noun: $\beta = -25.41, SE = 8.24, t = -3.08$), pitch range being wider when the target word was in the phrase non-final position (thus an adjective) than when the target word was in the phrase final position (thus a noun), see Figure 5.

![Figure 5. Pitch range by focus condition and word position (A-adjective, N-noun)](image)
Figure 2. No main effect of focus condition was found ($\chi^2 (2) = 2.288, p < 0.318$), nor a two-way interaction between word position and focus condition ($\chi^2 (2) = 1.513, p = .469$).

3.3. Gesture production

The results for gesture production showed that there was a significant effect of word position on head gesture presence (Intercept (Adjective): $\beta = 0.82, \text{SE} = 0.25, z = 3.26, p < .001$; Noun: $\beta = -0.42, \text{SE} = 0.14, z = -2.89, p < .01$). Like pitch range, more gestures were produced when the target word was in the phrase non-final position (Adjective) than when the word was in the phrase final position (Noun). However, no main effect of focus condition ($\chi^2 (2) = 2.965, p < 0.227$) and no two-way interaction effect was found, see Figure 3.

3.4. Gesture-Prosody coordination

In order to explore the coupling of gesture and prosody, we also investigated the effect of head gesture presence on pitch range and duration. There was a significant effect of gesture presence on pitch range ($\chi^2 (1) = 9.87, p < .01$; Intercept (no gesture): $\beta = 67.43, \text{SE} = 11.17, t = 6.04$; gesture: $\beta = 23.13, \text{SE} = 5.99, t = 3.86$), but no significant effect of focus condition ($\chi^2 (2) = 0.714, p = .699$), nor was there a significant interaction between head gesture presence and focus condition on pitch range values ($\chi^2 (2) = 3.369, p = .185$). Moreover, the analysis revealed no significant effects of focus condition or head gesture presence on duration, either on their own ($\chi^2 (2) = 3.208, p = .21$ and $\chi^2 (1) = 0.031, p = .861$, respectively) or in interaction ($\chi^2 (2) = 1.365, p = .505$), see Figure 4.

4. Conclusion & Discussion

This study set out to determine whether head gestures can serve as an anchor for prosodic during focus marking in the L2, as they have been shown to do in the L1, where gestures appeared entrainment prosody in this context. Our results show that although Catalan/Spanish learners of English do vary syllable duration, pitch range, and head gestures in a systematic way, they do not seem to do so to reflect focus marking. While they produced longer syllable durations in contrastive and corrective narrow focus conditions, the effect barely reached significance and will have to be replicated with analysis of the complete dataset (25 participants). Even if the effect of word position on duration holds up, it is likely to be an artifact of final lengthening, which is employed both in Spanish and English [23-24]. For pitch range, as well as head gesture presence we found higher values for noun phrase non-final words than noun phrase final words. In other words, adjectives generally were produced with a larger pitch range and more head gestures than nouns, irrespective of whether they were in focus or not. This finding is interesting because it is not in line with previous studies on prosodic transfer from the L1 to the L2 in general and L1 to L2 transfer of prosody to mark focus in Romance and Germanic languages in particular. These predict that if prosodic cues are not varied as a function of focus marking (as is typically the case in Romance languages), the last function word in the utterance will receive prosodic prominence [1-3, 14-15]. In our study, this means that we would expect the noun to receive prosodic prominence irrespective of its focal status. Instead, our participants produced adjectives with a larger pitch range and more head gestures. Perhaps they are aware of a difference in this context between their L1 and the L2 but hypercorrect themselves by “over-stressing” the adjective. More research is needed to investigate and explain this effect.

In sum, it appears that, contrary to the L1 learners, Catalan/Spanish learners of English do have trouble marking focus both prosodically and gesturally. Whereas the French 4 and 5-year-olds in [7] were unable to mark focus acoustically but did use head gestures to do so, the adult L2 learners in the current study did not systematically use duration or pitch range to mark focus, but neither did they vary their head gestures as a function of focus marking. Instead, head gesture production and pitch range appeared to be coupled (in line with [10-11]) and most prominent on non-final utterance elements (i.e., adjectives). This seems to imply that speech and gesture are coupled in this context, which results in both modalities sometimes being most prominent during the production of non-focused elements. While this leads us to conclude that head gestures do not function as a biomechanical bootstrapping mechanism in the L2 as they do in the L1, recent studies on the use of gesture as a practical resource in L2 classrooms [12, 25] suggest that combining both modalities in the acquisition of L2 prosody remains a promising avenue for future research.

5. Acknowledgements

All students from the Universitat Oberta de Catalunya gratefully acknowledged for their participation. We thank Matheus Azevedo for his help in preparing the data for acoustic analysis.
6. References


